

From Hand-Crafted to Deep Learning-based Cancer Radiomics: Challenges and Opportunities

Parnian Afshar[†], *Student Member, IEEE*, Arash Mohammadi[‡], *Senior Member, IEEE*, Konstantinos N. Plataniotis[‡], *Fellow, IEEE*, Anastasia Oikonomou*, and Habib Benali[†], *Member, IEEE*

[†]Concordia Institute for Information Systems Engineering (CIISE), Concordia University, Montreal, Canada

[‡] Department of Electrical and Computer Engineering, University of Toronto, Toronto, Canada

*Department of Medical Imaging, Sunnybrook Health Sciences Centre, University of Toronto, Toronto, Canada

[†]PERFORM Centre, Electrical and Computer Engineering Department, Concordia University, Montreal, Canada

Abstract—Recent advancements in **signal processing and machine learning coupled** with developments of electronic medical record keeping in hospitals and the availability of extensive set of **medical images** through internal/external communication systems, **have resulted in a recent surge of significant interest in “Radiomics”**. Radiomics is an emerging and relatively new research field, which refers to extracting semi-quantitative and/or quantitative features from medical images with the goal of developing predictive and/or prognostic models, and is expected to become a critical component for integration of image-derived information for personalized treatment in the near future. The conventional Radiomics workflow is typically based on **extracting pre-designed features (also referred to as hand-crafted or engineered features) from a segmented region of interest**. Nevertheless, recent advancements in deep learning have caused trends towards **deep learning-based Radiomics (also referred to as discovery Radiomics)**. Considering the advantages of these two approaches, there are also **hybrid solutions developed to exploit the potentials of multiple data sources**. Considering the variety of approaches to Radiomics, further improvements require a comprehensive and integrated sketch, which is the goal of this article. This manuscript provides a unique interdisciplinary perspective on **Radiomics by discussing state-of-the-art signal processing solutions in the context of Radiomics**.

Index Terms: Radiomics, Deep Learning, Hand-Crafted Features, Medical Imaging.

I. INTRODUCTION

The volume, variety, and velocity of medical imaging data generated for medical diagnosis are exploding. Generally speaking, medical diagnosis refers to determining the source and etiology of a medical condition. **Diagnosis is typically reached by means of several medical tests, among them biopsy and diagnostic imaging, in case of suspected cancer**. Although biopsy can **be very informative, it is invasive and by being focal, may not represent the heterogeneity of the entire tumor**, which is crucial in cancer prognosis and treatment. In contrast to biopsy, diagnostic imaging is not invasive and can provide information on tumor’s overall shape, growth over time, and heterogeneity, making it an attractive and favored alternative to biopsy. Interpretation of such a large amount of diagnostic

images, however, highly depends on the experience of the radiologist and due to the increasing number of images per study **can be time-consuming**.

Referred to as “Radiomics” [1]–[4], the ability to process such large amounts of data promises to decipher the encoded information within medical images; **Develop predictive and prognostic models to design personalized diagnosis; Allow comprehensive study of tumor phenotype [5], and; Assess tissue heterogeneity for diagnosis of different type of cancers**. More specifically, Radiomics refers to the process of **extracting and analyzing several semi-quantitative (e.g., attenuation, shape, size, and location) and/or quantitative features (e.g., wavelet decomposition, histogram, and gray-level intensity) from medical images with the ultimate goal of obtaining predictive or prognostic models**.

Although several challenges are in the way of bringing Radiomics into daily clinical practice, it is expected that **Radiomics become a critical component for integration of image-driven information for personalized treatment in the near future**.

It is worth mentioning that **computer aided diagnosis (CAD)** is not a new concept, and researchers have developed automatic systems to **investigate the link between imaging-based features and biological characteristics in the past**. However, **this field is formalized as “Radiomics”, since 2010 [6], and it has a few key differences with the traditional CAD systems**. First of all, **CAD systems incorporate much fewer number of features (typically, within 8 to 20 features), whereas in Radiomics, hundreds to thousands of features are extracted**. Second, **the application of the CAD systems is, typically, limited to the diagnosis of the diseases, such as distinguishing between benign and malignant masses**. **Nevertheless, Radiomics is a much broader field, including both predictive and prognostic applications [7]**. The first comprehensive clinical application of Radiomics was performed by Aerts *et al.* [5] with involvement of 1019 lung cancer patients. More than **400 different intensity, shape, texture, and wavelet features were extracted from Computed Tomography (CT) images and used together with clinical information and gene expression data to develop Radiomics heat map**, which shows the association between Radiomics and different clinical outcomes such

This work was partially supported by the Fonds de Recherche du Québec Nature et Technologies (FRQNT) Grant 206591 EQ. Corresponding Author is Arash Mohammadi, email: arash.mohammadi@concordia.ca

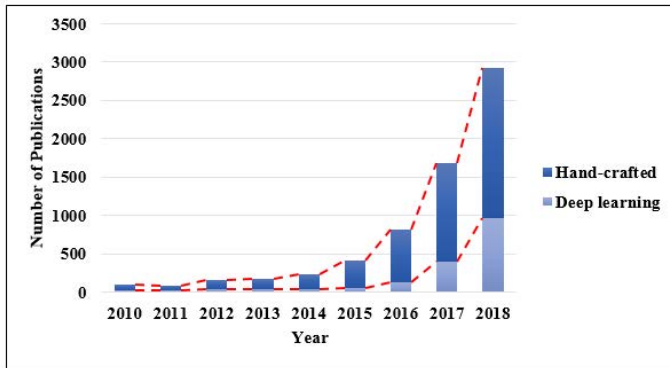


Fig. 1: Increasing interest in Radiomics based on data from Google Scholar (“Radiomics” is used as the keyword). It is observed that there is an increasing interest in both types of the Radiomics.

as cancer stage. This clinical study has illustrated/validated effectiveness of Radiomics for tumor related predictions and showed that Radiomics has the capability to identify lung and head-and-neck cancers from a single-time point CT scan. Consequently, there has been a recent surge of interest [8]–[12] on this multidisciplinary research area as Radiomics has the potential to provide significant assistance for assessing the risk of recurrence of cancer [13]; Evaluating the risk of radiation-induced side-effects on non-cancer tissues [14], and; Predicting the risk for cancer development in healthy subjects [14]. In a very recent article by Vallieres *et al.* [16], it is shown that the Radiomics features extracted in Reference [5] have a noticeable dependency on the tumor volume, which is a strong prognostic factor, and revised calculations are proposed that are less correlated to the tumor volume. In other words, more powerful Radiomics features and procedures are being introduced, illustrating the ongoing research potentials of Radiomics.

The key underlying hypothesis in the Radiomics is that the constructed descriptive models (based on medical imaging data, sometimes complemented by biological and/or medical data) are capable of providing relevant and beneficial predictive, prognostic, and/or diagnostic information. In this regard, one can identify two main categories of Radiomics. Conventional pipeline based on Hand-Crafted Radiomic features (HCR) that consists of the following four main processing tasks: (i) Image acquisition/reconstruction; (ii) Image segmentation; (iii) Feature extraction and quantification, and; (iv) Statistical analysis and model building. On the other hand, the Deep Learning-based Radiomics (DLR) pipeline has recently emerged which differs from the former category since deep networks do not necessarily need the segmented Region Of Interest (ROI), and their feature extraction and analysis parts are partially or fully coupled. We will elaborate on these properties in section IV.

More clinical studies are being approved and conducted to further investigate and advance the unparalleled opportunities the Radiomics posed to offer for clinical applications. Information Post I provides an overview of different screening technologies used within the Radiomics pipeline along with supporting data sources and available datasets to develop Radiomics-based predictive/prognostic models. While

Radiomics consists of a wide range of (partially interconnected) research areas with each individual branch possibly worth a complete exploration, the purpose of this article is to provide an inclusive introduction to Radiomics for the signal processing community, as such, we will focus on the progression of signal processing algorithms within the context of Radiomics. In brief, we aim to present an overview of the current state, opportunities, and challenges of Radiomics from signal processing perspective to facilitate further advancement and innovation in this critical multidisciplinary research area. As such, the article will cover the following four main areas:

- (i) **Hand-crafted Radiomics**, where we introduce and investigate different feature extraction, feature reduction, and classification approaches used within the context of Radiomics. Most of the techniques utilized in any of the aforementioned steps lie within the broad area of “Machine Learning,” where the goal is to improve the performance of different computational models using past experiences (data) [17]. In other words, the underlying models are capable of learning from past data, leading to the automatic process of prediction and diagnosis. Furthermore, since hundreds of Radiomics features are extracted, an appropriate feature selection/extraction strategy should be adopted to reduce the “curse of dimensionality” and overfitting of the prediction models. Most of these strategies, themselves, lie within the field of “Machine Learning”, as they are aimed to learn the best set of features, based on the available data.
- (ii) **Deep learning-based Radiomics**, where we provide an overview of different deep architectures used in Radiomics along with interpretability requirements.
- (iii) **Hybrid solutions** developed to simultaneously benefit from the advantages of each of the above two mentioned categories.
- (iv) **Challenges, Open Problems, and Opportunities**, where we focus on the limitations of processing techniques unique in nature to the Radiomics, and introduce open problems and potential opportunities for signal processing researchers.

Fig. 1 shows the increasing interest in Radiomics within the research community. Although there have been few recent articles [3], [18] reviewing and introducing Radiomics, to the best of our knowledge, most of them are from outside the signal processing (SP) community. References within the SP society such as the work by J. Edwards [19] have investigated recent advancements in medical imaging devices and technologies without reviewing the role of Radiomics in medical applications. Other existing papers outside SP community (e.g., [3]) have failed to clearly describe the underlying signal processing technologies and have narrowed down their scope only to hand-crafted Radiomics and its diagnosis capability. While Reference [18] has briefly touched upon the deep learning pipeline as an emerging technology that can extract Radiomics features, it has not studied applicability of different deep architectures [20] and left the interpretability topic untouched. Furthermore, Reference [21] has mostly focused on the hand-crafted Radiomics, while deep learning-based Radiomics is

explained briefly without addressing different architectures, interpretability, and hybrid models. Although both types of Radiomics are covered in Reference [22], combination of hand-crafted and deep learning-based features are not considered. Besides, challenges associated with Radiomics and the relation between Radiomics and gene-expression (Radiogenomics) are also not discussed thoroughly. Finally, the scope of Reference [23] is limited to deep learning-based Radiomics, without addressing hand-crafted features, their stability, hybrid Radiomics, and Radiogenomics. All these call for an urgent and timely quest to introduce Radiomics to our community

especially since SP is one of the main building blocks of the Radiomics.

The reminder of this article is organized as follows: first in Section II, we will discuss several applications of Radiomics in cancer-related fields, followed by Hand-Crafted solutions in Section III. The Deep learning-based Radiomics is presented in Section IV, where several aspects of DLR is investigated. In Section V, we explain different hybrid solutions to Radiomics, which aim to take advantage of both DLR and HCR. Finally in Section VI various challenges and opportunities of Radiomics, especially for SP community, are discussed. We conclude our work in Section VII.

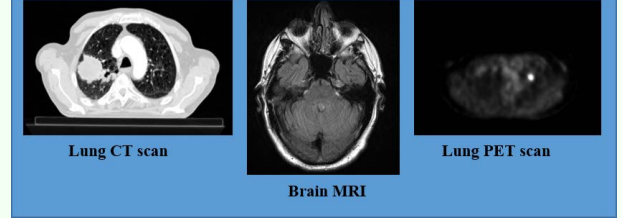
Information Post I: Radiomics Supporting Resources

Several potential medical resources provide information to the **Radiomics pipeline**, some of which are directly used to extract Radiomics features and some serve the decision making process, as complementary information sets. Below we review the most important data resources for Radiomics.

Screening Technologies: The Radiomics features can be extracted from several imaging modalities, among which the following are the most commonly used modalities:

- **Computed Tomography (CT) Scans:** The CT is the modality of choice for the diagnosis of many diseases in different parts of the body, and by providing high resolution images [1] paves the path for extracting comparable Radiomics features. Nonetheless, the CT imaging performance depends on different components of the utilized protocol including the following three main properties: (i) **Slice thickness**, which is the distance in millimeter (*mm*) between two consecutive slices; (ii) **The capability for projecting the density variations into image intensities**, and; (iii) **Reconstruction algorithm**, which aims at converting tomographic measurements to cross-sectional images. Although protocols for specific clinical indications are usually similar across different institutions, Radiomics features can even differ between different scanners with the same settings [24]. Therefore, there is still a considerable need to ensure consistency of Radiomics feature extraction amongst different scanners and imaging protocols [2]. CT images are typically divided into two categories [25]: screening and diagnostic. While screening CT uses low dose images, diagnostic CT utilizes high dose and is of higher quality contrast.
- **Positron Emission Tomography (PET) Scans:** The PET is a nuclear imaging modality that evaluates body function and metabolism [1], and since its performance depends on not only the scanner properties, but also the dose calibration, similar to the case with the CT scans, standardizing the PET protocols across different institutions is challenging. Furthermore, glucose level at the time of scanning can also affect the properties of PET images [2].

- **Magnetic Resonance Imaging (MRI):** Unlike CT, properties of MRI images are not directly associated with tissue density and specific methods are required to obtain the so-called signal intensity. Besides, several imager and vendor-dependant factors such as gradient and coil systems [26], pulse sequence design, slice thickness, and other parameters such as artifacts and magnetic field strength affect the properties of the MRI images [2], which should be consistent across different institutions.



Complimentary Data Sources: In addition to imaging resources, the following clinical data sources are typically combined with Radiomics features:

- **Gene expression:** The process of converting DNA to functional product to have a global insight of cellular function.
- **Clinical characteristics:** Patient's characteristics such as age, gender, and past medical and family history [2].
- **Blood Bio-markers:** Measurable characteristics from the patient's blood such as glucose level, cholesterol level and blood pressure.
- **Prognostic Markers:** Markers to evaluate the progress of the disease, response to treatment or survival, such as size, tumor stage, tumor recurrence, and metastasis.

Radiomics Image Databases: Large amount of data is typically required to get reliable results about tissue heterogeneity based on Radiomics [2]. Table I below introduces a few publicly available imaging data source that can be used to develop and test new Radiomics approaches:

TABLE I: Popular data sets for performing Radiomics.

Data Set	Reference	Year	Imaging Modality	Type of the Tumor	Number of Patients	Annotation Type
LIDC-IDRI	[27]	2015	CT	Lung Tumor	1010	Multiple experts
NSCLC-Radiomics	[5]	2014	CT	Lung Tumor	422	One expert
NSCLC-Radiomics-Genomics	[5]	2014	CT	Lung tumor	89	-
LGG-1p19qDeletion	[28]	2017	MRI	Brain Tumor	159	One expert
Head-Neck-PET-CT	[29]	2017	PET, CT	Head-and-Neck Cancer	298	-
BRATS2015	[30]	2015	MRI	Brain Tumor	65	Multiple experts

II. APPLICATIONS

It is undeniable that **automatic diagnosis systems are still in their infancy**, and **there is a long way before they can be reliably used in clinical applications**. Having said that, several recent studies have investigated automatic diagnosis systems and compared them against human experts. For instance, Esteve *et al.* [31] have developed a deep CNN for skin cancer classification, using a dataset of 129,450 clinical images. The performance of this system is tested against 21 board-certificated experts, and results show that the performance of the system is on a par with human experts. This study suggests

that the automatic diagnosis systems have the potential to achieve a human-level performance and can be utilized as one of the two experts. Similar to other emerging technologies, automatic diagnosis systems have different advantages and disadvantages. For instance, these systems have the potential to improve the quality of clinical care and decrease the number of medical errors. These systems, however, are associated with a major concern, i.e., **the risk of violating patients' privacy**, calling for strict regulations to ensure the privacy of the clinical information [32].

In recent years, Radiomics has been applied to many health-

TABLE II: Applications of Radiomics.

Reference	Imaging Modality	Information Sources	Application Domain	Radiomics Method	Number of Patients	Type of Annotation
Zhang <i>et al.</i> [8]	CT	-	Prediction of lung cancer recurrence and death.	Hand-crafted Radiomics (HCR)	112	One expert
Aerts <i>et al.</i> [5]	CT	Gene-expression and Clinical data.	Lung, and head & neck cancer survival prediction.	HCR	1019	Multiple experts
Griethuysen <i>et al.</i> [9]	CT	-	Lung cancer benign and malignant classification.	HCR	302	Multiple experts
Oikonomou <i>et al.</i> [4]	CT, PET	Standardized uptake value (ratio of image and body radioactivity concentration).	Lung cancer survival prediction.	HCR	150	One expert
Kumar <i>et al.</i> [64]	CT	-	Lung cancer benign and malignant classification.	Deep Learning-based Radiomics (DLR)	1010	Multiple experts
Kumar <i>et al.</i> [41]	CT	-	Lung cancer benign and malignant classification.	DLR	97	Multiple experts
Huynh <i>et al.</i> [42]	Mammogram	-	Classification of breast cancer: benign or malignant.	HCR, DLR, Combination of HCR and DLR	219	Semi-automatic
Li <i>et al.</i> [43]	MRI	-	IDH1 enzyme mutation prediction.	DLR	151	Automatic
Sun <i>et al.</i> [10]	CT	-	Lung cancer benign and malignant classification.	HCR, DLR	1018	Multiple experts
Jamaludin <i>et al.</i> [44]	MRI	-	Disc abnormality classification.	DLR	2009	
Liu <i>et al.</i> [48]	MRI	-	Prostate cancer diagnosis.	HCR, DLR	341	Not performed
Oakden-Rayner <i>et al.</i> [45]	CT	-	Longevity prediction.	HCR, DLR	48	Semi-automatic
Paul <i>et al.</i> [46]	CT	-	Lung cancer short/long-term survival prediction.	Combination of HCR and DLR	81	Semi-automatic
Fu <i>et al.</i> [47]	CT	-	Lung tumor detection.	Combination of HCR and DLR	1010	Not performed
Bickelhaupt <i>et al.</i> [68]	Mammogram	-	Classification of breast cancer: benign or malignant.	HCR	50	One expert

TABLE III: Applications of Radiomics (Continued).

Reference	Imaging Modality	Information Sources	Application Domain	Radiomics Method	Number of Patients	Type of Annotation
Lao <i>et al.</i> [49]	MRI	Clinical risk factors.	Brain cancer survival prediction.	Combination of HCR and DLR	112	Multiple experts
Antropova <i>et al.</i> [52]	Mammogram, Ultrasound, MRI	-	Breast cancer benign and malignant classification.	Combination of HCR and DLR	2060	Semi-automatic
Wang <i>et al.</i> [15]	CT, PET	Standardized uptake value.	Lung cancer benign and malignant classification.	HCR, DLR	168	One expert
Shen <i>et al.</i> [55]	CT	-	Prediction of lung tumor malignancy likelihood.	DLR	1010	Not performed
Emaminejad <i>et al.</i> [65]	CT	Genomics bio-markers.	Lung cancer recurrence risk prediction	HCR	79	Semi-automatic
Sun <i>et al.</i> [12]	CT	-	Lung cancer benign and malignant classification.	HCR, DLR	1018	One expert
Kim <i>et al.</i> [57]	CT	-	Lung cancer benign and malignant classification.	Combination of HCR and DLR	20	One expert
Shen <i>et al.</i> [58]	CT	-	Lung cancer malignancy probability estimation.	DLR	1010	Not performed
Ciampi <i>et al.</i> [59]	CT	-	Lung tumor classification as solid and non-solid.	DLR	1411	Not performed
Afshar <i>et al.</i> [37]	MRI	-	Brain tumor type classification.	DLR	233	One expert
Cha <i>et al.</i> [71]	CT	-	Bladder cancer treatment response prediction.	DLR	123	Automatic
Yu <i>et al.</i> [70]	CT	-	Kidney tumor type classification.	HCR	119	Multiple experts
Zhou <i>et al.</i> [69]	CT	Gene expression.	Liver cancer recurrence prediction.	HCR	215	One expert

care applications, including oncology, cardiology, and neurology. In cardiology, for instance, Radiomics is used in different

investigations, such as identifying the coronary plaques [33]. In neurology, it is widely applicable for detecting Alzheimer's disease [34] and Parkinson's disease [35]. However, among all the applications of the Radiomics, cancer-related topics have been the focus of interest. Below we briefly introduce and define different cancer-related applications in which Radiomics has been shown to be successful.

- 1) **Cancer diagnosis**, which refers to confirming the presence or absence of the cancer, is one of the most critical and sensitive decisions that has to be made as early as possible. However, most of the times cancers are diagnosed in late stages reducing the chance of receiving effective treatment, as there are typically few clinical symptoms in the early stages of cancer. Nevertheless, Radiomics has the potential to improve the accuracy of cancer early diagnosis.
- 2) **Tumor detection** refers to the identification of those lesions that are malignant, which is very important in order to guide targeted local treatment. For instance, Radiotherapy, the process of killing cancerous cells using ionizing radiation, can have much more efficient results if it is focused on the more "aggressive" areas of the tumor (usually the more heterogeneous areas are the more aggressive ones). Drug delivery, i.e., having an exact plan to deliver the drug to the target area, is another problem that requires precise information about the abnormality location.
- 3) **Tumor classification and attribute scoring**: Tumor classification refers to determining the type of the tumor. Typically, cancer is classified into the following main classes: (i) benign; (ii) primary malignant, and; (iii) metastatic malignant. Besides, tumors are associated with different attributes such as their border and sphericity. Analyzing these attributes contribute to a better understanding of the tumor's shape and behavior.
- 4) **Survival prediction**: The knowledge of the expected survival of a specific disease with or without a specific treatment is critical both for treating physicians and the patients. Physicians need to choose the best treatment plan for their patients and patients need to know their predicted survival time in order to make their own choices for the quality of their life. Radiomics can add significant information about patient's survival based on image properties and heterogeneity of the tumor and this has attracted a lot of attention recently.
- 5) **Malignancy prediction**: Tumors can be either malignant or benign based on several factors such as their ability to spread to other tissues. Benign tumors usually do not spread to other organs but may need surgical resection because occasionally they may grow in size. Pre-invasive lesions may be indolent for years, however, they may transform to aggressive malignant tumors and therefore need to be monitored closely or even be treated with lower dose of anti-cancer regimens. Malignant tumors are life threatening and may spread to distant organs, requiring more complicated treatments such as Chemotherapy. Prediction of tumor malignancy

likelihood with noninvasive methods such as Radiomics is, therefore, of paramount importance.

- 6) **Recurrence prediction**: Even the treated cancers have the potential to grow and reappear, which is referred to as "cancer recurrence". As the cancerous region is supposed to be removed or treated, there are not strong landmarks or evidences helping with predicting the recurrence. However, recently Radiomics is being employed to assist with such issue and has shown promising initial results.
- 7) **Cancer staging**: Cancers may be diagnosed in different stages, e.g., they may be in an early stage meaning that they are remaining in the tissue they have first appeared in, or they can be in an advanced stage, meaning that they are spread in other tissues. Knowing the stage of the tumor has significant impact on the choice of required treatment.

Based on the above categories, Tables II and III summarize different application domains of Radiomics introduced in various articles, along with their associated Radiomics method (HCR, DLR, or the combination of both). These tables also provide information on any complementary data source that has been utilized in combination with Radiomics.

III. STATE-OF-THE-ART IN HAND-CRAFTED RADIOMICS

In clinical oncology, tissue biopsy, which refers to the removal of a small focal part of the cancerous tissue (tumor), is considered as the state-of-the-art approach for diagnosing cancer. Although tissue biopsy has several diagnostic properties and is widely used for detecting and investigating cancerous cells, its reliability is limited by the fact that tumors are spatially and temporally heterogeneous, and as a result, biopsy cannot capture all the available information that is necessary for an inclusive decision. Besides, most of the biopsies are invasive which restricts the number of times this procedure can be performed, or sometimes biopsy is not an option due to the high risk for complication that pertains to specific patients.

Although biopsy remains the gold standard for cancer diagnosis, it can be combined with Radiomics, which is a non-invasive technique and can capture intra-tumoral heterogeneity. The resulting imaging-guided biopsy is a much less interventional procedure associated with fewer complications compared to surgical biopsy. In other words Radiomics can be used to facilitate biopsy by detecting more suspicious locations [3]. Furthermore, Radiomics can provide complementary information for diagnosis, and in case of "negative result of a biopsy" Radiomics prediction models may also provide additional information to the clinicians on whether re-biopsy is needed [78].

In this section, we focus on the state-of-the-art research on hand-crafted Radiomics (HCR). Studies on hand-crafted Radiomics features [1], [2], [4], typically, consist of the following key steps:

1. **Pre-processing**, introduced to reduce noise and artifacts from the original data and typically includes image smoothing and image enhancement techniques.
2. **Segmentation**, which is a critical step within the HCR workflow, as typically HCR features are extracted from

segmented sections and many tissues do not have distinct boundaries [3]. Although manual delineation of the gross tumor is the conventional (standard) clinical approach, it is time consuming and extensively sensitive to inter-server variability [2], resulting in a quest to develop advanced (semi) automated segmentation solutions of high accuracy that can also generate reproducible boundaries.

Automatic and semi-automatic segmentation techniques can be either conventional, meaning that pre-defined features are used to classify image pixels/voxels as tumorous or non-tumorous, or deep learning-based, referring to the use of a deep network to segment the image. Conventional techniques can, themselves, lie within three categories of intensity-based [82], model-based, and machine learning methods. In the former category, intensity is used as the main distinguishing feature of the pixels, while in the model-based approaches, the aim is to improve an initial contour, by optimizing an energy function. In machine learning methods, however, a set of features, including intensity and gradient, are extracted from the pixels. These features are then used as the inputs to a machine learning model, such as a Support Vector Machine (SVM), to classify the pixels. Nevertheless, conventional techniques are subject to several shortcomings. For instance, the intensity of the tumor can, sometimes, be similar to other tissues, and therefore, intensity can not be a good discriminator. Furthermore, the formulation of an energy function, in a model-based segmentation, may involve large number of parameters [86], which makes optimization of the energy function difficult and time-consuming. Deep learning methods, on the other hand, are capable of learning the features that can best distinguish tumorous and non-tumorous pixels, and can be trained in an end-to-end manner. Deep learning approaches, such as different variations of the U-Net [67], “LungNet” architecture [83], DenseNet [84], and hybrid dilated convolutions (HDC) [85] are currently used more often for medical image segmentation.

The most important metric for evaluating a segmentation method is to calculate its accuracy according to a ground truth. However, since ground truth is not always available for medical images, reproducibility metrics are often used to assess the performance of the segmentation algorithm [2], [3]. For instance, reference [2] has used a similarity metric based on the overlap of generated segments resulting in a better average for automatic methods compared with manual delineation.

3. **Feature extraction**, which is the main step in Radiomics workflow and will be discussed in details in Sub-section III-A.
4. **Feature reduction**, is another critical step in Radiomics as although a large number of quantitative features can be extracted from the available big image datasets, most of the features are highly correlated, irrelevant to the task at hand, and/or contribute to over-fitting of the model. To address these issues, Radiomics feature reduction

techniques are discussed in Sub-section III-B.

5. **Statistical analysis**, which refers to utilizing the extracted Radiomics features in a specific application as outlined in Section II. We will further elaborate on such Radiomics-based statistical analysis in Sub-section III-C.

In the reminder of this section, we focus on Steps 3-5 in Sub-sections III-A-III-C, respectively, starting by reviewing the key feature extraction methodologies recently used in HCR, followed by a review of the main feature reduction techniques and Radiomics-based statistical analytics.

A. Radiomics Feature Extraction

During the feature extraction step within Radiomics workflow, different types of features are extracted that can be generally classified into three main categories: (1) **First order (intensity-based and shape-based features)** [8]; (2) **Second order (texture-based features)** [8], and; (3) **Higher order features** [3]. Table IV provides a summary of different potential features. It is worth mentioning that HCR features are not limited to this list and can exceed hundreds of features (e.g., in Reference [5] 400 HCR features are initially extracted before going through a feature reduction process). Below, we further investigate the most commonly used categories of hand-crafted features:

1. Intensity-based Features: Intensity-based methods convert the multi-dimensional ROI into a single histogram (describing the distribution of pixel intensities), from which simple and basic features (e.g., energy, entropy, kurtosis, and skewness) are derived. Intensity features allow us to investigate properties of the histogram of tumor intensities such as sharpness, dispersion, and asymmetry. These features are, however, the most sensitive ones to image acquisition parameters such as slice thickness [25] (discussed in the Information Post I). Therefore, designing intensity-based features need special care and pre-processing. Among all intensity features, entropy and uniformity are the most commonly used ones in Radiomics [62]. Generally speaking, entropy measures the degree of randomness within the pixel intensities, and takes its maximum value when all the intensities occur with equal probabilities (complete randomness). Uniformity, on the other hand, estimates the consistency of pixel intensities, and takes its maximum value when all the pixels are of the same value.

Although intensity-based features are simple to calculate and have the potential to distinguish several tissues such as benign and malignant tumors [62], they suffer from some drawbacks. First, the selected number of bins can highly influence such features, as too small or too large bins can not resemble the underlying distribution correctly, and as such these features are not always reliable representatives. Besides, optimizing the number of histogram bins can also be problematic, because it leads to different number of bins for different ROIs, and makes it difficult to compare the results of various studies.

2. Shape-based Features: Shape-based features describe the geometry of the ROI and are useful in the sense that they

TABLE IV: Different categories of HCR features commonly used within the context of Radiomics.

Category	Description	Sub-category
First Order Radiomics	Concerned with the distribution of pixel intensities and use of elementary metrics to compute geometrical features.	
• Shape Features	Quantify the geometric shape of region or volume of interest [2]	Size of the Region of Interest (ROI); Sphericity; Compactness; Total volume; Surface area, Diameter, flatness and; Surface-to-volume ratio [2], [25].
• Intensity Features	Derived from a single histogram generated from the 2D region or the whole 3D volume [2].	Intensity Mean; Intensity Standard Deviation; Intensity Median; Minimum of Intensity; Maximum of Intensity; Mean of Positive Intensities; Uniformity; Kurtosis; Skewness; Entropy ; Normalized Entropy; Difference of Entropy; Sum of Entropy, and; Range [2], [25].
Second Order Radiomics (Texture Features)	Concerned with texture features and relations between pixels to model intra-tumor heterogeneity. Texture features are generated from different descriptive matrices [2].	
• Gray Level Co-occurrence (GLCM)	GLCM [25] is a matrix that presents the number of times that two intensity levels have occurred in two pixels with specific distance.	Contrast; Energy; Correlation; Homogeneity; Variance; Inverse Difference Moment; Sum of Average; Sum of Variance; Difference of Variance; Information Measure of Correlation; Autocorrelation; Dissimilarity; Cluster Shade; Cluster Prominence; Cluster Tendency, and; Maximum Probability.
• Gray Level Run-Length (GLRLM)	GLRLM [62] is a matrix that presents the length of consecutive pixels having the same intensity.	Short run emphasis; Long run emphasis; Gray Level Non-Uniformity; Run length non-uniformity; Run percentage; Low gray level run emphasis, and; High gray level run emphasis [2].
• Neighborhood Gray Tone Difference Matrix (NGTDM)	NGTDM [25] is concerned with the intensities of neighboring pixels instead of the pixel itself.	Coarseness; Contrast; Busyness; Complexity Texture Strength.
• Grey-Level Zone Length Matrix (GLZLM)	GLZLM [9] considers the size of homogeneous zones in every dimension.	Zone Percentage; Short-Zone Emphasis; Long-Zone Emphasis; Gray-Level Non-Uniformity for zone; Zone Length Non-Uniformity.
Higher Order Radiomics	Use of filters to extract patterns from images.	Wavelets ; Fourier features [25]; Minkowski functionals; Fractal Analysis [3], and; Laplacian of Gaussian (LoG) [9].

have high distinguishing ability for problems such as tumor malignancy and treatment response prediction [25]. Although radiologists commonly use shape features (also referred to as “Semantic Features” or “Morphological features”), the aim of Radiomics is to quantify them with computer assistance [3]. These features are extracted from either 2D or 3D structures of the tumor region to investigate different shape and size characteristics of the tumor.

Among different shape-based features, volume, surface, sphericity, compactness, diameter, and flatness are more commonly used in Radiomics. For instance, sphericity measures the degree of roundness of the volume or region of interest and it is specially useful for the prediction of tumor malignancy,

as benign tumors are most of the times more sphere compared to malignant ones. Compactness is itself defined based on sphericity and as such, these two need not to be calculated simultaneously, and one of them will be probably excluded by the feature selection methods, which are targeting feature redundancy.

3. Texture-based Features: Shape-based and intensity-based features fail to provide useful information regarding correlations between different pixels across a given image. In this regard, texture-based features are the most informative ones, specially for problems where tissue heterogeneity plays an important role, because texture-based features can catch

the spatial relationships between neighboring pixels [25]. In Radiomics, typically, texture-based features are extracted based on different descriptive matrices, among them gray level co-occurrence matrix (GLCM), gray level run length matrix (GLRLM), neighborhood gray tone difference matrix (NGTDM), and gray level zone length matrix (GLZLM) are the most commonly used ones [62], which are defined below:

- **The GLCM**, models the spatial distribution of pixels' intensities and can be calculated by considering the frequency of the occurrence of all pairs of intensity values. Features extracted from GLCM are the most commonly used textural features in Radiomics [62]. Each GLCM is associated with two predefined parameters θ and d , where $\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$, and d is any integer distance admissible within the image dimensions.
- **The GLRLM**, defines the number of adjacent pixels having the same intensity value, e.g., the (i, j) element of the $GLRLM_\theta$ matrix determines the number of times intensity value i has occurred with run length j , in direction θ .
- **The NGTDM**, which is based on visual characteristics of the image, is a vector whose k^{th} element is defined as the summation of differences between all pixels with intensity value k and the average intensity of their neighborhood (size of which is determined by the user).
- **The GLZLM**, which looks for zones in a matrix. A zone can be defined as the set of connected pixels/voxels sharing the same intensity. The $(i, j)^{\text{th}}$ element of the GLZLM corresponds to the number of zones with the intensity i , and the size j .

4. Higher Order Radiomics Features: Higher order features such as Wavelet and Fourier features capture imaging biomarkers in various frequencies [25]. Wavelet features are the mostly used higher order features in Radiomics. Wavelet coarse and fine coefficients represent texture and gradient features respectively, and is calculated by multiplying the image by a matrix including complex linear or radial "wavelet mother functions". Fourier features can also capture gradient information. Minkowski Functional (MF) is another common higher order feature extractor considering the patterns of pixels with intensities above a predefined threshold.

In brief, the MFs are computed by initially forming a binary version of the ROI through utilization of several thresholds within the minimum and maximum intensity limits. Although the number of utilized thresholds is a free parameter, for better results, it should be identified through a selection mechanism (typically empirical tests are used). Based on the binarized ROI, different MFs such as area and perimeter are computable as follows

$$\begin{aligned} MF_{\text{area}} &= n_s, \\ \text{and } MF_{\text{perimeter}} &= -4n_s + 2n_e, \end{aligned} \quad (1) \quad (2)$$

where n_s and n_e are the total number of white pixels (above the threshold) and edges, respectively. This completes our coverage of feature extraction methods used in Radiomics.

B. Radiomics Feature Reduction Techniques

Feature reduction is another critical step in Radiomics as although a large number of quantitative features can be extracted from the available image datasets, most of the features are highly correlated, irrelevant to the task at hand, and/or contribute to over-fitting of the model (making it highly sensitive to noise). Feature reduction techniques that are used in Radiomics can be classified into supervised and unsupervised categories [8], as summarized in Table V. Supervised approaches, such as filtering and wrapper methods, take the discriminative ability of features into account and favor features that can best distinguish data based on a pre-defined class. Unsupervised methods, on the other hand, aim to reduce feature redundancy and include Principle Component Analysis (PCA), Independent Component Analysis (ICA) and Zero Variance (ZV) [8].

In summary, various objectives can be defined when reducing the feature space in Radiomics. The following key characteristics can be defined for feature selection purposes [2], [3]:

- **Reproducibility:** Reproducible features (also referred to as "stable features") are the ones that are more robust to pre-processing and manual annotations. These features will be discussed in Sub-section III-D.
- **Informativeness and Relevancy,** which can be defined as features that are highly associated with the target variable [25]. For instance a χ^2 -test, calculates the chi-squared statistic between features and the class variable, and consequently features with low impact on the target are discarded. Another selection approach is a Fisher score test, where features with higher variance are treated as the more informative ones.
- **Redundancy:** Non-redundant features are the ones with small correlation with each other. Feature redundancy is defined as the amount of redundancy present in a particular feature with respect to the set of already selected features.

Below, supervised and unsupervised techniques commonly used in Radiomics are further discussed.

1. Supervised Feature Selection Methodologies: Supervised methods are generally divided into two categories as outlined below:

- **Filtering (Univariate) Methods:** These methods consider the relation between the features and the class label one at a time without considering their redundancy. Among all filtering approaches, Wilcoxon test based method has been shown to be more stable, resulting in more promising predictions in the field of Radiomics [62]. A Wilcoxon test is a nonparametric statistical hypotheses testing technique that is used to determine dependencies of two different feature sets, i.e., whether or not they have the same probability distribution.
- **Wrapper (Multivariate) Methods:** Filtering methods have the drawback of ignoring relations between features which has led to development of wrapper techniques. In contrary to the filtering methods, wrapper methods investigate the combined predictive performance of a subset

TABLE V: Feature reduction techniques commonly used within the Radiomics literature.

Category	Description	Methods
Supervised	Considers the relation of features with the class labels and features are selected mostly based on their contribution to distinguish classes.	
• Filtering (Univariate)	Test the relation between the features and the class label one by one.	Fisher score (FSCR); Wilcoxon rank sum test; Gini index (GINI); Mutual information feature selection (MIFS); Minimum redundancy maximum relevance (MRMR), and; Student <i>t</i> -test [62].
• Wrapper (Multivariate)	Considers both relevancy and redundancy.	Greedy forward selection, and Greedy backward elimination.
Unsupervised	Does not consider the class labels and its objective is to remove redundant features.	
• Linear	Features have linear correlations.	Principle Component Analysis (PCA), and; Multidimensional scaling (MDS)
• Nonlinear	Features are not assumed to be lied on a linear space.	Isometric mapping (Isomap), and; Locally linear embedding (LLE).

of features, and the scoring is a weighted sum of both relevancy and redundancy [62]. However, computational difficulties prevent such methods from testing all the possible feature subsets.

Wrappers methods include greedy forward selection and greedy backward elimination. In a forward feature reduction path, selection begins with an empty set and the correlation with class label is calculated for all features individually. Consequently, the most correlated feature is selected and added to the set. In the next step, the remaining features are added, one by one, to this set to test the performance of the obtained set, and the process continues until no further addition can increase the predictive performance of the set. A backward selection path works in contrary to the forward one, beginning with a set including all the available features, and gradually reduces them until no further reduction improves the performance.

Since supervised methods are based on class labels, they are subject to over-fitting and can not be easily applied to different applications once trained based on a given feature set.

2. Unsupervised Feature Selection Methodologies: Unsupervised approaches try to reduce the feature space dimensionality by removing redundant features (those who are correlated and do not provide any additional information). Although these methods are not prone to over-fitting, they are not guaranteed to result in the optimum feature space. Unsupervised techniques can be divided into linear and non-linear methods, where the former assumes that features lie on a linear space. Because, in the field of Radiomics, commonly, very simple and basic forms of unsupervised techniques, such as PCA, are used, they are not covered in this article. However, this presents an opportunistic venue for application of more advanced statistical-based dimensionality reduction solutions recently developed within signal processing literature.

C. Radiomics Statistical Analysis

Statistical analysis refers to utilizing the extracted Radiomics features in a specific task such as cancer diagnosis, tumor stage classification, and survivability analysis, as described in Section II. Although most statistical methods, initially, treat all the features equally and use the same weights over all predictors, in the area of Radiomics, the most successful methods are the ones that use a prior assumption (provided by experts) over the meaning of features [3]. One basic approach to analyze the Radiomics features adopted in [5], [9] is to cluster the extracted features and look for associations among clusters and clinical outcomes. For instance, patients belonging to one cluster may have similar diagnosis or patterns. Observations show that image biomarkers are associated with clinical outcomes such as tumor malignancy. Hierarchical clustering is most commonly used in Radiomics [2]. However, clustering techniques are not basically trained for target forecasting purposes, which necessitates the use of prediction tools that are specially trained based on a predefined class label. Prediction tools in Radiomics are categorized as either:

- (i) **Classification and Regression Models** that are mostly similar to other multi-media domains, trying to foresee a discrete or continues value. Random Forest (RF), Support Vector Machine (SVM) and Neural Network (NN) are among the most common regression and classification techniques used to make predictions based on Radiomics [8].
- (ii) **Survivability analysis:** Also referred to as time-related models, mostly try to predict the survival time associated with patients. These models are also useful when testing the effectiveness of a new treatment.

Table VI presents a summary of different Radiomics analysis techniques. As predictors belonging to the former category

TABLE VI: Common analysis methods in Radiomics.

Purpose	Description	Methods
Clustering	Similar patients are grouped together based on a distance metrics.	Hierarchical, Partitional
Classification	Models are trained to distinguish patients based on their associated clinical outcome.	Random Forest (RF); Support Vector Machine (SVM); Neural Network (NN); Generalized linear model (GLM); Naive Bayes (NB); k-nearest neighbor (KNN); Mixture Discriminant Analysis (MDA); Partial Least Squares GLM (PLS), and; Decision Tree (DT).
Time-related analysis	The survival time or the probability of survival is calculated based on the available set of data from previous patients.	Kaplan-Meier survival analysis; Cox proportional hazards regression model [4], and; Log-Rank Test.

are also common in other multi-media applications, they are not covered in this article. Survivability analysis (the latter category), however, is more specific to Radiomics, and as such, below, we discuss the three mostly used techniques from this category, i.e., **Kaplan-Meier Survival Curve (KMS)**, **Cox Proportional Hazards (regression) Model (PHM)**, and **Log-Rank Test**.

1. Kaplan-Meier Survival Curve (KMS): The KMS curve [5], [8] represents a trajectory for measuring the probability of survival $S(t)$ in given points of time t , i.e.,

$$S(t) = \frac{\text{Number of patients survived until } t}{\text{Number of patients at the beginning}}. \quad (3)$$

The KMS curve can be calculated for all Radiomics features to assess the impact of different features on patients' survival as follows:

- 1) A desired feature, for which the KMS curve is supposed to be calculated, is selected.
- 2) Based on the selected feature, one or more thresholds are considered that can partition patients into, e.g., low and high risk cancer subjects. Patients are then grouped based on whether their associated feature lies above or below the threshold.
- 3) The KMS curve is calculated for all the obtained groups, and the result can be used to compare the survivability among patients with, e.g., low and high risk cancer. For instance, in Reference [5] *high heterogeneity features* are associated with shorter survival time, while *high compactness features* are associated with longer survival.

2. Cox Proportional Hazards (Regression) Model (PHM) [5], is commonly used in medical areas to predict patient's survival time based on one or more predictors (referred to as covariates) such as Radiomics features. The output of the PHM model denoted by $h(t)$ is the risk of dying at a particular time t , which can be calculated as follows

$$h(t) = h_0(t) \times \exp \sum_{i=1}^{N_c} b_i x_i \quad (4)$$

where x_i , for $(1 \leq i \leq N_c)$, are predictors (covariates); b_i represent the impacts of predictors, and $h_0(t)$ is called the base-line hazard. The exponent term in Eq. (4) is referred to as

the "Risk" and is conventionally assumed to be a linear combination of the features (covariates), i.e., $\text{Risk} \triangleq \sum_{i=1}^{N_c} b_i x_i$. The Risk coefficients (b_i , for $(1 \leq i \leq N_c)$) are then computed through a training process based on historical data. More realistically, the risk can be modeled as a general non-linear function, i.e., $\text{Risk} \triangleq f(x)$, with the non-linearity being learned via deep learning architectures, which has not yet been investigated within the Radiomics context.

3. Log-Rank Test [5], which is used for comparing the survival of two samples specially when these two samples have undergone different treatments. This test is a non-parametric hypothesis test assessing whether two survival curves vary significantly. One limitation associated with the Log-Rank test is that the size of the groups can influence the results, therefore, larger number of patients should be included to form equal sized groups.

Evaluation of HCR: In summary, having a successful hand-crafted Radiomics pipeline requires an accurate design to choose the best combination of feature extraction, feature reduction, and analysis methods. Several studies have tried to find the most important factor leading to the performance variation. For instance, the effects of these design choices are recently investigated in [8] through an analysis of variance, where it has been shown that, e.g., **feature selection can significantly influence the final accuracy**. On the other hand, References [17], [74] have concluded that the classification method is the dominant source of performance variation for head-and-neck and lung cancer classification tasks. The difference between the concluding remarks of Reference [8] and References [17], [74] indicates that **impact of design choices may vary from one application to another, one dataset to another, and from a set of features to another**.

Finally, it should be noted that **reporting accuracy is not as informative measure in Radiomics as it is in other multi-media domains**. Because, in medical areas, making mistakes in, for instance, classifying positive and negative samples are not equal. **Therefore, in Radiomics, measures which are capable of distinguishing between False Positive (FP) and False Negative (FN) errors are more favored**. Such one measure is the **area under Receiver Operating Characteristic (ROC) curve**, which allows for investigating the impact of different decision thresh-

olds on FP and FN rates. Confusion matrix is another common and useful technique to report the performance of a Radiomics classifier in terms of its FP and FN rates. In practice, most of the decisions in medical areas cannot be made with a complete certainty, and physicians consider several factors such as harms and benefits of a specific judgment, when forming thresholds for their decisions. However, these factors are not quantified and utilized in Radiomics, which calls for a broad investigation into potential solutions to incorporate the factors commonly used by physicians.

D. Radiomics Stability

An important aspect of Radiomics is the stability of the extracted features, which quantifies the degree of dependency between features and pre-processing steps. Stability in Radiomics is generally evaluated based on either of the following two techniques:

- 1) **Test-Retest:** In this approach, patients undergo an imaging exam more than once and images are collected separately. Radiomics features are then extracted from all the obtained sets and analyzed. Here, being invariant across different set of images illustrates stability of Radiomics features.
- 2) **Inter-observer reliability,** which is referred to an experiment where multiple observers are asked to delineate the ROI from the same images, and Radiomics features are extracted from all different delineations to test their stability for variation in segmentation [9]. Here, being invariant across different segmentations illustrates stability of Radiomics features.

Different stability Criteria are used to find robust features in Radiomics as briefly outlined below:

- 1) **Intra-class Correlation Coefficient (ICC):** One approach to measure the stability of Radiomics features, which is used for both the aforementioned categories (i.e., test-retest and inter-observer setting) is referred to as the intra-class correlation coefficient (ICC) [5]. The ICC is defined as a metric of the reliability of features, taking values between 0 and 1, where 0 means no reliability and 1 indicates complete reliability. Defining terms BMS and WMS as mean squares (measure of variance) between and within subjects, which are calculated based on a one-way Analysis of variance (ANOVA), for a test-retest setting, the ICC can be estimated as

$$ICC_{\text{Test-Retest}} = \frac{BMS - WMS}{BMS + (N - 1)WMS}, \quad (5)$$

where N is the number of repeated examinations. By defining EMS as residual mean squares from a two-way ANOVA and M as the number of observers, for an inter-observer setting, the ICC can be calculated as

$$ICC_{\text{Inter-Observer}} = \frac{BMS - EMS}{BMS + (M - 1)EMS}. \quad (6)$$

- 2) **Friedman Test:** The Friedman test, which is specially useful for assessing the stability in an inter-observer

setting, is a nonparametric repeated measurement that estimates whether there is a significant difference between the distribution of multiple observations, and has the advantage of not requiring a Gaussian population. Based on this test, the most stable features are the ones with a stability rank of 1 [5].

In [5], it is declared that Radiomics features with higher stability have more prognostic performance, therefore, stability analysis can be interpreted as a feature reduction technique. According to Reference [9], Laplacian of Gaussian (LoG), intensity-based, and texture features are more stable for lung CT images, while wavelet and shape-based features are sensitive to variation in segmentation. However, there are also other sources of variation (other than the segmentation step), that can influence the stability of the features, one of which is the image intensity discretization strategy [72] that has a strong impact on the texture features, in particular. There are two main approaches to discretize the medical images. The first one is to adopt a fixed bin size for all of the images, and the second one is to use a fixed number of bins. While it is shown that both approaches lead to texture features that depend on the intensity resolution, the first method (fixed-sized bins) results in more stable and comparable features. Nevertheless, texture analysis requires standardized intensity discretization method, to serve as a meaningful and reliable Radiomics technique.

Finally, it is worth mentioning that the Image Bio-marker Standardization Initiative (IBSI) [73] is an international collaboration that seeks to provide unique definitions, guidelines, and Radiomics steps, in response to the reproducibility challenge of the Radiomics. The provided guideline covers several steps within the Radiomics pipeline, from image acquisition, pre-processing and segmentation, to feature calculation. To summarize our finding on HCR and elaborate on its applications, we have provided an example in Information Post III-E, where the problem of Radiomics-based lung cancer analysis is investigated.

E. Radiogenomics

Radiomics is typically combined with Genomic data, often referred to as Radiogenomics [3]. In other words, Radiogenomics refers to the relationship between the imaging characteristics of a disease and its gene expression patterns, gene mutations and other genome-related characteristics. Potential association between imaging outcome and molecular diagnostic data can serve as a predictor of patient's response to the treatments and provide critical support for the decision making tasks within the clinical care setting. In other words, Radiogenomics has the potential to investigate the Genomics of the cancer, without utilizing invasive procedures such as biopsy. Various association mining and clustering methods are used to identify the relationships between gene expressions and Radiomics, e.g., in [3] it was found that just 28 Radiomics features were able to reconstruct 78% of the global gene expressions in human liver cancer cells.

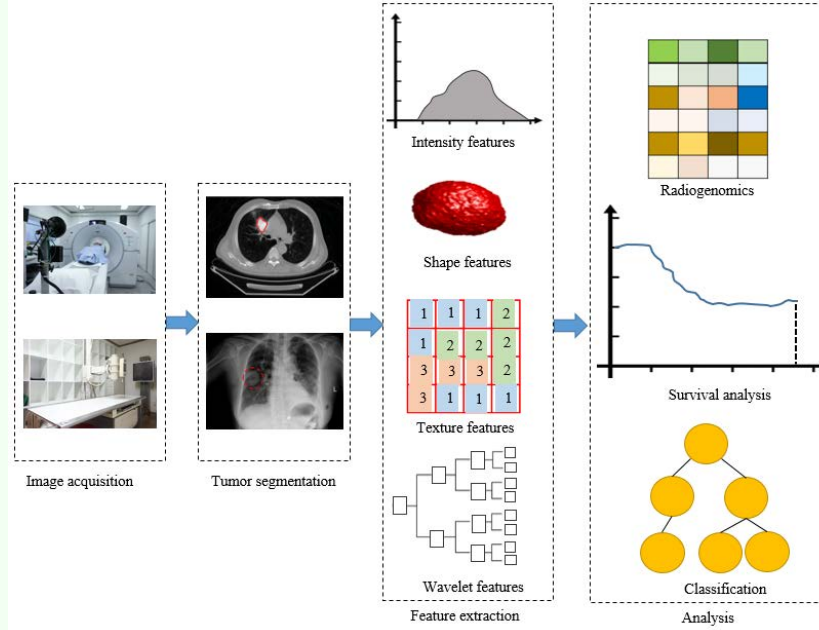
To assess the association between gene expression and discrete classes such as benign and malignant tumors, genes should be first sorted based on their discriminative ability.

However, the goal of Radiogenomics is to find associations between gene expression and Radiomics features, therefore, discriminative ability is indefinable. Spearman's Rank Correlation Coefficient (SRC) [38] can be used to measure the correlation between a specific Radiomics feature and gene expression. Genes are then sorted based on their SRC coefficient instead of their discriminative ability. The ordered genes are typically stored in a list L , and to extract meaning from this list, the traditional approach is to focus on the top and bottom genes in list L , representing genes with the strongest positive and negative correlations, respectively. This approach, however, is subject to several limitations, such as the difficulty in biological interpretation, which has led to the introduction of Gene-set Enrichment Analysis (GSEA) [39]. The GSEA is one of the mostly used Radiogenomics approaches [5], and is based on the definition of gene sets. Each gene set S is a group of genes that are similar in terms of the prior biological knowledge, such as involvement in common

biological pathways. The goal of the GSEA is to find out whether the members of a given set S tend to occur in the top or bottom of the list L . In this case, the expression of this gene set is associated with the specific Radiomics feature. The result of Radiogenomics analysis using GSEA is a heat map representing the degree of association between all gene sets and Radiomics features as shown in Fig. 2.

The main role of Radiogenomics is that it allows imperfect data sets, where clinical outcomes are difficult to be collected or require an extended period to be collected, to be leveraged based on prior knowledge of the relationship between clinical outcomes and genomics, in order to draw new conclusions. For example in a research project there may have been imaging data and genome-related data but no clinical outcomes. If there is prior knowledge of the relationship of the genomics with certain clinical outcomes, then by correlating the imaging data with the genomics, new conclusion can be drawn about the relationship of the imaging data with the clinical outcomes. In this scenario genomics can fill the gaps in knowledge.

Information Post II: Radiomics for Lung Cancer Analysis



Studies have shown that use of computer aided systems can significantly help with the early diagnosis and detection of lung cancer and consequently lead to dramatic decrease in lung cancer-related mortality [64]. For these computer aided systems to be successful in lung cancer early prognosis, distinguishing Radiomics features have to be extracted from the CT images, which are the main detectors of lung cancer. Radiomics can help with different tasks related to lung cancer such as: (i) Lung cancer patient survival prediction; (ii) Lung cancer malignancy prediction; (iii) Forecasting the patient's response to treatment; and (iv) Prediction of the stage of the cancer.

As illustrated in the figure above, the first stage in lung cancer Radiomics pipeline is commonly the segmentation of left and right lungs, followed by the **segmentation of the tumor** which can be performed automatically for tumors with high intensities located in low intensity backgrounds. However, segmentation can become problematic when lung tumors are attached to vessels or walls which makes automatic segmentation fail to generate reproducible and accurate outcomes. Due to this reason lung tumor segmentation is still an open problem [2].

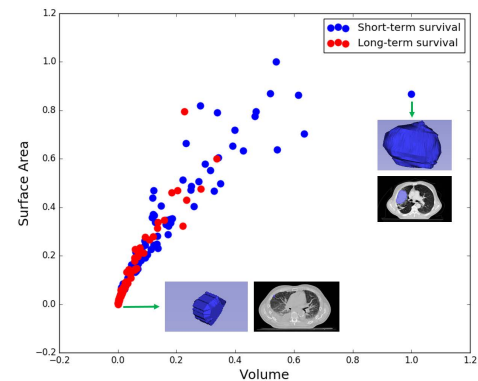
The next step after segmenting the tumor, is to **extract Radiomics features**. Types of features to extract depend on the problem at hand. For instance, according to [25], intensity features are highly correlated to lung cancer patient survival, while shape features are beneficial for lung cancer malignancy prediction and also forecasting the patient's response to treatment. More importantly, texture features are reported to be the most influential ones on lung cancer outcomes [3]. Selected features will then go through a **feature selection** process, where redundant and non-relevant features are excluded. Feature selection is followed by a **statistical analysis** step to perform the aforementioned tasks.

To further elaborate on the use of Radiomics in lung cancer analysis, we have implemented the Hand-Crafted pipeline on 157 patients, from NSCLC-Radiomics dataset [5]. We grouped these patients into two categories: Long-term survival, referring to patients who have survived more than 500 days, and Short term survival, referring to those who have survived less than 500 days. The following techniques (from possible options described in Section III) are employed in the proposed pipeline:

- 1) **Segmentation:** Manual annotations performed by an expert are utilized.
- 2) **Radiomics Feature Extraction:** 11 first-order (Number of pixels, ROI size in mm^2 , Mean gray level, Stan-

dard Deviation, Median gray level, Min ROI, Max ROI, Mean Positive Values, Uniformity, Kurtosis, Skewness) and 19 second-order Radiomics features (Contrast, Energy, Correlation, Homogeneity, Entropy, Normalized Entropy, Variance, Inverse Difference Moment, Sum of Average, Sum of Variance, Sum of Entropy, Difference of Variance, Difference of Entropy, Information Measure of Correlation, Autocorrelation, Dissimilarity, Cluster Shade, Cluster Prominence, Maximum Probability) are extracted from lung tumors [8].

- 3) **Feature Reduction:** Based on a χ^2 feature selection test, among all the extracted features, the surface area and volume of the tumors are selected as the most correlated ones to the survival outcome.
- 4) **Analysis:** Following figure illustrates the relation between survival and the two aforementioned features. As it can be inferred from this figure, although patients with short-term survival can be associated with various tumor sizes, those who have survived longer have smaller tumors in terms of surface area and volume. In other words, being small in size seems to be necessary for long-term survival, however, as so many other factors influence survival, not all small tumors lead to long survivals. For instance, the location of the tumor can have noticeable impact on survival, nevertheless, it can not be assessed based on hand-crafted Radiomics features, as they are extracted from the segmented ROI, without taking its location into account.



IV. STATE-OF-THE-ART IN DEEP LEARNING-BASED RADIOMICS

Deep learning-based Radiomics (DLR), sometimes referred to as “Discovery Radiomics” or “Radiomics Sequence Discovery” with “Sequence” referring to features [66], is the process of extracting deep features from medical images based on the specifications of a pre-defined task including but not limited to disease diagnostics; Cancer type prediction, and; Survival Prediction. In brief, the DLR can be extracted via different architectures (stack of linear and non-linear functions), e.g., Convolutional Neural Network (CNN) or an Auto-Encoder, to find the most relevant features from the input [64]. Fig. 3 illustrates the schematic of extracting deep features. The extracted features can then either go through the rest of the deep net for analysis and making decisions or they may exit the network and go through a different analyzer such as an SVM or a Decision Tree (DT). Commonly used deep architectures for Radiomics will be discussed in details later in Section IV-C. **Benefits of DLR vs. HCR:** An important advantage of DLR over its hand-crafted counterpart is that the former does not need any prior knowledge and features can be extracted in a completely automatic fashion with high level features extracted from low level ones [64]. Moreover, deep learning networks can be trained in a simple end-to-end process, and their performance can be improved systematically as they are fed with more training samples [60]. Another key benefit of using DLR instead of HCR is that the input to the deep networks to extract Radiomics features, can be the raw image without segmenting the region of interest, which serves the process in two ways:

- (i) Eliminating the segmentation step can significantly reduce the computational time and cost by taking the burden of manual delineation off the experts and radiologists, besides, manual annotations are highly observer-dependent, which makes them unreliable sources of information, and;
- (ii) Automatic segmentation methods are still highly error prone and inaccurate to be used in a sensitive decision making process.

Furthermore, the input to a deep network can also be the combination of the original and segmented image along with any other pre-processed input such as the gradient image (referred to as “multi-channel” input), all concatenated along the third dimension [10]. The variety of input types can even go further to include images from different angles such as coronal and axial [59].

Generally speaking, studies on DLR can be categorized from several aspects including:

- (i) **Input Hierarchy:** The input to the deep net can be the single slices, the whole volume, or even the whole examinations associated with a specific patient. Each of these cases require their own strategy, e.g., in case of processing the whole volume simultaneously, one should think of a way to deal with the inconsistent dimension size, as patients are associated with different number of slices. One common architecture that allows for utilization of inputs with variable sizes, such as various number

of slices, is the Recurrent Neural Network (RNN), which will be briefly discussed in Section IV-C;

- (ii) **Pre-trained and Raw Models:** Depending on the size of the available dataset and also the allocatable time, pre-trained models can be fine-tuned or raw models can be trained from scratch. This will be analyzed more specifically in Section IV-B, and;
- (iii) **Deep Learning Network Architectures:** Choice of the deep network is the most important decision one should make to extract meaningful and practical DLR, which will be discussed in Section IV-C.

In the reminder of this section, we will review the state-of-the-art in deep Radiomics from different perspectives such as input hierarchy, pre-trained vs. raw models, and deep learning network architectures. Fig. 4 illustrates a taxonomy of different DLR approaches providing a guide to the rest of this section.

A. Input Hierarchy

As shown in Fig. 5, input images for DLR studies can be divided into three main categories: Slice-level; Volume-level, and; Patient-level. Slice-level classification refers to analyzing and classifying image slices independent from each other, however, this approach is not informative enough as we typically need to make decisions based on the labels assigned to the entire Volume of Interest (VOI). Shortcomings of slice-level classification leads to another approach referred to as volume-level classification, where either the slice-level outputs are fused through a voting system, or the entire image slices associated with a volume is used as the input to the classifier. Finally, patient-level classification refers to assigning a label to a patient based on a series of studies (such as CT imaging follow-ups). For example, in Reference [58], patient-level classification is explored with the goal of estimating the probability of lung tumor malignancy based on a set of CT studies. To achieve this goal, initially, a simple three layer CNN is trained to extract DLR from tumor patches associated with individual CT series (volume-level classification) with the objective of minimizing the difference between the predicted malignancy rate and the actual rate. Then, by adopting a previously trained CNN, the malignancy rate is calculated for all the series belonging to the patient and the final decision is made by selecting the maximum malignancy rate. In other words, a patient is diagnosed with malignant lung cancer if at least one of the predicted rates is above a pre-determined rate for malignancy.

B. Pre-trained or Raw Models

Similar to the other medical areas, the DLR can be extracted based on either of the following two approaches:

Training from scratch: Training a deep network from scratch for extracting DLR has the advantage of having a network completely adjusted to the specific problem at hand. However, performance of training from scratch could be limited due to couple of key issues, i.e., over-fitting and class imbalance. Adhering to patients’ privacy and need for experts to provide ground truth typically limits the amount of medical datasets

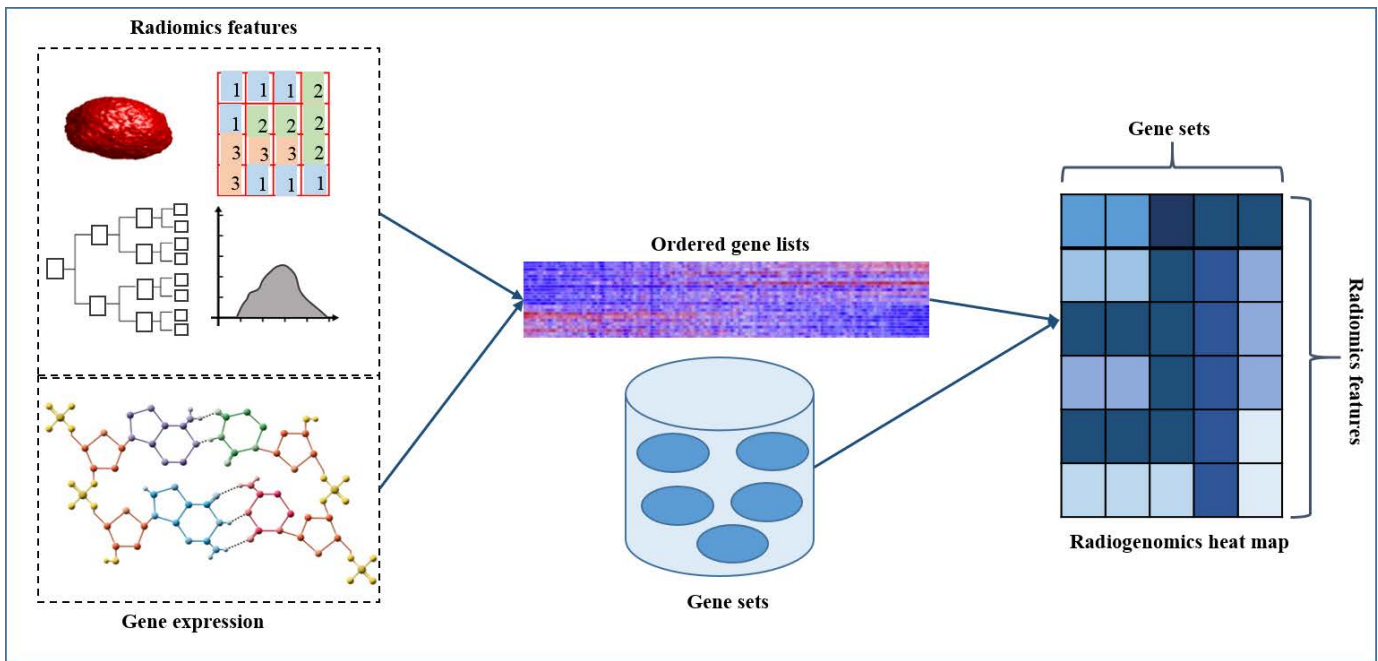


Fig. 2: **Radiogenomics analysis using GSEA**. Genes are first sorted based on their association with Radiomics features, and the final heat map is formed according to the pattern of gene sets within the ordered gene list.

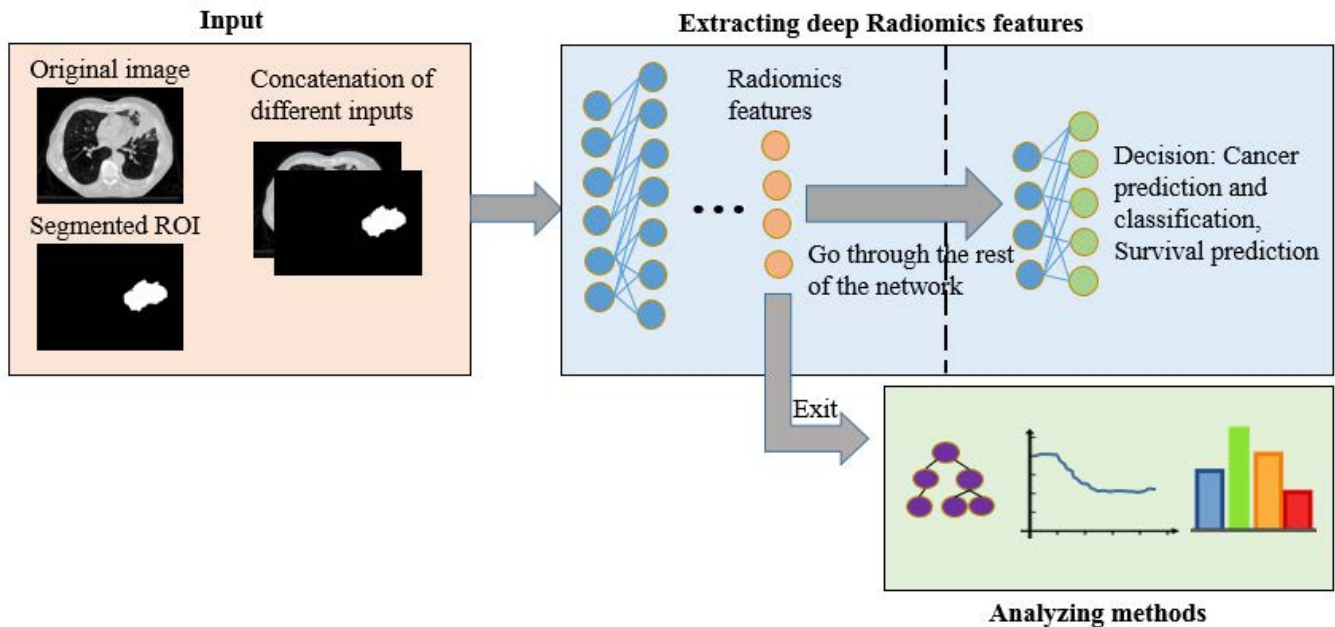


Fig. 3: Extracting deep Radiomics. The input to the network can be the original image, the segmented ROI, or the combination of both. Extracted Radiomics features are either utilized through the rest of the network, or an external model is used to make the decision based on Radiomics.

available for extracting DLR resulting in over-fitting of the deep nets. The second issue is the problem of class imbalance, i.e., unequal number of positive and negative classes. This happens as number of patients diagnosed with abnormalities is commonly less than the amount of data available from healthy subjects. More specifically, class imbalance in medical areas is due to the fact that typically number of positive labels is less

than the number of negative ones, making the classifier biased toward the negative class, which is more harmful than the other way around because, for instance, classifying a cancerous patient (positive label) as healthy (negative label) has worse consequences than classifying a healthy patient as cancerous [11]. The following strategies can be adopted to address these two issues:

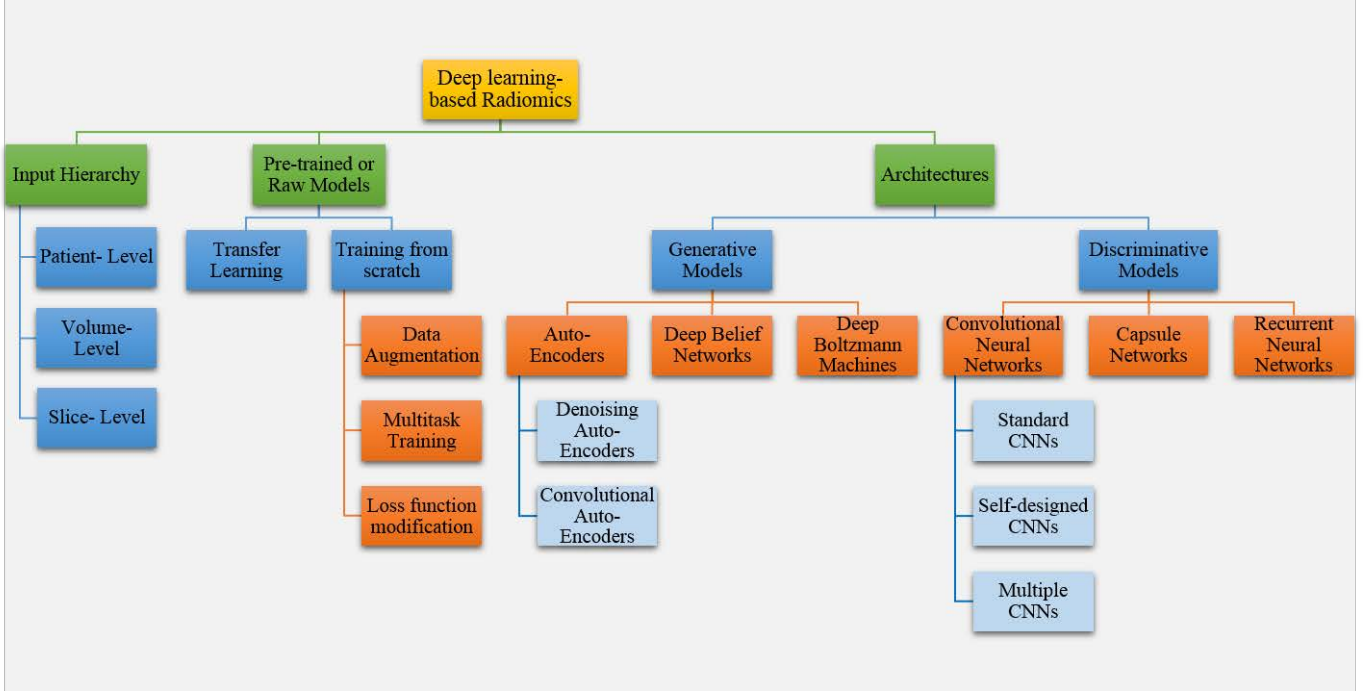


Fig. 4: Taxonomy of Deep Learning-based Radiomics (DLR).

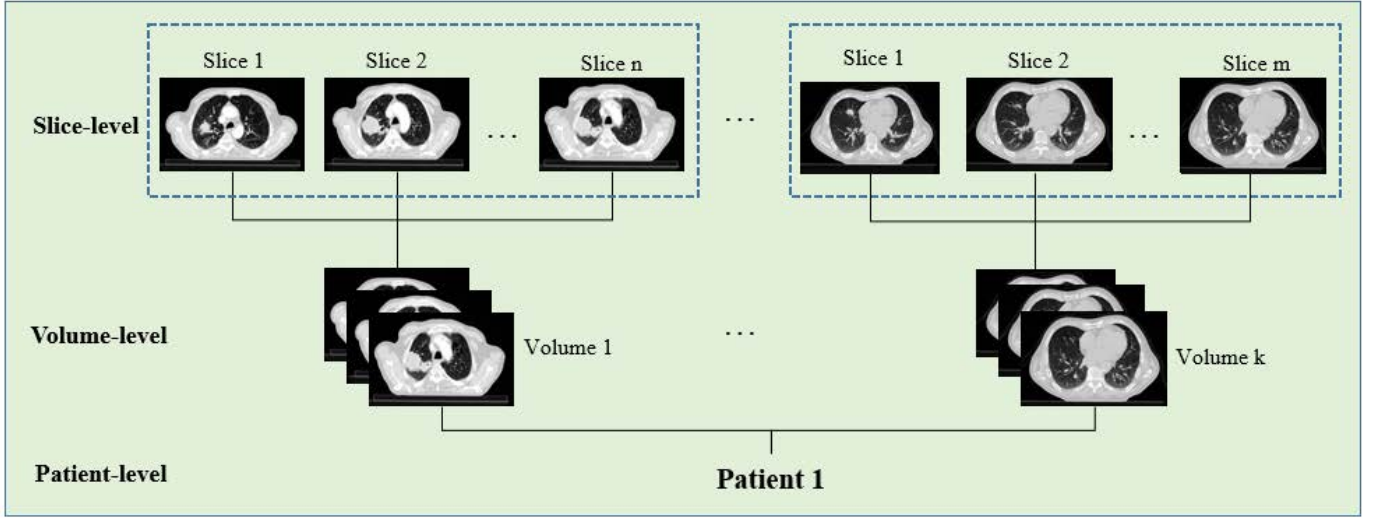


Fig. 5: Input hierarchy for one patient. In the top row the slice-level input is shown where the patients went through K examination visits during each of which $N^{(i)}$, for $(1 \leq i \leq K)$, number of slices is captured. The second row shows the volume-level where all slices associated with one visit is provided simultaneously as the input to the network. Finally, the third row shows the Patient-level analysis, where a single input consisting of all the volumes is provided.

- (i) **Data Augmentation**, where different spatial deformations (such as rotation [41]) are applied to the existing data in order to generate new samples for training purposes. Sub-patch Expansion [10] is another form of augmentation commonly adopted in Radiomics to handle the inadequate data situation via extracting several random fixed-sized sub-patches from the original images.
- (iii) **Multitask training** is another method introduced to handle class imbalance and inadequate data [20], which is achieved by decreasing the number of free parameters and consequently the risk of over-fitting. For instance,

this approach is adopted in [44] for spinal abnormality classification based on MRIs through training a multitask CNN. Multitask in this context refers to performing different classification tasks simultaneously via the same unified network (e.g., the network tries to classify disk grading and disk narrowing at the same time). The loss function is defined as the weighted summation of all the losses associated with different tasks. One important decision to make in multitask learning is the point that branching begins, e.g., in Reference [44], a unified CNN is trained, where all Convolutional layers are shared for

performing different tasks and tasks are separated from the point that fully connected layers begin.

- (iv) **Loss function modification:** Another common approach specific to handling class imbalance for DLR extraction is to modify the loss functions by giving more weight to the minority class [44].

Transfer Learning via a Pre-trained Network: A different solution to class imbalance and inadequate training data is “transfer learning” followed by “fine tuning” [10], [20], [42]. The transfer learning phase refers to training the deep net using a natural image data set, and then in the fine tuning phase, the trained network will be re-trained using the desired medical dataset. This strategy is adopted in Reference [42], where a pre-trained CNN is used for breast cancer classification based on mammographic images. The pre-trained CNN used is an Alexnet which is too complicated and prone to over-fitting for small datasets. Therefore, this network is first pre-trained using ImageNet database which consists of more than one million natural images. Pre-trained CNN based on ImageNet is also adopted in [46] for lung cancer survival prediction.

C. Deep Learning Architectures in Radiomics

Radiomics features can be extracted through both discriminative and/or generative deep learning networks. As is evident from its name, discriminative deep models try to extract features that make the classes (e.g., normal or cancerous) distinguishable, and thus these models can directly classify instances from the extracted features. On the other hand, generative models are unsupervised, meaning that they are trained without considering the class labels. Generally, the goal of these models is to learn the data distribution in a way that enables them to generate new data from the same distribution. In other words, generative models can extract the natural and representative features of the data, which can then be used as inputs to a classifier. Furthermore, in the field of Radiomics, it is common [60] to train a generative model and use the learned weights as initial weights of a discriminative model. Below, an introduction to widely used discriminative and generative deep models in Radiomics is provided.

1. Discriminative Models: Deep discriminative models try to extract features capable of distinguishing class labels, and the objective is to minimize the prediction error. Below, we will review the **Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs)**, which are the most popular discriminative architectures in Radiomics. Later, we will introduce a recently designed deep architecture referred to as the **Capsule network (CapsNet)** [36] and explain how this new architecture can contribute to the Radiomics.

1.1. Convolutional Neural Networks (CNNs): CNN is a stack of layers performing Convolutional filtering combined with nonlinear activation functions and pooling layers [20]. The fact that CNNs have recently resulted in promising outcomes have made them the mostly used architecture in medical areas including Radiomics. CNNs are more practical in the sense that shared weights are utilized over the entire input, which

reduces the number of trainable parameters. Unlike extracting hand-crafted features, kernels used in Convolutional layers are not pre-determined and are automatically learned through the training process. This property makes CNNs suitable methods for extracting DLR features as they are flexible and can be applied without requiring a prior knowledge. In [55], it has been shown that the DLR extracted from a CNN can visually distinguish benign and malignant lung tumors when projected into a 2D space, while the original pixel values completely fail to provide such distinction.

When adopting CNNs in the field of Radiomics, output of the fully connected layers is typically treated as DLR features. These features are then either used within the original CNN to provide the desired (classification and/or regression) output such as cancer type, or exist the network to be provided as the input to the rest of the Radiomics pipeline. As an example, in Reference [46] DLR are extracted from the layer just before the classification (SoftMax) layer of the CNN with the goal of lung cancer survival prediction. These features are referred to as the “preReLU” and “postReLU” features as they are extracted both before and after applying the ReLU activation function. The DLR features are then used as inputs to four classifiers (i.e., Naive Bayes, Nearest Neighbor, Decision tree and Random Forest) after going through a feature selection algorithm.

The CNN architectures used in Radiomics can be divided into three main categories: (i) Standard architectures; (ii) Self-designed architectures, and; (iii) Multiple CNNs. Below, we describe each of these categories with examples from Radiomics:

1.1.1. Standard CNN Architectures: As the name suggests, standard architectures are those that have been previously designed to solve a specific problem, and due to their success are now being adopted in the Radiomics. Two of such architectures that have been used in Radiomics are LeNet and AlexNet. The LeNet is one of the simplest CNN architectures, having a total of 7 layers, that has been used in Radiomics. However, researchers have some times modified this network to achieve higher performance, e.g., the CNN used in Reference [10] is a LeNet architecture with a total of 9 layers including 3 Convolutional layers, 3 pooling layers and one fully connected layer followed by the classification layer to classify lung tumors as either benign or malignant.

Another commonly used standard architecture in Radiomics is the 11 layers CNN called Alexnet, which has been adopted in [42] to extract DLR features from breast mammographic images. Features are extracted from all 11 layers and used as inputs to 11 support vector machine (SVM) with the goal of classifying breast tumors as either benign or malignant. Since it is not obvious which set (output of which of the 11 underlying layers) of DLR features are more practical, these SVMs are compared and the one with the largest area under the curve is chosen for predictive analysis of breast cancer. The results of [42] concluded that the features extracted from the 9th layer (a fully connected layer before the last fully connected

layer and the classification layer) are the best predictors of breast cancer and they are of lower dimension compared to previous ones, which reduces the computational cost. In other words and in contrary to [46], the output of the last Convolutional layer, right before the fully connected layer, is selected as the DLR features.

Although AlexNet is a powerful network, it has too many parameters for small datasets and is, therefore, prone to over-fitting. As a result, Reference [15] has used an Alexnet with number of layers reduced to 5 in order to avoid the over-fitting problem. The input to this network is a combination of CT and PET images, each having 3 channels: One slice corresponding to the center of the lung nodule, specified by an expert, and the two immediate neighbors. The goal of this article is to classify lung tumors as benign or malignant, and although it has been shown that the adopted CNN does not result in significantly higher accuracy than classical methods (HCR), it is more convenient as it does not require the segmented ROI.

Inception network [75], [77] is another CNN adopted in Radiomics. This network involves parallel convolutions with different kernel sizes, and poolings within the same layer, with the overall aim of allowing the network to learn the best weights and select the most useful features. The Inception CNN is used in [76], for the detection of diabetic retinopathy. This paper is the first work on deep learning-based detection of diabetic retinopathy that has been approved by the Food and Drug Administration (FDA).

- 1.1.2. **Self-designed CNNs:** As opposed to researchers that have used standard CNNs with or without modifications, some have designed their own architectures based on the specification of the Radiomics problem at hand. For example, Reference [41] has used a CNN with three Convolutional layers to extract DLR features, and although the CNN itself is trained to use these features for classifying benign and malignant tumors, they are used as inputs to a binary decision tree.

In a similar fashion, Reference [43] has used a CNN with 6 Convolutional layers and one fully connected layer for DLR extraction in the problem of brain tumor classification. The designed network, however, is different from previously mentioned articles as it is developed for tumor segmentation, and *features are extracted from the last Convolutional layer since they are more robust to shifting and scaling of the input*. In other words, the CNN was designed for segmentation and once trained, the output of the last Convolutional layer is used as the DLR features. The claim here is that the quality of extracted features depends on the accuracy of segmentation, and when segmentation is precise the quality of Radiomics features is guaranteed. Due to the high importance of the segmentation, more advanced and efficient CNN architectures have been developed, one of which is the Fully Convolutional Neural Network (FCNN) [23]. In

an FCNN, fully connected layers are rewritten as convolutional layers, having the advantage of not requiring fixed-sized inputs. This network is also extended to 3D image segmentation, to segment multiple targets at once. To decrease the false positive rate, FCNN is further combined with graphical models such as Markov Random Fields (MRFs) and Conditional Random Fields (CRFs). Finally, to improve the resolution of the output, U-Nets [67] are proposed, which include up-convolutions to increase the image size, and skip-connections to recover spatial information.

Lung cancer detection using CNNs is also investigated in [47], with the difference that the input to the network is not only the original image but also the nodule-enhanced and vessel-enhanced images, stating that providing the network with more information on tumor and vessels reduces the risk of misplacing these two by the network. The main focus here is to reduce the false positive rate while keeping the sensitivity high, therefore, a significant number of nodule candidates are selected at the beginning. Use of CNNs is further investigated in [45], where a 7 layer architecture is fed with down-sampled volumetric CT images along with their segmentation masks for longevity prediction. In [48] an architecture called XmasNet is provided that can maximize the accuracy of prostate cancer diagnosis. This network consists of 4 Convolutional layers, 2 fully connected layers, 2 pooling layers and one SoftMax layer for cancer prediction. The inputs to this network are 3D MRI images.

In summary, self-designed CNNs are developed by varying the depth of the network (number of the Convolutional and non-Convolutional layers); the order the layers are cascaded one after another; the type of the input to the network (e.g., single channel or different form of multi-channel), and/or; the layer whose output is treated as the DLR features.

- 1.1.3. **Multiple CNNs:** Beside using single standard or self-designed CNNs, some researchers have proposed to use multiple networks, which has the advantage of benefiting from multiple inputs, having various modalities, scales and angles as shown in Fig. 6, or different architectures with different properties.

“Scale” is a significant factor to consider when designing the input structure. For example to distinguish tumors from vessels, a large enough region should be included in the input patch, while to differentiate between solid and non-solid tumors, the nodule region should be the main core of the patch. Having this in mind, Reference [59] has designed a CNN architecture for lung tumor classification, where inputs are patches not only from different angles (sagittal, coronal, and axial) but also in different scales. Following a similar path, Reference [50] has also designed a multiple CNN architecture, where each CNN takes a lung tumor patch at a specific scale (illustrated in Fig. 6) as input and generates the associated DLR features. Features extracted from all the CNNs are

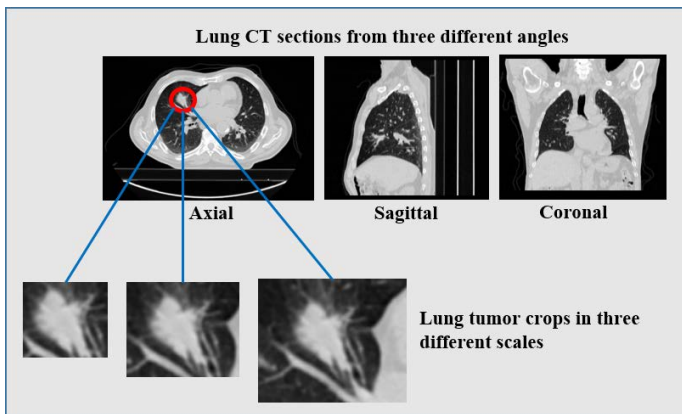


Fig. 6: Different angles of lung CT scan along with tumor crops in three different scales.

then concatenated and used for lung tumor malignancy prediction through a conventional classifier (SVM). The idea here is that segmenting the tumor regions is not always feasible. Furthermore using a tumor patch provides information on not only the tumor itself but also the surrounding tissues, and since tumor sizes can vary significantly among patients, using multi-scale patches instead of the single ones will improve the overall performance of the extracted DLR features. An interesting property of such multiple CNN architecture is that since the constituent CNNs share parameters, training can be performed in a reasonable time. Another benefit of using a multiple CNN architecture is that the network becomes robust to addition of small noise to the input.

Similar to the work in [50], Reference [56] has designed a CNN called “Multi-view CNN”, which uses 7 patches at different scales as inputs, with the difference that these patches are resized to have the same dimension, and therefore, a single CNN can be used instead of multiple CNNs. This work has also extended the binary lung tumor classification to a ternary classification to classify lung tumors as benign, primary malignant, and metastatic malignant. Furthermore, this article has adopted another validation approach called “separability” besides the common terms such as accuracy and AUC (area under curve). Separability refers to the extent that different classes are distinguishable based on the learned features, and according to the aforementioned article, the proposed multi-view CNN has a higher Separability compared to a single scale CNN. In addition to that, as the layers go deeper, features with higher separability are learned.

The idea of using multi-scale image patches is further expanded in Reference [55] through designing a novel CNN architecture called “Multi-crop CNN”, where instead of taking inputs in various scales, multi-scale features are extracted through parallel pooling layers, one of which applies pooling to a cropped version of the input from the previous layer. Features from multiple pooling layers are then concatenated and fed to the next layer. 3D lung CT images are inputs to this network, and since multiple

CNNs are replaced with one single CNN, the training can be performed in a more time effective manner. Besides forecasting the lung tumor malignancy, this work has also predicted other attributes associated with tumor such as diameter, by replacing the final SoftMax layer with a regression one. It is worth mentioning that this network is not performing all the assigned tasks simultaneously. Instead they are performed one after another, which distinguishes this network from a multitask training one discussed in section IV-B.

Radiomics through multiple CNNs is further explored recently in [51] for Alzheimer’s disease diagnosis using MRI, where in the first stage several landmarks are detected based on the comparison between normal and abnormal brains. These landmarks are then used to extract patches (separately around each individual landmark), and consequently each CNN is trained taking patches corresponding to a specific landmark position as input. Final decision is made based on a majority voting among all the CNNs. Here, the idea behind using a multiple architecture is the fact that detecting Alzheimer’s disease requires the examination of different regions of the brain.

In summary, multiple CNNs methods developed for DLR feature extraction are designed by either fusing the outputs of several CNNs which are trained based on a specific input, or multi-path layers are embedded within a single network to modify the output from previous layers differently.

One challenge shared among all the aforementioned CNN architectures is that they do not take the spatial information between objects into account. As an example, they may fail to consider the location of abnormality within the tissue as an indicator of its type. The newly proposed deep architecture called CapsNets, described next, is introduced to overcome this drawback.

1.2. Capsule Networks: Although CNNs are the state of the art in many medical and non-medical classification problems, they are subjected to several drawbacks including their low explainability and their negligence in preserving the spatial relationships between elements of the image leading to misclassification. Besides, CNNs have low robustness to some types of transformation. Loss of spatial relation information, which is associated with the pooling layers, is resolved by the newly proposed Capsule networks (CapsNets) [36] consisting of both convolutional and capsule layers that can handle more types of transformation. These deep architectures have the ability to consider the relationships between the location of objects and tolerate more types of transformation, through their routing by agreement process, which dictates that an object will not be classified as a specific category unless the lower level elements of this object agree on the existence of that category. Another important property of CapsNets is that they can handle smaller datasets, which is typically the case in most medical areas. Here we explain the architecture of Capsule networks, as illustrated in Fig. 7, and their routing by agreement process.

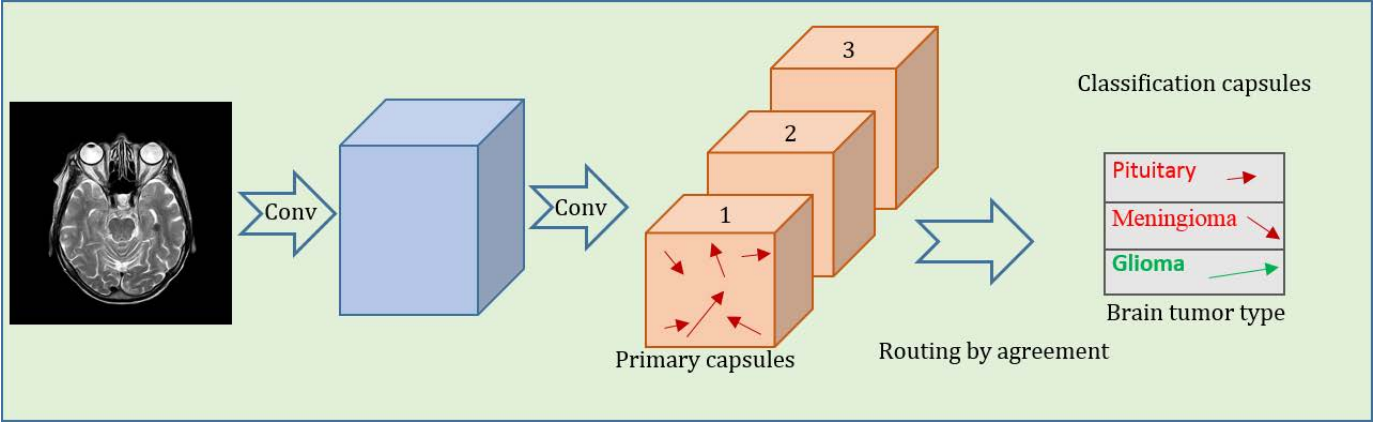


Fig. 7: Capsule network architecture. A convolutional layer is used to form the primary capsules, and the decision is made based on the agreement among these capsules.

Capsules are group of neurons whose activity vectors consist of various instantiation parameters, and the length of the activity vectors represent the probability of a specific instance being present. Each Capsule in the primary capsules layer tries to predict the outcome of all the capsules in the next layer (parent capsules), however, these predictions are sent to the next layer with different coefficients, which is based on how much the actual output agrees with the prediction. This process of looking for agreement between capsules before coupling them is called routing by agreement and it is the most important property of capsule networks, making them consider spatial relations among objects, therefore, being robust to several types of transformations such as affine transformation and rotation. Defining \mathbf{u}_i as the output of capsule i in the primary capsules layers, the prediction for capsule j , $\mathbf{u}_{j|i}$ is calculated as follows, where \mathbf{W}_{ij} is the weight matrix to be learned in back propagation

$$\mathbf{u}_{j|i} = \mathbf{W}_{ij} \mathbf{u}_i. \quad (7)$$

Consequently C_{ij} , the coupling coefficient of capsules i and j , is calculated based on the degree of conformation between these two capsules, based on the idea that if two vectors agree they will have a bigger inner product, and the output of parent capsule j , \mathbf{s}_j , is estimated as follows

$$\mathbf{s}_j = C_{ij} \mathbf{u}_{j|i}. \quad (8)$$

Finally, a non-linear function is applied to \mathbf{s}_j to always keep its length equal or smaller than one as this length should represent probability of an object being present.

Below we further investigate the initial utilization of CapsNets in Radiomics [37], for the first time. In [37] we have explored various CapsNet architectures to select the one that maximizes the prediction accuracy resulting in a network that has fewer convolutional filters (64 filters) compared to the original Capsule network which has 256 filters. This architecture consisting of one convolutional, one primary capsule, and a classification layer results in 86.56% accuracy.

Furthermore, separate networks are trained based on two types of inputs: the original brain images and the segmented

tumor regions, observing that CapsNet performs better when being fed with tumor masks, probably because the brain images have miscellaneous backgrounds, distracting the network from extracting important distinguishing features. Nevertheless, Capsule network is of higher accuracy for both of the input types compared to the CNN, which has 78% accuracy for tumor images. Several factors may have enabled the CapsNet to provide a better performance including its ability to handle small datasets, and being robust to transformations and rotations, resulted from the *routing by agreement* process.

1.3. Recurrent Neural Networks: Most of the deep network architectures need fixed-sized inputs, which makes them ineffective for Radiomics analysis of volumetric images (volume-level classification), i.e., when the whole volume is needed to be processed at once (such as tumor classification based on the 3D volume). In these scenarios, the RNNs can be adopted as they are capable of processing sequential data such as CT or MR slices, and they take both the present image slice and result of processing the previous ones as inputs. RNNs are also useful to monitor the medical images resulted from follow-up examinations (patient-level classification).

Since RNNs are associated with the vanishing gradient problem, a new type called long-short-term-memory (LSTM) is proposed which has the ability to decide what to store and what to forget. Although it seems that RNNs and LSTM are computationally more expensive than other architectures, their training time and cost is greatly reduced by using the same weights over the whole network [20]. Use of LSTMs is explored in Reference [61] for prostate cancer benign and malignant classification based on sequences of ultrasound images, where it has been shown that the predictive accuracy of this sequential classification is higher than making decision based on independent single images.

This completes the overview of deep discriminative models with application to Radiomics. Next, we briefly review the generative models.

2. Generative Models: The objective of most of the deep generative models is to learn abstract yet rich features from the data distribution in order to generate new samples from

the same distribution. What makes these models practical in Radiomics is the fact that the learned features are probably the best descriptors of the data, and thus have the potential to serve as Radiomics features and contribute to a consequent tasks such as tumor classification. Auto-encoder networks, deep belief networks, and deep Boltzmann machines are among the deep generative models that have been utilized in Radiomics works as outlined below:

2.1. Auto-Encoder Networks: An auto-encoder network consists of two main components: An encoder which takes as input N_s medical images denoted by $f^{(i)}$, for $(1 \leq i \leq N_s)$, and converts each into a latent space $\phi(\mathbf{W}f^{(i)} + \mathbf{b})$, i.e., Radiomics features. The second component, the decoder, takes the latent space and tries to reconstruct the input image with the objective of minimizing the difference between the original input and the reconstructed one $\phi(\mathbf{W}^T \phi(\mathbf{W}f^{(i)} + \mathbf{b}) + \mathbf{c})$ [64] given by

$$\min_{\mathbf{W}, \mathbf{b}, \mathbf{c}} \sum_{i=1}^{N_s} \|\phi(\mathbf{W}^T \phi(\mathbf{W}f^{(i)} + \mathbf{b}) + \mathbf{c}) - f^{(i)}\|, \quad (9)$$

where $\phi(\cdot)$ is the network's activation function; \mathbf{W} denotes the weight matrix of the network used by both the encoder and the decoder; Term \mathbf{b} denotes the encoder's bias vector; \mathbf{c} is the decoder's bias vector, and; superscript T denotes the transpose operator. The reason that the encoded variables can be treated as Radiomics features is that they are the most important representatives of the input image that can be used to reproduce it. Although an auto-encoder can be trained completely in an end-to-end manner, to begin training with good initial weights and thus avoid the vanishing gradient problem, one can first train layers one by one, and use the obtained weights as the auto-encoder starting point [20]. Depending on the application, Auto-encoders have several extensions including:

2.1.1. Denoising Auto-Encoders (DAEs): To make auto-encoders capture more robust features of the input, one common strategy is to add some noise to the input. This kind of auto-encoder is called a denoising auto-encoder (DAE) [20]. Reference [57] has adopted DAE for extracting Radiomics features that are fed to an SVM to classify lung tumors as benign or malignant. Reference [10] has also adopted a five layer denoising auto-encoder which takes the corrupted lung images as inputs and tries to recover the original image. In particular, 400 Features extracted by the encoder part of this network are treated as Radiomics to train another neural network for lung cancer classification (identify the type of the tumor such as benign or malignant).

2.1.2. Convolutional Auto-Encoders (CAEs): This type of auto-encoders are specially useful for Radiomics (image type inputs) as the spatial correlations are taken into account. In these networks, nodes share weights in a local neighborhood [20]. A CAE with 5 Convolutional layer is adopted in Reference [11] for lung cancer diagnosis (identify the presence of

cancer).

There are two common strategies to leverage Auto-encoders in Radiomics:

- The first and most frequent approach is to directly use the extracted features to train a classifier. For instance, [64] has extracted Radiomics features using a 5 layer auto-encoder, which receives the segmented region of interest as the input. These features go through a binary decision tree in the next step to produce the output which is the classified lung nodule in this case.
- Auto-encoders can also serve as a pre-training stage to make the network extract representative features before trying to perform the actual classification. For instance, Reference [60] has first trained a DAE based on resized (down-sampled images to facilitate training) lung CT patches. In the next stage, a classification layer is added to the network and the whole network is re-trained taking both resized images and the resizing ratio as inputs.

2.2. Deep Belief Networks (DBNs): DBNs are stack of Restricted Boltzmann Machines (RBMs) on top of each other where the RBM is an unsupervised two layer stochastic neural network that can model probabilistic dependencies with the objective of minimizing the reconstruction error. More importantly RBM is a bipartite graph allowing value propagation in both directions. Although DBNs are composition of RBMs, only the top two layers have undirected relations. DBNs are first trained in a greedy fashion meaning that RBM sub-networks are trained individually followed by a fine-tuning phase [20]. In Reference [10], a DBN consisting of 4 hidden layers is designed with the goal of extracting the DLR from the top layer which has 1600 nodes. This last layer is connected to an external neural network to classify lung nodules. Besides, to have multi-channel input (original image, segmented tumor, and gradient image), these channels are concatenated vector wise before being fed to the network.

2.3. Deep Boltzmann Machine (DBMs): DBMs are also based on RBMs, but they differ from DBNs in the sense that DBMs include undirected relations between all layers which makes them computationally ineffective, though they are trained in a layer wise manner [20]. Due to the two-way relations, however, RBMs can capture complicated patterns from the data [34]. DBMs are adopted in [34] for Alzheimers disease diagnosis. In this work, a classification layer is added to the last layer of the DBM allowing to extract not only hierarchical (generative) but also discriminative features.

This completes an overview of different deep discriminative and generative models used within the Radiomics workflow. Next, we consider a critical drawback of such architectures, i.e., acting as a black-box.

D. Explainability of Deep Learning-based Radiomics

Explainability of deep networks refers to revealing an insight of what has made the model to come into a specific

decision, helping with not only improving the model by knowing what exactly is going on in the network, but also detecting the failure points of the model. No matter how powerful DLR are, they will not be utilized by physicians, unless they can be interpreted and related to the image landmarks used by the experts. Besides, not even a single mistake is allowed in medical decisions as it may lead to a irreparable loss or injury, and having an explanation of the logic behind the outcome of the deep net is the key to prevent such disasters. This subsection will present an overview on recently developed techniques to increase the explainability of deep Radiomics.

One simple approach to ensure the accuracy of the automatic prediction, is to double-check the results with an expert. For instance, Reference [45], which has used a CNN for longevity prediction using CT images, has reviewed the outcomes with experts leading to the fact that people predicted with longer lives are indeed healthier. However, this approach is time consuming and needs complete supervision and investigation, which is in conflict with the concept of automatizing and personalized treatment, which is the whole point of Radiomics. Therefore, nowadays several criteria are being presented to reduce the time and complexity of explaining deep Radiomics. One of these approaches is “feature visualization” which tries to gain knowledge on the network behavior by visualizing what kinds of features the network is looking for. This technique can be applied to different layers of the model. For example, to visualize the first layer features, the associated filters are applied to the input and the resulting feature maps are presented. However, as the last layer is the most responsible one in the network’s output, paying attention to the features learned in this layer is more informative. For instance, Reference [10] has visualized the final weights of a DBN, showing that the network is looking for meaningful features such as curvity. Nevertheless, these features are not as meaningful as they are for simple image recognition tasks as clinicians themselves are sometimes unsure about the distinctive properties of the images.

One other method to provide the user with an explanation on the decision made by a deep architecture is called “sensitivity analysis” referring to generating a heat-map highlighting the image regions responsible for the output [44]. In the heat-map, the brighter areas are the ones that have influenced the prediction. This can be achieved by determining and measuring the effect of changing each individual input pixel on the output. In a CNN this effect can be estimated by determining the weight associated with each input pixel through back propagation. This approach can discover the cause for the prediction [44], however, the drawback of this approach is that not all the detected pixels through the heat-map are necessarily the ones leading to the specific decision, and besides, as the depth and complexity of the deep net increases, it becomes more difficult to measure the contribution of each individual pixel on the output.

A third proposed approach to understand the learned features is to project the high-dimensional feature space from the deep network to a bi-dimensional plane. Reference [59] has adopted this strategy by using t-Distributed Stochastic Neighbor Embedding (t-SNE) algorithm to visualize the features

learned by a CNN for lung tumor classification. The resulted plane presents clearly defined clusters of lung tumors, which shows that the networks has successfully learned discriminating features. However, although this method can verify the accuracy of the network, it does not provide information on the exact reason behind making the decision.

The interpretability of meaningless weights is improved in the newly proposed Capsule networks through reconstructing the input image based on the features learned by the network. CapsNet includes a set of fully connected layers that take the final calculated features, based on which the final classification is made, as inputs, and reproduce the original image with the objective of minimizing the difference between the original and the reconstructed image. This objective function is added to the classification loss with a smaller weight not to distract network from extracting discriminative features. If the trained CapsNets is not only of high accuracy but also capable of resembling the input image, it has been successful in extracting representative features. Besides, visualizing these features provides insight on the explainability of the model. Interestingly, CapsNets are equipped with a powerful feature visualization technique through their input reconstruction part, which works as follows:

1. If CapsNet is in the training phase, the feature vector associated with the true class label is selected, otherwise, the one with the higher probability is used.
2. The selected feature vector is tweaked, meaning that small numbers are added to or subtracted from the feature values leading to a slightly changed new feature vector.
3. The new feature vector is fed to the reconstruction part and the input image is reproduced. However this reconstructed image is not supposed to exactly resemble the input image as it is generated using the tweaked features not the actual ones learned by the network.
4. By repeating the process of tweaking and reconstructing process over and over again, one can understand what features are learned by observing the influence of changing them on the generated images.

This strategy is adopted for explaining the output of a CapsNet trained to classify brain tumors, where it is shown that the network is probably making the decision based on features such as size and deformation. Fig. 8 shows the reconstructed brain tumor images based on tweaked feature vectors helping to gain insight on the nature of features.

This completes our discussion on deep learning-based Radiomics. Next, we consider hybrid solutions in which hand-crafted and deep Radiomics are jointly used.

V. HYBRID SOLUTIONS TO RADIOMICS

To summarize our findings on HCR and DLR features, tables VII(a) and (b) provide different comparisons between these two categories from various perspectives. Fig. 9 shows the number of publications in four distinct applications of Radiomics, based on either hand-crafted or deep learning techniques. As it can be inferred from this figure, although deep learning techniques are much newer methods, compared

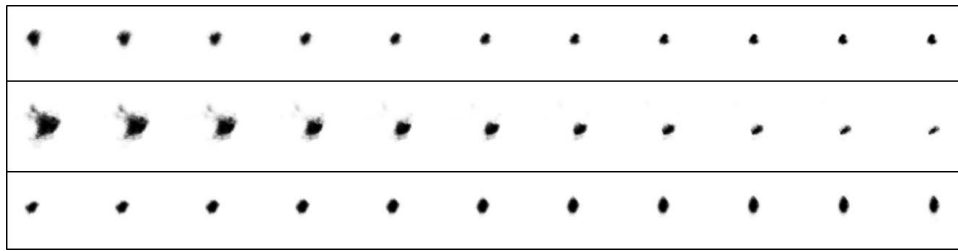


Fig. 8: Effect of tweaking the final feature vector on the reconstructed brain tumor image. Each row corresponds to a single feature which is tweaked 11 times.

TABLE VII: A Comparison between hand-Crafted and Deep Radiomics.

Hand-Crafted Radiomics (HCR)	Deep Radiomics (DLR)
Needs a prior knowledge on types of features to extract.	Can learn features on its own and without human intervention.
Features are typically extracted from the segmented ROI.	does not necessarily require a segmented input.
It is generally followed by a feature selection algorithm.	Feature selection is rarely performed.
As features are defined independent from the data, does not require big datasets.	Requires huge datasets, since it has to learn features from the data.
Processing time is not normally significant.	Can have high computational cost depending on the architecture and size of the dataset.
Since features are pre-designed, they are tangible.	The logic behind the features and decisions is still a black box.

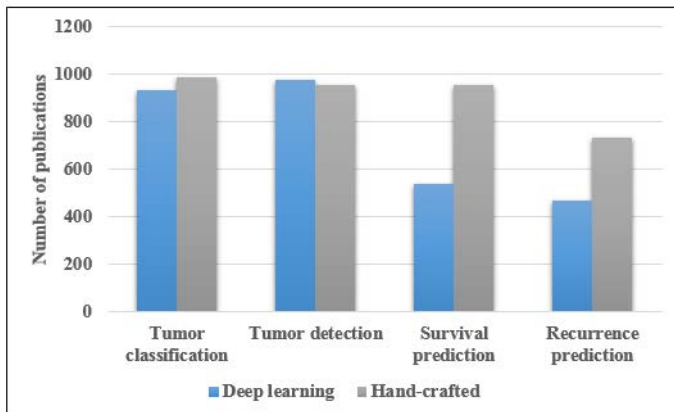


Fig. 9: Differences between the popularity of Hand-crafted and Deep learning-based Radiomics in four distinct applications, based on data from Google Scholar.

to hand-crafted ones, these two techniques are on a par with each other, in tumor classification and detection. However, the number of publications on deep learning methods is relatively less than the number of publications on hand-crafted methods, in survival and recurrence prediction, possibly because the domain knowledge, which is not accessible in deep learning-based Radiomics, is more important in survival and recurrence prediction. This calls for techniques that can utilize both the advantages of deep learning (not requiring segmentation, feature selection, and human intervention), and the domain knowledge available in hand-crafted approaches. Furthermore, in scenarios where neither of the above two mentioned categories are capable of providing informative Radiomics features

with high predictive capacity, one can resort to hybrid strategies. Here, potential hybrid solutions to Radiomics [65] are reviewed from different points of view including combination of Radiomics with other data sources and combination of HCR and DLR features.

A. Combination of Radiomics and Other Data Sources

Physicians, normally, do not rely on a single input for their diagnosis of diseases and disorders. To come into a conclusive decision, inputs from different sources are compared and combined including Radiomics (image bio-markers from different imaging modalities); Blood bio-markers; Clinical outcomes; Pathology, and; Genomics results [1]. In Information Post I, we have provided an overview of various Radiomics' imaging modalities and data sources which are typically combined with Radiomics features. Below, we discuss two different ways to fuse/combine Radiomics with other available resources of information along with the rationales and potentials behind such combinations:

- i. **Extracting Radiomics from Different Imaging Modalities:** As stated previously, Radiomics can be extracted from different imaging modalities each of which can only capture/provide particular information on tissues' properties. For instance, although the CT scan is among the most common and informative imaging modalities allowing to observe the body internal organs, the CT can not provide information on body function and metabolism. This type of information is available through PET scan, which calls for studying the effect of combining Radiomics extracted from different modalities. For example, in [4] Radiomics features are extracted from both CT and PET images,

and the concatenated feature vector is fed to a classifier for lung cancer survival prediction, resulting in a higher accuracy compared to each modality separately.

Combining different imaging modalities is also tested on brain tumor classification in [43]. Since MRI can output different images varying mostly in terms of their contrast and relaxation parameters, these images can be fused to provide complementary information. Based on [43], extracting Radiomics from this combination of MRIs outperforms the single modal classifier for brain tumor classification.

ii. **Integration of Extracted Radiomics with Other Data**

Sources: Radiomics features are combined with other resources only after the extraction process. The best descriptive or predictive models in the field of Radiomics are the ones that utilize not only imaging bio-markers, but also other information such as Genomics patterns and tumor histology [3]. In [5], it is reported that combining Radiomics features with lung cancer staging information, which is obtained based on the tumor location and dispersion, can improve the prognostic performance of Radiomics alone or staging alone. In other words, Radiomics can provide a complementary information for lung cancer prognosis [2]. It is also shown that combining Radiomics with other prognostic markers in head-and-neck cancer leads to a more inclusive decision. Combining Radiomics with clinical data is further investigated in [49] for brain cancer survival prediction. The interesting output of this work is a nomogram based on both Radiomics and clinical risk factors such as age that can be used to visually calculate survival probability. Developing such a nomogram is further investigated in references [40] and [63] for prediction of Hepatitis B virus and lymph node metastasis, respectively.

In brief, the first step to build a Radiomics-based nomogram, is calculating a linear combination of selected Radiomics features based on a logistic regression, which results in a Radiomics score to exploit further for the desired prediction task. Consequently, by training a multivariate logistic regression, Radiomics score is fused with other influential factors to make the final prediction. Fig. 10 presents the nomogram introduced in [63] along with an example illustrating how the lymph node metastasis prediction is made.

Reference [65] has adopted a fusion approach based on both Radiomics and Genomics bio-markers for predicting the recurrence risk of lung cancer. In this study, 35 hand-crafted features are extracted from segmented lung CT images and reduced to 8 after a feature selection phase. These features are then used to train a Naive Bayesian network. The same classifier is also trained using two Genomics bio-markers, and the outputs of two classifiers are fused through a simple averaging strategy. Results demonstrate that the combination of classifiers not only leads to higher prediction accuracy compared to individual ones, but also resembles the Kaplan-Meier plot of survival more precisely.

B. Fusion of HCR with DLR (i.e., Engineered Features Coupled with Deep Features)

As mentioned in Table VII, engineered (hand-crafted) and deep Radiomics both have their own advantages and disadvantages. As a result, combining these features has the promise of benefiting from both domains and incorporating different types of features [47] potentially results in significantly improved performance. As shown in Fig. 11, the following two categories of data fusion have been used in Radiomics most recently:

1. Decision-level Fusion: One common approach to combine HCR with DLR is to first use them separately to train separate classifiers and then adopt a kind of voting between the outputs to make the final decision. The voting or fusion approaches in Radiomics include:

- (i) **Soft Voting**, which is combining the probability outputs, for instance, through a simple averaging. Soft voting is adopted in [42], where two individual SVMs are trained on hand-crafted and deep Radiomics features (extracted using a pre-trained CNN), and consequently breast cancer prediction is performed based on averaging the output probabilities. Results of this article shows that the combined features are associated with higher prediction accuracy. Fusion of separately trained classifiers through soft-voting based on deep and hand-crafted Radiomics is also examined in [52] for breast cancer classification, where it has been shown that the combined SVM model outperforms individual classifiers in term of accuracy for mammogram, ultrasound, and MRI images.
- (ii) **Hard Voting**, which is combining outputs, for example, through a majority vote.
- (iii) **Adaptive Voting**, where a weight of importance for each model (HCR and DLR) is learned for example using a separate neural network. In Reference [53], a different kind of voting is adopted for lung cancer classification. This voting is based on the idea that not all the classifiers contribute equally to the final decision, and contribution weights are parameters that should be optimized through a learning process. The aforementioned article has first trained a CNN and several traditional classifiers such as SVM and logistic regression to independently predict the type (benign or malignant) of lung cancer. Predictions are consequently utilized to train a second-stage classifier to generate the final outcome. Any classifier such as SVM and NN can be used as the second-stage classifier.

2. Feature-level Fusion: Second widely used approach to combine deep and hand-crafted Radiomics is to first concatenate the feature vectors and then feed them to a classifier, referred to as feature-level fusion [57]. Reference [46] has shown that this combination lead to the highest performance in lung cancer survival prediction using Random Forest and Naive Bayes classifiers. The efficiency of this approach is also verified in [47] for lung tumor detection. Although mixing deep and hand-crafted Radiomics has several advantages such

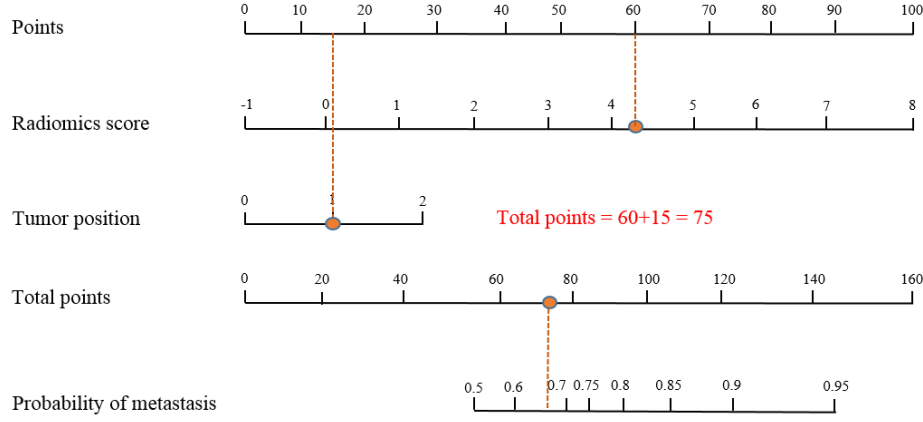


Fig. 10: Radiomics-based nomogram to predict lymph node metastasis. Tumor position is considered as extra information to assist with making a more reliable prediction.

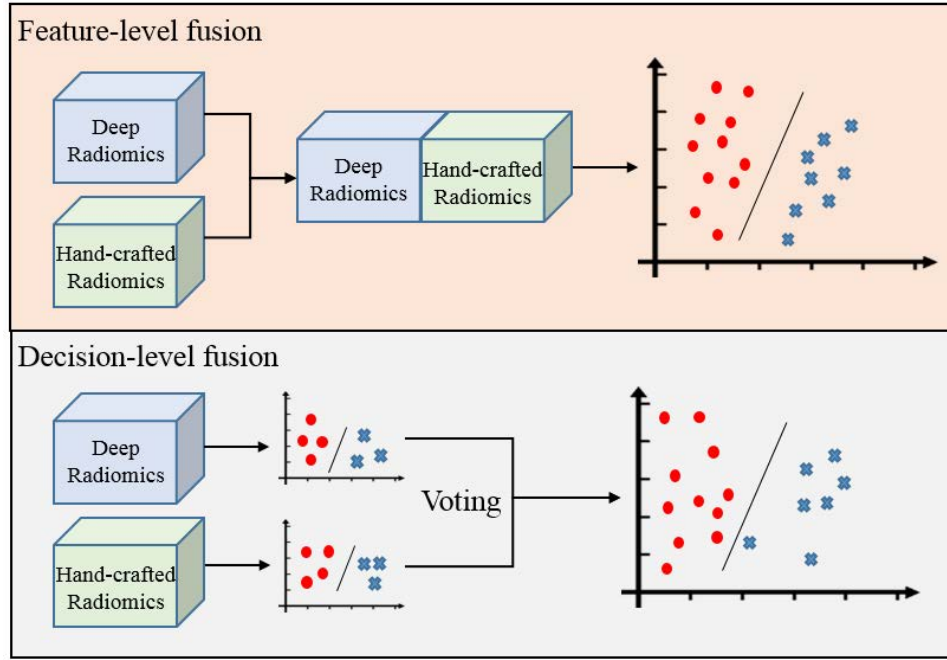


Fig. 11: Combining deep and hand-crafted Radiomics through feature-level or decision-level fusion.

as ensuring the heterogeneity of the features, it may cause over-fitting as the number of training data is relatively less than the number of features. Therefore, Reference [49] has examined this large set of features in terms of stability, informativeness, and redundancy leading to a dramatic dimension reduction and increase in the accuracy of brain cancer survival prediction. To further reduce the number of features, a Cox regression is adopted that can determine the impact of features on survival, and as a result, those with small weights can be removed as effectless.

Reference [54] has leveraged the idea of concatenating deep and hand-crafted features for lung tumor attributes (such as spiculation, sphericity and malignancy) scoring through a multi-task learning framework. For extracting deep features, 9 CNNs corresponding to each of the 9 task at hand, and a 3 layer DAE are trained, where each CNN generates 192 Ra-

diomics features extracted from the last fully connected layer before the SoftMax layer, and DAE results in 100 features. Deep features are further combined with hand-crafted features consisting of Haar and Histogram of oriented gradients (HoG) features, and the resulting vector is used as input to a multi-task linear regression model, which can consider the inter-task relations, in order to calculate the score of each of the 9 lung cancer attributes. This completes our discussion on hybrid Radiomics.

VI. CHALLENGES, OPEN PROBLEMS, AND OPPORTUNITIES

In this section, we will focus on the limitations of processing techniques unique in nature to the Radiomics, and then introduce open problems and signal processing opportunities as outlined in the following subsections.

A. Challenges of Hand-Crafted Radiomics

Extraction of hand-crafted features in Radiomics is more problematic in comparison to multimedia domains as, typically, very limited distinct visual variation exists to differentiate, for example, cancerous tissues. To address this issue, the common approach in Radiomics is to first extract hundreds and hundreds of different low level and high level features without devising a systematic mechanism and taking into account the end goal. Then, to resolve the resulting curse of dimensionality, simple and basic reduction techniques (e.g., basic principal component analysis (PCA)) are used. Another major issue with existing hand-crafted features in Radiomics (opportunity for signal processing researchers) is that they are extracted without using the information that can be obtained from other sources of data such as gene expressions and clinical data, which further limits their discriminative abilities for cancer prediction.

More importantly, most of the hand-crafted Radiomics require the segmented ROI. Providing annotations might not be a significant problem in other multi-media domains, as this project can be easily crowd sourced. However, when it comes to Radiomics, only experts have the ability to provide segmentations, which is both time and cost ineffective.

B. Challenges of Deep Radiomics

Although deep Radiomics has several advantages including its generalization capability and its independence from the supervision of experts, it is also associated with some disadvantages such as its need for huge data sets, and its lack of robustness to some kinds of transformation. Besides, another important challenge with deep learning-based Radiomics is that there is still no strategy to choose the optimum architecture.

Establishing the requirements of an appropriate deep architecture for Radiomics is a main challenge in development of DLR and another venue for future works. In particular, performing sensitivity analysis is a critical step to explain the connection between the designed choices and the achieved results. It is essential to identify different conditions under which the results are obtained. In short, what is happening when a specific architecture is used? For example, consider the simple issue of choosing the size of the input image. The question here is if one should provide the original image size or downsize the image and why? One simple answer could be that downsizing is applied for computational savings. Another intuitive answer could be that down sampling makes the data invariant to translation and rotation, i.e., when the image is down sampled, we take averages and as such the model will be less prone to outliers. The latter intuition could be a reasonable idea for natural images but is this the case for Radiomics-based features obtained from medical images?

Intuitively speaking, a fundamental challenge and an open problem for development of DLR is that the models and solutions developed for natural images should be modified before being applied to medical images as the nature of these images are totally different (e.g., consider MRI images). As the nature of the input signal is different, parts of the model

has to change but this issue has not yet been investigated systematically and thoroughly.

It is worth mentioning that while hand-crafted Radiomics normally need less amount of data, the number of images required for effective training of deep architectures, such as the CNNs, depends on the complexity of the underlying model. In other words, the amount of required training data increases when the number of trainable parameters increase. In the case of lung cancer diagnosis, e.g., the LIDC Dataset [27] is widely used, which consists of 244,527 CT images of 1010 patients. Normally, training complex deep models based on such dataset cannot be performed in a timely fashion using common processors and one needs to resort to one or more GPU processors. One common approach to reduce the computational cost in deep learning methods is to crop the input image to include just the object of interest. Although this approach does not cause any information loss in non-medical areas, it is harmful for classifying medical images as size is an important discriminator feature of, for instance, normal and abnormal tissues. This problem is investigated in [12], where it is shown that the miss-classified lung tumors based on features extracted from a DBN are 4% larger than correctly classified ones, possibly because the cropped images do not contain size information.

On the other hand, although limited access to training data is a common problem in other multimedia domains, it becomes significantly more critical in Radiomics as typically access to patients data is subject to several ethical regulations making it hard to collect the required amount of data for training purposes.

Finally, interpretability of deep Radiomics is of paramount importance as human lives are at stake, and without providing appropriate explainability, utilization of deep-Radiomics in clinical practice will remain limited. The relation between deep features and genetic patterns is also not established yet and requires further studies.

C. Open Problems and Signal Processing Opportunities

Despite recent advancements in the field of Radiomics and increase of its potential clinical applications, there are still several open problems which require extensive investigations including:

1. Most Radiomics models need rich amounts of training images, however, due to strict privacy policies, medical images are usually hard to collect.
2. Even without considering the privacy issues, it is difficult to find the required amount of data with similar clinical characteristics (e.g., corresponding to the same cancer stage).
3. Radiomics analysis need ground truth which is scarce as labels can only be provided by clinical experts (this is in contrary to other multimedia domains). This calls for development of weakly or semi-supervised solutions taking into account the specifics of the Radiomics domain.
4. Properties of medical images such as their contrast and resolution varies significantly from one institute to

another (from one dataset to another), because each institute may use different types of scanners and/or use different technical parameters. Development of novel and innovative information fusion methodologies together with construction of unifying schemes are critical to compensate for lack of standardization in this field and produce a common theme for comparing the Radiomics results. Furthermore ground truth and annotations provided by different experts can vary significantly as experts, depending on their area of specialty (such as oncology and surgery), may consider and look for different details and landmarks in the image.

5. Unbalanced data refers to a problem where classes are not equal in a classification problem rendering the classifier biased toward the majority class. This is almost all of the time the case for Radiomics analysis as the number of positive classes (existence of disease) is typically smaller than the negative ones. Therefore, proper care is needed when working with medical data. Although several solutions, such as modifying the metric function to give more weight to minority class, are provided to deal with the aforementioned issue, it is still an unsolved problem that needs further investigations.
6. Dealing with image noise is another challenging problem, which is common in all multi-media domains, but it is more severe in Radiomics as there may be more unpredictable sources of variation in medical imaging. As an example, patient's breathing in the CT scanner can cause change of lung tumor location in consecutive slices bringing about difficulty in extracting stable Radiomics features. Therefore to achieve reliable personalized diagnosis and treatment, careful strategies should be developed to address the effects of these kinds of variations.

Furthermore, there are several factors, such as imaging environments, capabilities of the scanners [79] and other shortcomings of radiological images (e.g., radiations during acquisition, noisy acquisitions), that limit the resolution of the obtained medical images. For instance, the range of the captured frequencies is limited by the maximum sampling rate of the scanner, and increasing the rate, will increase the resolution, at the cost of an increased noise [80]. Since access to high-quality images is necessary to achieve an early and accurate diagnosis/detection, there is an ongoing research on improving the quality of the medical images via development of advanced computational models to overcome the aforementioned shortcomings. One of such computational techniques is known as "Super-Resolution [81]", aiming at reconstructing a high-resolution image, using several low-resolution instances. Deep learning networks, and CNNs in particular, are widely used in super-resolution problems, and so far have shown promising results [81].

7. The biggest challenge in combining various data sources (such as imaging and clinical) is that not all data is provided for all the patients. In other words, Radiomics analysis model should be equipped with the ability to work with sparse data [3]. Besides, The currently used

fusion strategies within the Radiomics are still in their infancy and development of more rigorous fusion rules is necessary. For instance, feature-level fusion results in a vector and how to sort/combine the localized feature vectors is an open challenge. Giving the superiority of the initial results obtained from hybrid Radiomics, this issue becomes an urgent matter calling for advanced multiple-model solutions.

VII. CONCLUSION

During the past decades, medical imaging made significant advancements leading to the emergence of automatic techniques to extract information that are hidden to human eye. Nowadays, the extraction of quantitative or semi-quantitative features from medical images, referred to as Radiomics, can provide assistance in clinical care especially for cancer diagnosis/prognosis. There are several approaches to Radiomics including extracting hand-crafted features, deep features, and hybrid schemes. In this article, we have presented an integrated sketch on Radiomics by introducing practical application examples; Basic processing modules of the Radiomics, and; Supporting resources (e.g., image, clinical, and genomic data sources) utilized within the Radiomics pipeline, with the hope to facilitate further investigations and advancements in this field within signal processing community.

REFERENCES

- [1] P. Lambin, E. Rios-velazquez, *et al.*, "Radiomics: Extracting more Information from Medical Images using Advanced Feature Analysis," *European Journal of Cancer*, vol. 48, no. 4, pp. 441-446, 2012.
- [2] V. Kumar, Y. Gu, *et al.*, "Radiomics: The Process and the Challenges," *Magnetic Resonance Imaging*, vol. 30, no. 9, pp. 1234-1248, 2012.
- [3] R. Gillies, P. Kinahan, *et al.*, "Radiomics: Images Are More than Pictures, They Are Data," *Radiology*, vol. 278, No. 2, pp. 563-577, 2016.
- [4] A. Oikonomou, F. Khalvati, *et al.*, "Radiomics Analysis at PET/CT Contributes to Prognosis of Recurrence and Survival in Lung Cancer Treated with Stereotactic Body Radiotherapy," *Scientific Reports*, vol. 8, no. 1, 2018.
- [5] H.J. Aerts, E.R. Velazquez, *et al.*, "Decoding Tumour Phenotype by Noninvasive Imaging using a Quantitative Radiomics Approach," *Nature Communications*, vol. 5, 2014.
- [6] R.J. Gillies, A.R. Anderson, *et al.*, "The Biology Underlying Molecular Imaging in Oncology: from Genome to Anatome and Back Again," *Clinical Radiology*, vol. 65, no. 7, pp. 517-521, 2010.
- [7] M. Avanzo, J. Stancanella, I. El Naqa, "Beyond imaging: The promise of radiomics," *Physica Medica*, vol. 38, pp. 122-139, 2017.
- [8] Y. Zhang, A. Oikonomou, *et al.*, "Radiomics-based Prognosis Analysis for Non-small Cell Lung Cancer," *Scientific Reports*, vol. 7, 2017.
- [9] J. Griethuysen, A. Fedorov, *et al.*, "Computational Radiomics System to Decode the Radiographic Phenotype," *Cancer Research*, vol. 77, no. 21, pp. 104-107, 2017.
- [10] W. Sun, B. Zheng, *et al.*, "Automatic Feature Learning using Multichannel ROI based on Deep Structured Algorithms for Computerized Lung Cancer Diagnosis," *Computers in Biology and Medicine*, vol. 89, no. 1, pp. 530-539, 2017.
- [11] O. Echaniz, and M. Grana, *et al.*, "Ongoing Work on Deep Learning for Lung Cancer Prediction," *Biomedical Applications Based on Natural and Artificial Computing*, vol. 10338, pp. 42-48, 2017.
- [12] W. Sun, B. Zheng, *et al.*, "Computer Aided Lung Cancer Diagnosis with Deep Learning Algorithms," *Proc.SPIE*, vol. 9785, 2016.
- [13] H. Li, Y. Zhu, *et al.*, "MR Imaging Radiomics Signatures for Predicting the Risk of Breast Cancer Recurrence as Given by Research Versions of MammaPrint, Oncotype DX, and PAM50 Gene Assays," *Radiology*, vol. 281, no. 2, pp. 382-391, 2016.

- [14] A. Cunliffe, S. G. Armato III, *et al.*, "Lung Texture in Serial Thoracic Computed Tomography Scans: Correlation of Radiomics-based Features With Radiation Therapy Dose and Radiation Pneumonitis Development," *International Journal of Radiation Oncology*Biophysics*, vol. 91, no. 5, pp. 1048-1056, 2015.
- [15] H. Wang, Z. Zhou, *et al.*, "Comparison of Machine Learning Methods for Classifying Mediastinal Lymph Node Metastasis of Non-Small Cell Lung Cancer from 18F-FDG PET/CT Images," *EJNMMI Research*, vol. 7, no. 1, 2017.
- [16] M. Vallieres, D. Visvikis, M. Hatt, "Dependency of a Validated Radiomics Signature on Tumor Volume and Potential Corrections," *The journal of nuclear medicine (JNM)*, vol. 59 (supplement 1), 2018.
- [17] C. Parmar, P. Grossmann, *et al.*, "Machine Learning Methods for Quantitative Radiomic Biomarkers," *Scientific reports*, vol. 5, 2015.
- [18] J. Tian, D. Dong, *et al.*, "Radiomics in Medical Imaging-Detection, Extraction and Segmentation," *Artificial Intelligence in Decision Support Systems for Diagnosis in Medical Imaging*, vol. 140, pp. 267-333, 2018.
- [19] J. Edwards, "Medical Optical Imaging: Signal Processing Leads to New Methods of Detecting Life-Threatening Situations [Special Reports]," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 17-20, 2017.
- [20] D. Ravi, C. Wong, *et al.*, "Deep Learning for Health Informatics," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 4-21, 2017.
- [21] J. Constanzo, L. Wei, *et al.*, "Radiomics in Precision Medicine for Lung Cancer," *Translational lung cancer research*, vol. 6, no. 6, pp. 635-647, 2017.
- [22] C. Parmar, J.D. Barry, *et al.*, "Data Analysis Strategies in Medical Imaging," *Clinical cancer research*, vol. 24, no. 15, pp. 3492-3499, 2018.
- [23] G. Litjens, T. Kooi, *et al.*, "A Survey on Deep Learning in Medical Image Analysis," *Medical Image Analysis*, vol. 42, pp. 60-88, 2017.
- [24] R. Berenguer, M.D.R. Pastor-Juan, *et al.*, "Radiomics of CT Features May Be Nonreproducible and Redundant: Influence of CT Acquisition Parameters," *Radiology*, vol. 288, no. 2, pp. 407-415, 2018.
- [25] R. Thawani, M. McLane, *et al.*, "Radiomics and Radiogenomics in Lung Cancer: A Review for the Clinician," *Lung cancer*, vol. 115, pp. 34-41, 2017.
- [26] M. Sasaki, K. Yamada, *et al.*, "Variability in Absolute Apparent Diffusion Coefficient Values Across Different Platforms may be Substantial: A Multivendor, Multi-institutional Comparison Study," *Radiology*, vol. 249, no. 2, pp. 624-630, 2008.
- [27] S. Armato, G. McLennan, *et al.*, "Data From LIDC-IDRI. The Cancer Imaging Archive".
- [28] Z. Akkus, I. Ali, *et al.*, "Predicting Deletion of Chromosomal Arms 1p/19q in Low-Grade Gliomas from MR Images Using Machine Intelligence," *Journal of Digital Imaging*, vol. 30, no. 4, pp. 469-476, 2017.
- [29] M. Vallieres, E. Kay-Rivest, *et al.*, "Radiomics Strategies for Risk Assessment of Tumour Failure in Head-and-Neck Cancer," *Scientific Reports*, vol. 7, no. 1, 2017.
- [30] B. Menze, A. Jakab, *et al.*, "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," *IEEE Transaction on Medical Imaging*, vol. 34, no. 10, 2015.
- [31] A. Esteva, B. Kuprel, *et al.*, "Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks," *Nature*, vol. 542, pp. 115-118, 2017.
- [32] N. Menachemi, T.H. Collum, "Benefits and Drawbacks of Electronic Health Record Systems," *Risk Management and Healthcare Policy*, vol. 4, pp. 47-55, 2011.
- [33] M. Kolossvary, J. Karady, *et al.*, "Radiomic Features Are Superior to Conventional Quantitative Computed Tomographic Metrics to Identify Coronary Plaques With Napkin-Ring Sign," *Circulation: Cardiovascular Imaging*, vol. 10, no. 12, 2017.
- [34] H. Suk, S. Lee, *et al.*, "Hierarchical Feature Representation and Multimodal Fusion with Deep Learning for AD/MCI Diagnosis," *Neuroimage*, vol. 101, 2014.
- [35] A. Rahmim, P. Huang, *et al.*, "Improved Prediction of Outcome in Parkinson's Disease using Radiomics Analysis of Longitudinal DAT SPECT Images," *NeuroImage: Clinical*, vol. 16, pp. 539-544, 2017.
- [36] S. Sabour, N. Frosst, G.E. Hinton, "Dynamic Routing Between Capsules," *Conference on Neural Information Processing Systems (NIPS)*, pp. 3859-3869, 2017.
- [37] P. Afshar, A. Mohammadi, K.N. Plataniotis, "Brain Tumor Type Classification via Capsule Networks," *IEEE International Conference on Image Processing (ICIP)*, 2018.
- [38] C. Parmar, "Machine Learning Applications for Radiomics: Towards robust non-invasive predictors in clinical oncology," *Maastricht University*, 2017.
- [39] A. Subramanian, P. Tamayo, *et al.*, "Gene Set Enrichment Analysis: A knowledge-based Approach for Interpreting Genome-Wide Expression Profiles," *Proceedings of the National Academy of Sciences*, vol. 102, no. 43, pp. 15545-15550, 2005.
- [40] J. Peng, J. Zhang, *et al.*, "A Radiomics Nomogram for Preoperative Prediction of Microvascular Invasion Risk in Hepatitis B Virus-Related Hepatocellular Carcinoma," *Diagnostic and Interventional Radiology*, vol. 24, no. 3, pp. 121-127, 2018.
- [41] D. Kumar, A. Chung, *et al.*, "Discovery Radiomics for Pathologically-Proven Computed Tomography Lung Cancer Prediction," *Image Analysis and Recognition*, pp. 54-62, 2017.
- [42] B. Huynh, H. Li, *et al.*, "Digital Mammographic Tumor Classification Using Transfer Learning from Deep Convolutional Neural Networks," *Journal of Medical Imaging*, vol. 3, no. 3, 2016.
- [43] Z. Li, Y. Wang, *et al.*, "Deep Learning Based Radiomics (DLR) and Its Usage in Noninvasive IDH1 Prediction for Low Grade Glioma," *Scientific Reports*, vol. 7, no. 1, 2017.
- [44] A. Jamaludin, T. Kadir, *et al.*, "SpineNet: Automatically Pinpointing Classification Evidence in Spinal MRIs," *International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 9901, pp. 166-175, 2016.
- [45] L. Oakden-Rayner, G. Carneiro, *et al.*, "Precision Radiology: Predicting Longevity Using Feature Engineering and Deep Learning Methods in a Radiomics Framework," *Scientific Reports*, vol. 7, no. 1, 2017.
- [46] R. Paul, S. Hawkins, *et al.*, "Deep Feature Transfer Learning in Combination with Traditional Features Predicts Survival Among Patients with Lung Adenocarcinoma," *Tomography: a journal for imaging research*, vol. 2, no. 4, pp. 388-395, 2016.
- [47] L. Fu, J. Ma, *et al.*, "Automatic Detection of Lung Nodules: False Positive Reduction Using Convolutional Neural Networks and Handcrafted Features," *Proc.SPIE*, vol. 10134, 2017.
- [48] S. Liu, H. Zheng, *et al.*, "Prostate Cancer Diagnosis Using Deep Learning with 3D Multiparametric MRI," *Proc.SPIE*, vol. 10134, 2017.
- [49] J. Lao, Y. Chen, *et al.*, "A Deep Learning-Based Radiomics Model for Prediction of Survival in Glioblastoma Multiforme," *Scientific Reports*, vol. 7, no. 1, 2017.
- [50] W. Shen, M. Zhou, *et al.*, "Multi-scale Convolutional Neural Networks for Lung Nodule Classification," *International Conference on Information Processing in Medical Imaging*, 588-599, 2015.
- [51] M. Liu, J. Zhang, *et al.*, "Anatomical Landmark based Deep Feature Representation for MR Images in Brain Disease Diagnosis," *IEEE Journal of Biomedical and Health Informatics*, 2018.
- [52] N. Antropova, B. Huynh, *et al.*, "A Deep Feature Fusion Methodology for Breast Cancer Diagnosis Demonstrated on Three Imaging Modality Datasets," *The International Journal of Medical Physics Research and Practice*, vol. 44, no. 10, pp. 5162-5171, 2017.
- [53] S. Liu, Y. Xie, *et al.*, "Pulmonary Nodule Classification in Lung Cancer Screening with Three-Dimensional Convolutional Neural Networks," *Journal of Medical Imaging*, vol. 4, no. 4, 2017.
- [54] S. Chen, J. Qin, *et al.*, "Automatic Scoring of Multiple Semantic Attributes With Multi-Task Feature Leverage: A Study on Pulmonary Nodules in CT Images," *IEEE Transactions on medical imaging*, vol. 36, no. 3, pp. 802-814, 2017.
- [55] W. Shen, M. Zhou, *et al.*, "Multi-crop Convolutional Neural Networks for Lung Nodule Malignancy Suspiciousness Classification," *Pattern Recognition*, vol. 61, pp. 663-673, 2017.
- [56] K. Liu, G. Kang, "Multiview Convolutional Neural Networks for Lung Nodule Classification," *International journal of imaging systems and technology*, vol. 27, no. 1, pp. 12-22, 2017.
- [57] B. Kim, Y. Sung, *et al.*, "Deep Feature Learning for Pulmonary Nodule Classification in a Lung CT," *2016 4th International Winter Conference on Brain-Computer Interface (BCI)*, 2016.
- [58] W. Shen, M. Zhou, *et al.*, "Learning from Experts: Developing Transferable Deep Features for Patient-Level Lung Cancer Prediction," *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 124-131, 2016.
- [59] F. Ciompi, K. Chung, *et al.*, "Towards Automatic Pulmonary Nodule Management in Lung Cancer Screening with Deep Learning," *Scientific Reports*, vol. 7, 2017.
- [60] J. Cheng, D. Ni, *et al.*, "Computer-Aided Diagnosis with Deep Learning Architecture: Applications to Breast Lesions in US Images and Pulmonary Nodules in CT Scans," *Scientific Reports*, vol. 6, 2016.

- [61] S. Azizi, S. Bayat, *et al.*, "Deep Recurrent Neural Networks for Prostate Cancer Detection: Analysis of Temporal Enhanced Ultrasound," *IEEE Transactions on Medical Imaging*, 2018.
- [62] V. Parekh, M. Jacobs, "Radiomics: A New Application from Established Techniques," *Expert Review of Precision Medicine and Drug Development*, vol. 1, no. 2, pp. 207-226, 2016.
- [63] C. Shen, Z. Liu *et al.*, "Building CT Radiomics Based Nomogram for Preoperative Esophageal Cancer Patients Lymph Node Metastasis Prediction," *Translational Oncology*, vol. 11, no. 3, pp. 815-824, 2018.
- [64] D. Kumar, A. Wong, *et al.*, "Lung Nodule Classification Using Deep Features in CT Images," *International Conference on Computer and Robot Vision*, 2015.
- [65] N. Emaminejad, W. Qian, *et al.*, "Fusion of Quantitative Image and Genomic Biomarkers to Improve Prognosis Assessment of Early Stage Lung Cancer Patients," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 5, pp. 1034-1043, 2016.
- [66] M.J. Shafiee, A.G. Chung, *et al.*, "Discovery Radiomics via Evolutionary Deep Radiomic Sequencer Discovery for Pathologically Proven Lung Cancer Detection," *Journal of medical imaging*, vol. 4, no. 4, 2017.
- [67] O. Ronneberger, Ph. Fischer, T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 234-241, 2015.
- [68] S. Bickelhaupt, D. Paech, *et al.*, "Prediction of Malignancy by a Radiomic Signature from Contrast Agentfree Diffusion MRI in Suspicious Breast Lesions found on Screening Mammography," *Journal of Magnetic Resonance Imaging*, vol. 46, no. 2, pp. 604-616, 2017.
- [69] Y. Zhou, L. He, *et al.*, "CT-based Radiomics Signature: a Potential Biomarker for Preoperative Prediction of Early Recurrence in hepatocellular carcinoma," *Abdominal Radiology*, vol. 42, no. 6, pp. 16951704, 2017.
- [70] H. Yu, J. Scalera, *et al.*, "Texture Analysis as a Radiomic Marker for Differentiating Renal Tumors," *Abdominal Radiology*, vol. 42, no. 10, pp. 24702478, 2017.
- [71] K.H. Cha, L. Hadjiiski, *et al.*, "Bladder Cancer Treatment Response Assessment in CT using Radiomics with Deep-Learning," *Scientific reports*, vol. 7, no. 1, 2017.
- [72] R.T.H. Leijenaar, G. Nalbantov, *et al.*, "The Effect of SUV Discretization in Quantitative FDG-PET Radiomics: The Need for Standardized Methodology in Tumor Texture Analysis," *Scientific reports*, vol. 5, 2015.
- [73] A. Zwanenburg, S. Leger, *et al.*, "Image Biomarker Standardisation Initiative," *arXiv preprint arXiv:1612.07003*, 2018.
- [74] C. Parmar, P. Grossmann, *et al.*, "Radiomic Machine-Learning Classifiers for Prognostic Biomarkers of Head and Neck Cancer," *Frontiers in oncology*, vol. 5, no. 272, 2015.
- [75] C. Szegedy, V. Vanhoucke, *et al.*, "Rethinking the Inception Architecture for Computer Vision," *arXiv preprint arXiv:1512.00567v3*, 2015.
- [76] V. Gulshan, L. Peng, *et al.*, "Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs," *Journal of the American Medical Association (JAMA)*, vol. 16, no. 22, pp. 2402-2410, 2016.
- [77] C. Szegedy, W. Liu, *et al.*, "Going Deeper with Convolutions," *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
- [78] Y. Liu, J. Kim, *et al.*, "Radiomic Features are Associated with EGFR Mutation Status in Lung Adenocarcinomas," *Clinical lung cancer*, vol. 17, no. 5, pp. 441-448, 2016.
- [79] D. Trinh, M. Luong, *et al.*, "Novel Example-Based Method for Super-Resolution and Denoising of Medical Images," *IEEE Transactions on Image Processing*, vol. 23, no. 4, pp. 1882-1895, 2014.
- [80] H. Greenspan, "Super-Resolution in Medical Imaging," *The Computer Journal*, vol. 52, no. 1, pp. 43-63, 2009.
- [81] C. Dong, C. Change Loy, *et al.*, "Learning a Deep Convolutional Network for Image Super-Resolution," *European Conference on Computer Vision (ECCV)*, pp. 184-199, 2014.
- [82] J. Dehmshki, H. Amin, *et al.*, "Segmentation of Pulmonary Nodules in Thoracic CT Scans: A Region Growing Approach" *IEEE transactions on medical imaging*, vol. 27, no. 4, pp. 467-480, 2008.
- [83] M. Anthimopoulos, S. Christodoulidis, L. Ebner, T. Geiser, A. Christe, S. Mougiakakou, "Semantic Segmentation of Pathological Lung Tissue with Dilated Fully Convolutional Networks" *arXiv*, preprint arXiv: 1803.06167, 2018.
- [84] G. Huang, Z. Liu, L. van der Maaten, K. Q. Weinberger, "Densely Connected Convolutional Networks" *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [85] P. Wang, P. Chen, Y. Yuan, D. Liu, Z. Huang, X. Hou, G. Cottrell, "Understanding Convolution for Semantic Segmentation" *arXiv*, preprint arXiv: 1702.08502, 2018.
- [86] A. A. Farag, H. E. Abd El Munim, *et al.*, "A Novel Approach for Lung Nodules Segmentation in Chest CT Using Level Sets" *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 5202-5213, 2013.

Parnian Afshar (p_afs@encs.concordia.ca) is a Ph.D. candidate at Concordia Institute for Information System Engineering (CIISE). Her research interests include signal processing, biometrics, image and video processing, pattern recognition, and machine learning. She has extensive research/publication record in medical image processing related areas.

Arash Mohammadi (arash.mohammadi@concordia.ca) is an Assistant Professor with Concordia Institute for Information System Engineering (CIISE), at Concordia University. He is the Director-Membership Services of IEEE Signal Processing Society (SPS), and was General co-chair of the Symposium on Advanced Bio-Signal Processing and Machine Learning for Medical Cyber-Physical Systems, under IEEE GlobalSIP 2018, and the lead Organizer of the 2018 IEEE SPS Video and Image Processing (VIP) Cup. He also was the General Co-Chair of 2016 IEEE SPS Winter School on Distributed Signal Processing for Secure Cyber Physical Systems. His research interests include statistical signal processing, biomedical signal processing, image/video processing, and machine learning.

Konstantinos N. Plataniotis (kostas@ece.utoronto.ca) Bell Canada Chair in Multimedia, is a Professor with the ECE Department at the University of Toronto. He is a registered professional engineer in Ontario, Fellow of the IEEE and Fellow of the Engineering Institute of Canada. Dr. Plataniotis was the IEEE Signal Processing Society inaugural Vice President for Membership (2014-2016) and the General Co-Chair for the IEEE GlobalSIP 2017. He co-chaired the 2018 IEEE International Conference on Image Processing (ICIP 2018), October 7-10, 2018, Athens Greece, and co-chairs 2021 IEEE International Conference in Acoustics, Speech & Signal Processing (ICASSP 2021), Toronto, Canada.

Anastasia Oikonomou (anastasia.oikonomou@sunnybrook.ca), is the head of the Cardiothoracic Imaging Division at Sunnybrook Health Science Centre; Site Director of the Cardiothoracic Imaging Fellowship program at University of Toronto, and an Assistant Professor with the Department of Medical Imaging at the University of Toronto. Her research interests include imaging of pulmonary malignancies and interstitial lung diseases, Radiomics and machine learning methods in imaging of pulmonary disease.

Habib Benali (habib.benali@concordia.ca) is a Canada Research Chair in "Biomedical Imaging and Healthy Aging," is the Interim Scientific Director of the PERFORM Centre, and a Professor with the ECE Department at Concordia University. Dr. Benali was the director of Unit 678, Laboratory of Functional Imaging of the French National Institute of Health and Medical Research (INSERM) and Paris 6 University (UPMC). He was also the director of the International Laboratory of Neuroimaging and Modelisation of the INSERM-UPMC and Montreal University. His research interests include biological signal processing, neurology and radiology, biomedical imaging, biostatistics, and bioinformatics.