






Article

Radiomics-Based Prediction of Overall Survival in Lung Cancer Using Different Volumes-Of-Interest

Natascha Claudia D'Amico ^{1,2}, Rosa Sicilia ^{1,*} , Ermanno Cordelli ¹, Lorenzo Tronchin ¹ , Carlo Greco ³, Michele Fiore ³ , Alessia Carnevale ³, Giulio Iannello ¹ , Sara Ramella ³ and Paolo Soda ¹ 

¹ Unit of Computer Systems and Bioinformatics, Department of Engineering, Università Campus Bio-Medico di Roma, Via Alvaro del Portillo 21, 00128 Rome, Italy; n.damico@unicampus.it (N.C.D.); e.cordelli@unicampus.it (E.C.); lorenzotronchin@gmail.com (L.T.); g.iannello@unicampus.it (G.I.); p.soda@unicampus.it (P.S.)

² Department of Diagnostic Imaging and Sterotactic Radiosurgery, Centro Diagnostico Italiano S.p.A., Via S. Saint Bon 20, 20147 Milan, Italy

³ Radiation Oncology, Università Campus Bio-Medico di Roma, Via Alvaro del Portillo 21, 00128 Rome, Italy; c.greco@unicampus.it (C.G.); m.fiore@unicampus.it (M.F.); a.carnevale@unicampus.it (A.C.); s.ramella@unicampus.it (S.R.)

* Correspondence: r.sicilia@unicampus.it; Tel.: +39-06-225419622

Received: 14 August 2020; Accepted: 11 September 2020; Published: 15 September 2020



Featured Application: The manuscript aims to provide a signature to predict Overall Survival in patients with Locally Advanced Non-Small Cell Lung Cancer. The results could offer the physicians advanced software tools to early evaluate the disease evolution before the treatment start so to personalize each patient's therapy. Moreover, this work provides insight into the use of the different segmentation volumes usually applied in radiation oncology.

Abstract: Lung cancer accounts for the largest amount of deaths worldwide with respect to the other oncological pathologies. To guarantee the most effective cure to patients for such aggressive tumours, radiomics is increasing as a novel and promising research field that aims at extracting knowledge from data in terms of quantitative measures that are computed from diagnostic images, with prognostic and predictive ends. This knowledge could be used to optimize current treatments and to maximize their efficacy. To this end, we hereby study the use of such quantitative biomarkers computed from CT images of patients affected by Non-Small Cell Lung Cancer to predict Overall Survival. The main contributions of this work are two: first, we consider different volumes of interest for the same patient to find out whether the volume surrounding the visible lesions can provide useful information; second, we introduce 3D Local Binary Patterns, which are texture measures scarcely explored in radiomics. As further validation, we show that the proposed signature outperforms not only the features automatically computed by a deep learning-based approach, but also another signature at the state-of-the-art using other handcrafted features.

Keywords: radiomics; NSCLC; Local Binary Patterns; multi-VOI analysis

1. Introduction

Lung cancer is worldwide recognised as the most common type of cancer, with about 18.4% of all cases [1], also accounting for the largest amount of deaths. Every year, there are about 470 thousand new cases in Europe and the Non-Small Cell Lung Cancer (NSCLC) is the most frequent one with about 85–90% of all cases [2,3] and, due to the lack of apparent symptoms in early stages, approximately 70% of the new cases are diagnosed with stage III or IV. There are several treatment options that are

selected depending on the patient's characteristics; they include the radiotherapy, the chemotherapy, the surgical resection, the immunotherapy, but also targeted therapy [1,4,5].

We acknowledge that, nowadays, the standard procedure for stage III patients with NSCLC accounts chemoradiotherapy followed by immunotherapy [6]. However, when considering the guidelines at the time of this work, for stage III NSCLC, whether they were resectable or not, chemoradiotherapy (even concurrent followed by immunotherapy) was the recommended therapeutic approach [7,8], because showed better results than those that were attained by radiotherapy or by radiotherapy followed by chemotherapy [9,10]. Indeed, Curran Jr. et al. showed that patients randomly assigned to two different concurrent chemoradiotherapies had a longer survival time than patients who underwent sequential radio and chemotherapy [9].

The Overall Survival (OS), defined as the length of time from the date of diagnosis until the patient death, helps to identify patient subgroups with better or worse prognosis, so to hypothesize intensified treatment options for the latter group. Unfortunately, local recurrence and distant metastasis are still an issue [11] and, therefore, the strategies to improve OS are urgently needed.

In order to help the personalised medicine to make further steps facilitating the identification of the correct treatment for each patient, in recent years we have observed a raising interest in the development and application of Artificial Intelligence (AI) methods to oncology, similarly to what already happened with success in other medical fields [12–16]. It turns out this has first favoured the birth of radiomics, and now it is boosting the interests of the scientific community on its potential. Radiomics is based on the extraction of quantitative biomarkers, also called features, from medical images, such as CT, MRI, X-ray and PET, in order to diagnose, predict clinical outcome, or guide clinical decisions [17,18]. The first studies reported promising performances in different clinical applications for diagnostic and prognostic purposes [19–23].

Several radiomics studies have been conducted on NSCLC, but, to the best of our knowledge, the discovery of a signature associated to the Overall Survival is still an issue deserving more research efforts. In this respect, the contribution of this manuscript is twofold. First, we investigate the use of different Volumes-Of-Interest (VOIs) by exploiting quantitative biomarkers computed not only from the Gross Tumour Volume (GTV), which is what can be seen, palpated, or imaged, but also from other two larger VOIs commonly used in radiation treatment clinical practice, which are Clinical Target Volume (CTV) and Planning Target Volume (PTV). The first take into account the subclinical disease spread and the last allows for uncertainties in planning or treatment delivery. The idea, inspired by the radiotherapy practice adopted when planning the treatments, is that the area surrounding the tumour may contain important information that can be quantitatively exploited by a computer-based approach, although it cannot be discerned by the visual interpretation. In fact, this area around the tumour could catch also how the cancer infiltrates the nearby healthy tissues and, hence, potentially it could help predicting the disease behaviour and patient survival. The other contribution of this manuscript introduces the Local Binary Pattern (LBP) in the field of radiomics. LBP is a texture measure that is successfully used in computer vision, that here we extend to a three-dimensional (3D) environment and we apply with prognostic values in oncology, extending the set of texture features typically used in the radiomics literature, which are based on the two-dimensional (2D) and 3D second-order joint probability functions and their derivatives. To deepen our analysis, the approach presented here is compared to a deep approach, where features are automatically computed, as well as to the signature proposed by Aerts, et al. [19], which is the most cited at the state-of-the-art.

2. Background

Radiomics is an emerging trend that meets artificial intelligence and medical applications with prognostic perspective. Examples are mostly related with oncology, with efforts directed towards glioblastoma [24–26], oesophageal cancer [27], gastric cancer [28], soft tissue sarcoma [29], and lung tumours [19,30,31]. Nevertheless, recent examples outside the oncological field are related to COVID-19

outcome [32,33], cardiological issues [34], meningioma subtype, pneumonia, and parkinson outcome prediction [35–37].

Because of the topic discussed, hereinafter we shortly overview the work concerned with the prediction of the Overall Survival in patients suffering from the NSCLC [19,20,30,38,39] (Table 1). Furthermore, the interested readers can deepen AI applications in lung cancer starting from two reviews that discuss the prediction of lung cancer prognosis [40,41].

The literature explored so far shows a clear predominance of studies that focus on radiomics descriptors that were extracted from the GTV.

Moreover, the analysis of the literature reveals to us that most of the work computed both 2D and 3D shape and texture descriptors that mostly leverage on the Grey-Level Co-occurrence Matrices or wavelets. This can be justified by the fact that healthy tissue often differs from tumour tissue, mainly at the microstructure level.

On these grounds, the contribution of this manuscript is twofold: first, we investigate how the use of different volumes of interest impacts the Overall Survival prediction. Second, we introduce the 3D version of the local binary pattern (LBPs), a texture descriptor that is widely used for texture analysis in 2D images in other fields [42], but still not extended to the radiomic field.

Table 1. Summary of the background on the radiomics approach to predict the Overall Survival in lung cancer.

Author	Images	Segmented Volume	Study Population	Number of Patients	Number of Features	Features Description	Learning Model	Performance
Huynh et al.	CT	GTV	NSCLC, I and II stages	113	1605	shape, intensity, texture, Laplacian of Gaussian filter and wavelets features	Kaplan-Meier method	AUC = 0.67
Aertset al.	CT	GTV	lung and neck cancer	1091	440	tumour intensity, shape, texture and wavelet features	multivariate Cox proportional hazards regression model	AUC = 0.69
Parmar et al.	CT	GTV	NSCLC, I-III/IV stages	464	440	shape, intensity, texture, and wavelets features	Wilcoxon test for feature selection and a Random Forest as classification model	AUC = 0.66
Fave et al.	CT	GTV	NSCLC, III stage	107	212	shape and texture features	multivariate model	AUC = 0.675
Li et al.	CT	GTV	NSCLC, I and II stages	100	722	morphology and texture based features	Cox proportional hazards regression model	AUC = 0.64

3. Materials

This work studies 97 patients with NSCLC stage III that were treated with definitive concurrent chemoradiotherapy. The enrollment protocol was approved by Ethical Committee of Campus Bio-Medico University on 30 October 2012 and registered at ClinicalTrials.gov on 12 July 2018 with Identifier NCT03583723 after an initial exploratory phase. The Institutional review board approved this analysis, and a written informed consent was collected from all the patients. For each patient, the simulation CT images were acquired before the treatment using a Siemens Somatom Emotion, with 140 Kv, 80 mAs, and 3 mm for slice thickness. Subsequently all of the images were preprocessed using a lung filter (kernel B70) and a mediastinum filter (kernel B31). The patients were then clinically followed, for a median follow-up of 18.55 months after that it was possible to divide patient into two classes: 53 dead (class 0) and 44 alive (class 1). The patients had a mean Overall Survival (OS) time of 28.7 ± 26.4 months (min 4.8 months, max 142.6 months).

In this study we consider three different VOIs for each patient, the Gross Tumour Volume (GTV), the Clinical Target Volume (CTV), and the Planning Target Volume (PTV) segmented by radiation oncologists, as already introduced in Sections 1 and 2. For GTV, the expert radiation oncologists included all the macroscopic disease defined at CT or PET-CT and nodes PET positive and/or larger than 1 cm in the short axis at CT imaging. Starting from GTV, the other volumes can be segmented: CTV former contains GTV plus a margin for sub-clinical disease spread, which therefore cannot be fully imaged, and it represents the volume that must be adequately treated to achieve cure [43]. PTV is a geometric concept that is designed to ensure that the radiotherapy dose is actually delivered to CTV. It is determined expanding CTV with a safety margin, which is necessary to manage internal motion and set-up reproducibility [44].

It is worth noting that the ability to segment tumour mass and the surrounding tissue involved with the tumour to a certain extent is an important professional skill for a medical speciality that is radiation oncology. Several methods are proposed in the literature to reduce interpersonal variability [45,46]. This is why, in our work, the same radiation oncologist segmented all the images in our repository, and then another lung cancer expert radiation oncologist proofread the segmentations. Furthermore, to mitigate the inter-reader variability the segmentations were performed according to the international guidelines [47].

Figure 1 shows an example of these VOIs: the GTV is drawn in red, and it is the smallest segmentation that precisely delineates the visible tumour; CTV is represented in yellow and it includes areas, where, in the image, there is no evidence of tumour, but where sub-clinical disease may be. PTV, in blu, is the largest segmentation.

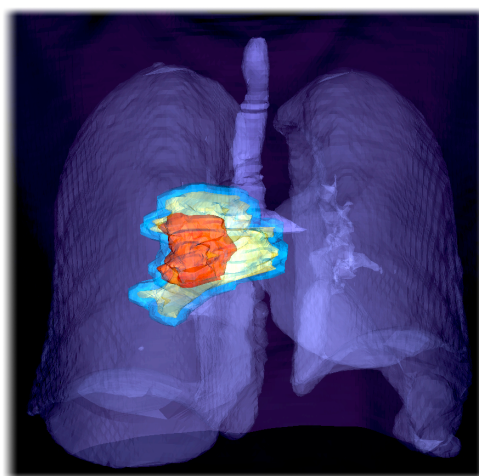


Figure 1. Example of a CT image of a patient suffering from Non-Small Cell Lung Cancer (NSCLC), where we highlight the segmentations corresponding to Gross Tumour Volume (GTV) (red), Clinical Target Volume (CTV) (yellow), and Planning Target Volume (PTV) (blue).

4. Methods

Figure 2 details the pipeline designed in this work, which can be divided in three main blocks: data processing with the feature computation, feature selection and final performance computation. First, the data processing flow is represented in the uppermost green block: it consists in the segmentation of the three VOIs (GTV, CTV, and PTV) from patients' CT scans and, then, in the computation of the descriptors outlined in the next Section 4.1. Second, the feature selection procedure and the final performance computation are both included in the grey block. Proceeding from left to right, the descriptors are standardised and ranked in the blue box, then we find the final subset selection in the purple box that is further detailed in the bottommost diagram. Third, the final model performance is shown in the rightmost grey blocks. All of these processes are fully described in the followings.

4.1. Feature Computation

For each patient, several quantitative biomarkers were extracted from all the VOIs. The features were computed using an in-house software tool coded in MATLAB R2019a (The MathWorks, Inc., Natick, MA, USA) that calculates the first order statistical features and two families of textural features: 3D Grey Level Co-occurrence Matrix (GLCM) and Three Orthogonal Planes-Local Binary Patterns (TOP-LBP). The TOP-LBP implementation of 3D LPBs computationally simplifies the extraction process, as discussed below.

First-order features: all of these features are computed from the grey levels' histogram of a ROI and therefore try to describe the statistical distribution of tissue's density inside the volume. From such histograms, we extracted 12 descriptors, which are the moments from first to fourth-order, namely the mean, the standard deviation, the skewness and the kurtosis, the histogram width, the energy, the entropy, the value of the histogram absolute maximum and the corresponding grey-level value, the energy around such maximum, and the number of relative maxima in the histogram and their energy.

3D Grey Level Co-occurrence Matrix: the healthy and the cancerous tissues usually present different structural patterns: for this reason, we go beyond the grey levels distribution of the VOIs voxels performing an additional textural analysis. In this respect, we computed the 3D Grey Level Co-occurrence Matrix, which generalises the 2D GLCM to the third dimension, being able to catch the differences of the tissues at the micro scale.

Given a 3D grey-scale image I and a 3D Cartesian reference system $O(x, y, z)$, whose origin is located in the top-front-left corner of I , the position of each voxel can be identified by a vector $\mathbf{p} = p_x\hat{i} + p_y\hat{j} + p_z\hat{k}$, with p_x, p_y and $p_z \in \mathbb{N}$. We also denote a displacement vector $\mathbf{d} = d_x\hat{i} + d_y\hat{j} + d_z\hat{k}$, with d_x, d_y and $d_z \in \mathbb{N}$. If we define m equal to the number of bit used to represent I , a *GLCM3* is a square matrix of size $N = 2^m$, where each entry (g_i, g_j) , with both g_i and $g_j \in [0, 2^m - 1]$, represents the number of times a voxel in \mathbf{p} with intensity g_i is separated by a displacement \mathbf{d} from another voxel with intensity g_j , therefore located in $\mathbf{p} + \mathbf{d}$. Hence, a *GLCM3* matrix synthesises the count of how many values co-occur between neighbouring pixels according to specific displacements. Those quantities vary, depending on the VOI structure. Denoting as d_h the h -th component of \mathbf{d} , in order to assess all possible directions, we considered the combination of $d_h \in \{-1, 0, 1\}$ as displacements, without considering the $(0, 0, 0)$ vector.

This leads to 26 different displacement directions. Furthermore, when considering that each direction $+\mathbf{d}$ produces a *GLCM3* matrix that is the transpose of the one produced by its opposite direction $-\mathbf{d}$, thus giving a redundant informative contribution, we only considered the 13 different directions given by $+\mathbf{d}$. More details on *GLCM3* can be deepened in [48]. Finally, from each *GLCM3*, we extract seven second order statistical features, referred to as Haralick descriptors [49]. Namely, they are autocorrelation, homogeneity, entropy, energy, covariance, inertia, and absolute contrast. Their formal definition is reported in the Appendix A. Concatenating such measures for each *GLCM3* we get $13 \times 7 = 91$ textural descriptors per patient.

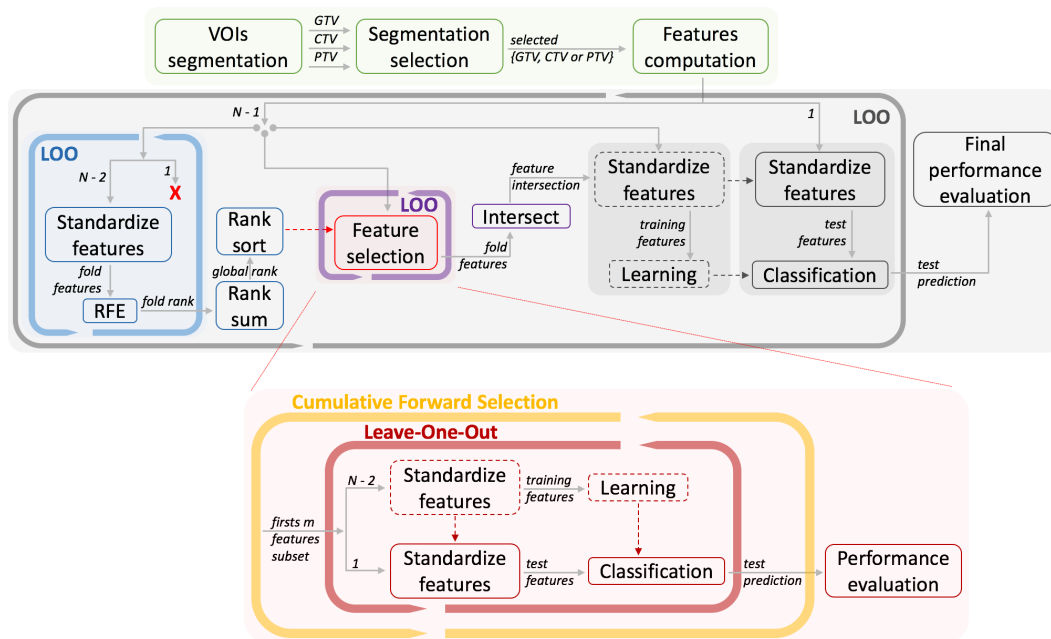


Figure 2. Graphical representation of the method, where the different colours refer to different steps pipeline described in Section 4. List of abbreviations: VOI: Volume of Interest, LOO: Leave-One-Out, RFE: Recursive Feature Elimination, N: number of patients, GTV: Gross Tumour Volume, CTV: Clinical Target Volume, PTV: Planning Target Volume.

Three Orthogonal Planes-Local Binary Patterns: this is a relatively new descriptor in radiomics scenario, and it is one of the possible generalisation of the well-known 2D LBP to the 3D. Now, we will briefly recall the computation of a basic bi-dimensional LBP. If we consider a bi-dimensional image I , then we can compare the intensity I_p of each pixel p with the intensity I_j of all its j th neighbour pixels that lay on a circle centred in p with radius r . If $I_j > I_p$, the j th pixel is set to 1, 0 otherwise. Next, it is possible to process all p 's neighbours in a circular direction, interpreting the sequence of 0s and 1s as a binary string and setting the value of p to the equivalent decimal value. Proceeding as described through all of the pixels of I , we obtain a new image encoding the intensity distribution of each pixel respect to its neighbours that can describe the texture of the original image. The number of patterns for this 2D implementation is 2^P , with P denoting the number of local neighbouring points around the central point. In order to extend a 2D LBP to the 3D environment, a first solution consists in considering a helix neighbourhood of each voxel [42]. However, to avoid the high increase of the computational burden, associated to the very large number of patterns for volume LBP, when P increases, as 2^{3P+2} , we chose another 3D implementation of LBP transformation that was also shown to work fine [42]. It considers the co-occurrence on three orthogonal planes crossing the centre of the analysed volume, as depicted in Figure 3.

It is based on the computation of three 2D LBPs, each of one is derived from each of the three planes; their histograms are then computed and concatenated to obtain a unique histogram for the specific volume. This conspicuously alleviates the computational load, since the number of patterns for TOP-LBP is 3×2^P . Furthermore, in our LBP implementation, we consider two more variants to cope with other two issues of 2D LBP definition. First, we computed rotation invariant LBP, i.e., all of the binary strings obtained as the circular shift of a fundamental string are considered the same. Second, we implemented a uniform version of LBP, i.e., all binary strings containing more than two crossings from 0 to 1 or from 1 to 0 are considered not uniform and coded with a specific string. In our case, setting $P = 8$, we get 48 features by computing first-order measures from the each of the three 2D LBP. The interested readers can find more details about LBPs in [42,50].

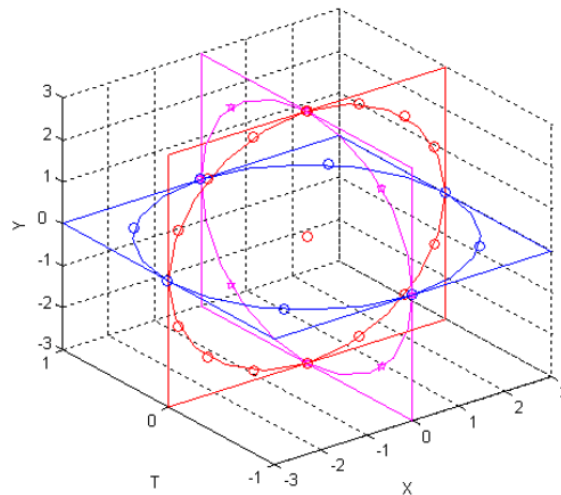


Figure 3. Visual representation of a TOP-LBP example: each ellipsoid shows a planar LBP calculated on a specific orthogonal plane.

4.2. Features Selection and Classification

An important step during a radiomics approach is the features selection step: it helps in reducing the dimensionality of the problem, lowering the curse of dimensionality and the risk of overfitting, and it helps finding the most informative set of descriptors for the problem at hand.

The feature selection pipeline, represented in the blue, purple and red blocks of Figure 2, consists in a wrapper-based approach, which searches and evaluates the best subset of features maximising the performance of a given classification algorithm. To avoid any bias, features are normalised before each step, using a standard scaler, as represented in Figure 2. This approach scales features while using the following equation: $z = (x - \mu) / \sigma$, with μ mean value, σ standard deviation and with x and z as the original feature and the scaled feature, respectively.

Starting from the blue box in Figure 2, the feature set was analysed by the Recursive Feature Elimination (RFE) approach [51], making use of four different classification algorithms, namely AdaBoost, CART Decision Tree (DT), Random Forest (RF), and XGBoost (XGB) [52]. The use of four learners and the availability of three different types of VOIs gave rise to 12 runs of the wrapper, which are listed in Table 2.

The goal of the RFE is to select a feature subgroup, by initially considering all of the features and recursively examining a smaller and smaller dataset according to an assigned feature importance attribute. During this process, the selected classifier is trained on the standardised descriptor set to compute the feature importance attribute based on the accuracy level of the model. The feature with the lowest attribute is excluded from further analysis. This recursive analysis is performed until the best feature is found resulting in a final descriptor ranking.

To avoid any bias during the features selection approach, the RFE step was applied according to a nested leave-one-out (LOO) procedure: indeed, referring to Figure 2, there is an outer grey loop for final performance computation and an inner blue loop for RFE application. This means that this analysis was repeated $N - 1$ times, to obtain a global view of the feature importance. Hence, all of the $N - 1$ rankings obtained were summed up to get a final rank of each descriptor.

From this step on we considered two alternative paths referred to as single ranking and total ranking. On the one hand, in the first path, the ranking procedure was performed independently for each single experiment, i.e., considering a single classifier and a single type of VOI. Straightforwardly, each of the 12 experiments listed in Table 2 uses a specific features rank and, in turn, this permits us to optimise the feature set for each learning algorithm and for each VOI.

On the other hand, in the second path, the wrapper first computes the rank of the features for the GTV, CTV, and PTV; then, such rankings computed using the same classifier on the three volumes

were summed up to gain a global ranking for every learning algorithm. This implies that we finally have four different feature sets, one for each classifier used in the wrapper, regardless of the VOI used. This permitted designing experiments that are comparable within the same classifier.

Turning our attention to the purple and red boxes in Figure 2, these represent the feature evaluation workflow to find the best subset and it was performed independently for both single ranking and total ranking analyses. To find out the best descriptors, we tested different combinations of features subsets whose performance was evaluated according to a given metric. To this goal, all of the features were initially sorted by their rank and, then, the subsets were sequentially inspected using a cumulative forward selection approach (yellow loop). The cumulative forward selection search procedure starts from a subset with one feature with the highest rank position and then incrementally adds descriptors with lower ranks, until the set contains all the descriptors. To avoid any bias, each subset was then evaluated while using a nested LOO configuration, where the inner loop (red LOO) evaluates the best subset maximising the Area Under the ROC Curve (AUC) metric, whereas the outer loop is again the grey LOO for final performance computation. Hence, the output of the purple box is a list of $N - 1$ best subsets found with the nested procedure. We chose as final set the intersection among these intermediate subsets, which represents the radiomic signature of the specific combination of VOI and classifier in Table 2.

Finally, with this set of descriptors, the outer grey leave-one-out computes the ultimate performance of the system.

Table 2. Overview of all possible combinations of volume of interests (VOIs) and learning algorithms.

VOI	Classifier			
	<i>AdaBoost</i>	<i>Decision Tree</i>	<i>Random Forest</i>	<i>XGBoost</i>
<i>GTV</i>	AdaBoost-GTV	DT-GTV	RF-GTV	XGB-GTV
<i>CTV</i>	AdaBoost-CTV	DT-CTV	RF-CTV	XGB-CTV
<i>PTV</i>	AdaBoost-PTV	DT-PTV	RF-PTV	XGB-PTV

4.3. Comparative Analysis

As reported in Section 1, in this work we deepen the analysis comparing the proposed approach with the state-of-the-art and with an approach that leverages on image features automatically computed. Therefore, this last section of the methods introduces the methodology of the competitors.

4.3.1. Comparison with the State-Of-The-Art

As a first point, we compare our approach with the most cited one at the state-of-the-art, represented by the radiomic signature that was proposed by Aerts et al. [19]. This signature includes four features: (i) the “statistics energy” that describes the overall density of the tumour volume, (ii) the “shape compactness” quantifying how compact the tumour shape is, (iii) the “grey level nonuniformity” that measures for intratumour heterogeneity, and (iv) the “grey level nonuniformity HLH” measuring intratumour heterogeneity after decomposing the image in midfrequencies through wavelet analysis. The interested readers can find the formal definition of such measures in the supplementary material of [19], while their implementation is available in Python [53]. We compute this signature on our dataset and, for a fair comparison, all of the tests are performed in LOO fashion for all of the classifier-VOI combinations reported in Table 2.

4.3.2. Comparison with a Deep Learning Approach

Besides the comparison with the handcrafted signature at the state-of-the art, we investigate the possibility to use automatic feature extractors. To this goal, we consider two well-established deep networks, i.e., the AlexNet [54] and the ResNet50 [55], and we train them from scratch on our dataset considering only the CTVs, since they yield the largest performance in the machine learning

pipeline proposed here, as shown in Section 5. On the one hand, the AlexNet has a feature extraction step composed of three consecutive blocks containing two convolutional layers and one max pooling layer. After each convolutional layer, we included a batch normalization layer and a dropout layer, to prevent exploding gradient values and overfitting, respectively. Subsequently, the classification step is designed with two dense layers, one with 256 neurons and the other with one neuron, as it is the output layer. All of the layers implement a LeakyRelu activation function, except for the output that uses a sigmoid for the final classification. On the other hand, the ResNet50 was designed, as reported in [55], except for the first max pooling layer, which was eliminated in order to maximally preserve the information contained in the images, and the output layer that is a dense layer with one neuron with a sigmoid activation function. In both cases, due to the large computational efforts needed, the test procedure was conducted in a 10-fold cross validation, with eight folds as training, one as validation, and one as test. Here, cross validation is preferred to LOO to alleviate the computational burden, while having enough samples in the training and validation sets. In this case, the CNNs classify each slice and then patient label is given by majority voting.

As a further direction of investigation, we also study what happens using the CNNs only as feature extractors that feed the same four classifiers that were used in our approach. In practice, from both networks pretrained on our dataset, we use as feature set the last layer in the feature extraction blocks: this yields to 256 descriptors from the dense layer of AlexNet and to 8192 descriptors from the average pooling layer of ResNet50. These two feature sets separately fed the four classifiers used in the proposed pipeline, which were trained in order to classify each 2D slice. Again, patient label is given by majority voting and the performance are computed in 10-fold cross validation

5. Experimental Results

The main experiment described in Section 4 was applied to all possible combinations of classifiers and VOIs where the features are computed, summarised in Table 2. Furthermore, the experiments are run also using both the single ranking and the total ranking methods for feature selection. The use of four classifiers, three VOIs, and two feature selection ranking approaches results in twenty-four experiments, whose results are presented in Table 3. We measured the following performance metrics: $accuracy = \frac{TP+TN}{P+N}$, the *area under the ROC curve (AUC)*, the $precision = \frac{TP}{TP+FP}$ and the $recall = \frac{TP}{TP+FN}$. Straightforwardly, TP , TN , P , and N stands for the true positive, true negative, total number of positive samples, and total number of negative samples, respectively. Hereinafter, we also use FP and FN to denote the false positive and false negative, respectively. Let us recall that the positive and negative classes correspond dead and alive patients, respectively, as described in Section 3.

In Table 3, the upper panel shows the performance of the twelve combinations of the classifiers and the VOIs when the single ranking method is used, whereas the lower panel reports the performance for the total ranking case. For the single ranking, the accuracy and AUC values range from 60.82% to 83.51% and from 61.06% to 82.78%, respectively. For the total ranking, the accuracy ranges from 44.33% to 77.32%, and the AUC ranges from 44.43% to 75.96%. The subsets of features selected in each combination range from two to nine descriptors for the single ranking and from two to 20 descriptors for the total ranking. Note that, in the total ranking case, we obtain a worse range of accuracy, AUC values and larger subsets of selected feature in comparison to the single ranking. This could be expected, since the single ranking approach provides a best feature subset optimised for both the classification algorithm and the considered VOI.

A more detailed analysis of Table 3 suggests us other three observations. The first is that the performance on GTV in the single ranking is stable across the four classification algorithms. The same does not happen for the total ranking case. We deem this indicates that a system built on this VOI is particularly stable when a volume-specific optimization occurs for the best feature set selection. Secondly, the overall largest performance was obtained combining the classifier Adaboost with the CTV in the single ranking procedure. Compared to this result, the best result for the total ranking procedure was obtained combining the Random Forest Classifier with, again, the CTV. Such best

results for single and total rankings are highlighted in bold in Table 3. As a third observation, we notice that, although both the best experiments exploit the CTV, the largest performance are given by the single ranking approach, where the best feature set was customized on the specific VOI analysed.

Table 3. System performance in all possible combinations. The upper panel refers to the single ranking strategy for feature selection, whereas the lower panel shows the results considering the total ranking option.

<i>Single</i>	AdaBoost			Decision Tree			Random Forest			XGBoost		
	Ranking	GTV	CTV	PTV	GTV	CTV	PTV	GTV	CTV	PTV	GTV	CTV
Accuracy	70.10	83.51	71.13	70.10	60.82	63.92	73.20	78.35	77.32	68.04	71.13	76.29
AUC	70.13	82.78	70.11	70.13	61.06	64.09	71.81	77.29	76.54	66.70	70.11	75.02
Precision	69.81	90.57	81.13	69.81	58.49	62.26	86.79	88.68	84.91	67.19	70.49	73.44
Recall	74.00	81.36	70.49	74.00	65.96	68.75	70.77	75.81	76.27	81.13	81.13	88.68
<i>Total</i>	AdaBoost			Decision Tree			Random Forest			XGBoost		
	Ranking	GTV	CTV	PTV	GTV	CTV	PTV	GTV	CTV	PTV	GTV	CTV
Accuracy	64.95	72.16	68.04	44.33	58.76	65.98	75.26	77.32	74.23	64.95	72.16	74.23
AUC	64.84	71.83	67.47	44.43	59.18	65.59	73.89	75.96	73.33	62.52	70.09	72.94
Precision	66.04	75.47	73.58	43.40	54.72	69.81	88.68	90.57	83.02	62.67	68.06	71.88
Recall	68.63	74.07	69.64	48.94	64.44	68.52	72.31	73.85	73.33	88.68	92.45	86.79

We also run the ANOVA test among the VOIs both for the results attained by single and total ranking, whatever the classification paradigm used. When we get a significant result, i.e., at least one of the groups tested differs from the other groups, we proceed with Least Significant Difference (LSD) that detects the statistical differences among a group of results. In the case of results provided by the single ranking, we obtain a p-value equal to 0.0334, suggesting that the performance achieved using the different VOIs are not the same. The LSD tests show that the results of CTV+Adaboost are significant different ($p < 0.1$) from the others in all cases, except for CTV+Random Forest, PTV+Random Forest, and PTV+XGBoost. In the case of the results provided by the total ranking, the p-value is 1.586×10^{-5} , suggesting again that there are differences between the VOIs. The method with the best AUC, i.e., CTV+Random Forest, shows performance statistically different from all the others except for CTV+AdaBoost, PTV+AdaBoost, GTV+Random Forest, PTV+Random Forest, CTV+XGBoost, and PTV+XGBoost.

Let us now turn the attention to the radiomics signature underlying the best case, i.e., CTV+Adaboost in single ranking: as shown in the first column of Table 4 it is composed of eight features belonging to the LBP set, and one belonging to the GLCM set. In detail, such features are: “uniform 3D LBP kurtosis”, which estimates the shape of the distribution of all patterns discharging the noisy patterns, “3D LBP energy”, which represents the homogeneity of the patterns, “rotation invariant 3D LBP absolute maximum”, which is the absolute maximum value among the patterns, regardless of their rotation, “3D LBP energy around the absolute maximum”, which estimates the homogeneity of patterns around the absolute maximum value, “rotation invariant 3D LBP energy”, which is the homogeneity of the patterns without considering their rotation, “uniform 3D LBP energy around the relative maximum”, which estimates the homogeneity of pattern around a relative maximum value discharging the noisy patterns, “uniform 3D LBP entropy”, which is a measure of the variety of patterns discharging the noisy ones, and “uniform 3D LBP skewness”, which measures the asymmetry of distribution of all patterns discharging noisy patterns, and “inverse GLCM in direction $(-1,-1,0)$ ”, which measures the inverse of grey level attenuation in the ROI over the direction $(-1,-1,0)$.

Turning our attention to the newly introduced set of features, i.e., 3D LBPs, and still focusing on the single ranking approach, these descriptors were always chosen in the best feature subsets, regardless of the classifier and the VOIs under consideration. For the sake of brevity, we hereby do not report the full list of features chosen in each of the 24 experiments. However, to provide a synthetic analysis of the best feature sets, the second and the third columns of Table 4 show how many times each feature in the first column (i.e., the radiomic signature of the best case) was included in the best feature set using the

Adaboost with the GTV and PTV (second column) and the DT, RF, and XGBoost with the CTV (third column). Note that finding zeros in the occurrences of Table 4 does not exclude the presence of other LBP features in the best subsets besides those chosen in the experiment with largest performance.

Table 4. Best radiomic signatures. The first column lists the feature selected in the best performance. The other two columns indicate the number of times that each descriptor was selected in the best feature set. Abbreviations: U—uniform, RI—rotation invariant.

Features Selected for AdaBoost CTV	#Occurrences in AdaBoost GTV and PTV	#Occurrences in DT, RF and XGBoost with CTV
U 3D LBP kurtosis	0	3
3D LBP energy	0	1
RI 3D LBP maxAss	0	0
3D LBP energy around maxAss	0	1
RI 3D LBP energy	0	1
U 3D LBP energy around maxRel	0	1
U 3D LBP entropy	1	3
U 3D LBP skewness	0	3
inverse GLCM (−1,−1,0)	0	1

To deepen the discussion on the effectiveness of 3D LBPs, we excluded them from each signature selected by our approach and train again each learner. The results are reported in Table 5, where “-” indicates that the classifier was not trained, since the best set included only 3D LBP descriptors. When comparing these results with Table 3 we notice that there is not an evident pattern in the performance differences between the two experiments. Despite that, the use of 3D LBPs improves the largest accuracy obtained: indeed, without them, the overall best accuracy score decreases from 83.51% to 76.29% and from 77.32% to 75.26% for the single ranking and the total ranking, respectively. This trend is also manifested by the other performance measures reported for the best cases, indicating that LBPs actually give an important contribution to the classification task at hand. Furthermore, we statistically pairwise compared the results reported in Table 3 with those in Table 5 using the Wilcoxon signed rank test. In the case of single ranking, for AdaBoost+CTV, DT+CTV, RF+CTV, RF+PTV, and XGBoost+GTV the use of 3D LBPs provides performance larger and statistically different from those attained excluding these descriptors ($p < 0.1$). In the case of total ranking, at the same level of confidence, the results of the test show that the use of 3D LBP provides larger and statistically different performance for AdaBoost+PTV and XGBoost+CTV.

Table 5. System performance excluding LBP descriptors from the final feature subset in all possible combinations of Table 2 for single ranking (upper panel) and total ranking (lower panel). The notation ‘-’ indicates that no feature was left in the final set after excluding LBP descriptors. The largest performance is highlighted in bold.

Single Ranking	AdaBoost			Decision Tree			Random Forest			XGBoost		
	GTV	CTV	PTV	GTV	CTV	PTV	GTV	CTV	PTV	GTV	CTV	PTV
Accuracy	76.29	69.07	62.89	71.13	48.45	72.16	70.10	69.07	69.07	55.67	69.07	71.13
AUC	75.99	68.22	62.75	70.50	47.81	72.02	68.01	66.87	68.42	51.71	68.22	69.14
Precision	77.78	69.49	66.67	71.93	52.73	75.00	66.67	65.75	70.18	55.56	69.49	67.61
Recall	79.25	77.36	64.15	77.36	54.72	73.58	90.57	90.57	75.47	94.34	77.36	90.57
Total Ranking	AdaBoost			Decision Tree			Random Forest			XGBoost		
	GTV	CTV	PTV	GTV	CTV	PTV	GTV	CTV	PTV	GTV	CTV	PTV
Accuracy	74.23	73.20	52.58	61.86	-	-	75.26	72.16	71.13	56.70	32.99	67.01
AUC	73.52	72.77	52.17	61.23	-	-	74.08	70.28	68.95	54.97	36.36	64.02
Precision	74.14	74.55	56.60	64.29	-	-	73.02	68.57	67.12	58.21	0	62.96
Recall	81.13	77.36	56.60	67.92	-	-	86.79	90.57	92.45	73.58	0	96.23

5.1. Comparison with The-State-Of-The-Art

The results were compared against the signature presented by Aerts et al. [19] for OS prediction in NSCLS, as introduced in Section 4. Further, to be one of the most cited works in the state-of-the-art, such a radiomic signature was also used by Lambin et al. [17] and Kwan et al. [56].

Table 6 reports the performance computed with Aerts et al.'s signature on our dataset: the largest AUC value was obtained considering the CTV with Random Forest as classifier, and it is equal to 76.92%. Using the same VOI our signature gets an AUC equal to 82.78%, which suggests that the inclusion of the 3D LBP descriptors boosts the discrimination power and leads to larger performance. Despite the fact that the best results of Tables 3 and 6 are obtained with different classifiers, comparing the corresponding columns of the two tables further validates the previous assertion. Also note that the best combination in Table 6 is based on a Random Forest classifier with the CTV, which is also one of two best configurations in our experiments: this confirms our results that the CTV is the most informative volume of interest.

Table 6. Performance with the signature presented by Aerts et al. [19] combined with the classification algorithms used here. The largest performance is highlighted in bold.

	Adaboost			Decision Tree			Random Forest			XGBoost		
	<i>GTV</i>	<i>CTV</i>	<i>PTV</i>	<i>GTV</i>	<i>CTV</i>	<i>PTV</i>	<i>GTV</i>	<i>CTV</i>	<i>PTV</i>	<i>GTV</i>	<i>CTV</i>	<i>PTV</i>
Accuracy	63.16	67.37	60.00	63.16	66.32	63.16	73.68	77.89	73.68	46.32	46.32	46.32
AUC	62.88	67.42	59.94	63.13	66.13	62.88	72.84	76.92	72.68	50.00	50.00	50.00
Precision	65.38	70.83	63.27	67.39	68.63	65.38	71.67	74.19	70.97	0	0	0
Recall	66.67	66.67	60.78	60.78	68.63	66.67	84.31	90.20	86.27	0	0	0

5.2. Comparison with a Deep Learning Approach

We conducted a comparison with two deep learning approaches to further assess the power of the handcrafted features, as described in Section 4. Hence, we performed two experiments: the first tests the AlexNet and the ResNet50 as a whole classification approach, whereas the second tests the two deep networks as feature extractors feeding the same four classifiers used herein-before. All of the results are summarised in Tables 7 and 8 and, in general, they show that the most performing network is the ResNet50.

In the first experiment, this deep network achieves an AUC equal to 76.49% and an accuracy equal to 71.11%. Furthermore, the performance differences between the ResNet50 and the AlexNet are statistically significant according to the Wilcoxon signed rank test ($p = 0.0396$).

In the second experiment, when the networks are used as feature extractors, we find that the automatic features they compute combined with the Adaboost classifier provide accuracy and AUC equal to 73.33% and 73.42%, respectively. Furthermore, using this learner, the performance differences using the features computed by the AlexNet and by ResNet50 are statistically significant with $p < 0.1$ according to the Wilcoxon signed rank test. We do not find statistical differences in the pairwise comparison between the automatic features while using the other three classifiers.

As it is evident, these values were lower than those achieved by the proposed approach (accuracy = 83.51% and AUC = 82.78%), showing the successful discriminative power of hand-crafted features. We also assess this difference using again the Wilcoxon signed rank test, comparing our proposal (AdaBoost+CTV in Table 3) with the best one provided by deep learning approach (ResNet50 features in Table 7). The results of such two approaches are statistically different with $p < 0.05$. A possible reason for this finding could be the size of our dataset: indeed, this could hinder the power of deep approaches, leading to the superiority of handcrafted features with respect to measures automatically learned.

Table 7. Performance of the deep features extracted with two state-of-the-art deep neural networks. The results are computed evaluating the four classification algorithms that were used in the proposed pipeline in 10-fold cross validation. The largest performance is highlighted in bold.

	ResNet50 features				AlexNet features			
	AdaBoost	DT	RF	XGB	AdaBoost	DT	RF	XGB
Accuracy	73.33	66.67	70.0	71.11	63.33	66.67	63.33	70.0
AUC	73.42	66.08	69.5	71.17	63.08	67.67	64.00	70.67
Precision	72.81	68.02	68.12	69.94	63.86	70.52	63.89	72.16
Recall	85.67	82.17	89.67	89.67	78.17	79.00	89.17	89.17

Table 8. Results of the AlexNet and the ResNet50 trained from scratch on our dataset.

	Accuracy	AUC	Precision	Recall
ResNet50	71.11	76.49	69.49	83.67
AlexNet	57.78	64.37	60.38	65.31

6. Conclusions

In this manuscript, we propose a radiomics approach to predict the Overall Survival in a cohort of 97 patients suffering from locally advanced non-small cell lung cancer. NSCLC has been already studied in few other works and, differently from the literature, here we introduce two novel aspects which have not been analysed in this way so far. The first is the analysis of three different VOIs commonly used in clinical practice and segmented on CT images routinely collected. The second, but not secondary, is the introduction of 3D Local Binary Pattern descriptors, already used in computer vision problems. The both constitute novel and challenging directions of this research.

Among the different combinations of classification algorithms and VOIs, the best accuracy of proposed approach is equal to 83.51%, achieved using the Clinical Target Volume, the Adaboost learner, and nine features selected from the pool. The fact that the VOI providing the best results is the CTV confirms our initial hypothesis that there is information in the surroundings of the visible tumour and that this is important to predict patients' OS. Furthermore, the LBPs features proved to be determinant, improving the classification performance in comparison with respect to other texture measures.

The promising results described so far suggest three future directions worthy of investigation. First, the system performance and its variability could be explored for small variations of the segmentations of the target tumour volume, considering not only the GTV, but also the CTV and PTV. This will permit us to assess how important it is to get a very accurate automatic segmentation. Second, we could further explore the potential of LBPs, validating such descriptors on an external dataset, so to confirm their stability and effectiveness. Third, we are working to enlarge the size of the dataset and to extend the research in the deep learning area, even when considering deep approaches to artificially enlarge the size of the training set, such as the Generative Adversarial Networks.

Author Contributions: Conceptualization, E.C., R.S., S.R. and P.S.; Methodology, E.C., N.C.D., L.T., R.S. and P.S.; Software, E.C., N.C.D., L.T., R.S. and P.S.; Validation, E.C., N.C.D., L.T., R.S. and P.S.; Formal analysis, E.C., N.C.D., L.T., R.S. and P.S.; Resources, C.G., M.F., A.C. and S.R.; Data Curation, C.G., M.F., A.C. and S.R.; Writing—original draft preparation, N.C.D., R.S., E.C., L.T., and P.S.; Writing—review and editing, N.C.D., R.S., E.C., L.T., C.G., M.F., A.C., G.I., S.R., and P.S.; Visualization, E.C., N.C.D., L.T., R.S. and P.S.; Supervision, S.R., G.I. and P.S.; Funding acquisition, S.R. and P.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially funded by Universita' Campus Bio-Medico di Roma under the programme "University Strategic Projects 2018 call" within the project "a CoLIaborative multi-sources Radiopathomics approach for personalised Oncology in non-small cell lung cancer (CLARO)".

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A. Haralick Features

This appendix presents in Table A1 the formal definition of Haralick features listed in Section 4. The content of such tables adopts the following notation:

- G is the number of grey levels in the image I ;
- g_i and g_j denotes two grey level values, with i and j lying in $[0, G - 1]$;
- \mathbf{d} is a displacement vector;
- $p(g_i, g_j, \mathbf{d})$ denotes the grey level co-occurrence matrix (GLCM) of I ;
- $\mu_{p(g_j, \mathbf{d})} = \sum_{i=0}^{G-1} p(g_i, g_j, \mathbf{d})$ and $\mu_{p(g_i, \mathbf{d})} = \sum_{j=0}^{G-1} p(g_i, g_j, \mathbf{d})$ denote the mean value of the one-dimensional marginal distributions of the GLCM;
- $\partial(g_i, g_j) = (g_i - g_j)^2$ is a measure of dissimilarity.

Table A1. Definition of the Haralick measures.

Feature	Definition
Autocorrelation	$\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} [p(g_i, g_j, \mathbf{d}) - \mu_{p(g_i, \mathbf{d})} \mu_{p(g_j, \mathbf{d})}]$
Homogeneity	$\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{p(g_i, g_j, \mathbf{d})}{1 + \partial(g_i, g_j)}$ for $g_i \neq g_j$
Entropy	$-\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} p(g_i, g_j, \mathbf{d}) \cdot \log[p(g_i, g_j, \mathbf{d})]$, for $p(g_i, g_j, \mathbf{d}) > 0$
Energy	$\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} p(g_i, g_j, \mathbf{d})^2$
Covariance	$\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (g_i - \mu_{p(g_i, \mathbf{d})})(g_j - \mu_{p(g_j, \mathbf{d})})p(g_i, g_j, \mathbf{d})$
Inertia	$\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \partial(g_i, g_j) p(g_i, g_j, \mathbf{d})$
Absolute contrast	$\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} g_i - g_j \cdot p(g_i, g_j, \mathbf{d})$

References

1. Word Health Organisation. LUNG. Available online: <https://gco.iarc.fr/today/data/factsheets/cancers/15-Lung-fact-sheet.pdf> (accessed on 24 July 2020).
2. Molina, J.R.; Yang, P.; Cassivi, S.D.; Schild, S.E.; Adjei, A.A. Non-small cell lung cancer: Epidemiology, risk factors, treatment, and survivorship. *Mayo Clin. Proc.* **2008**, *83*, 584–594. [CrossRef]
3. Cancer.net. Lung Cancer—Non-Small Cell: Statistics. Available online: <https://www.cancer.net/cancer-types/lung-cancer-non-small-cell/statistics> (accessed on 28 July 2020).
4. Lemjabbar-Alaoui, H.; Hassan, O.U.; Yang, Y.W.; Buchanan, P. Lung cancer: Biology and treatment options. *Biochim. Biophys. Acta (BBA)-Rev. Cancer* **2015**, *1856*, 189–210. [CrossRef]
5. Squibb, B.M. Lung Cancer by the Numbers. Available online: <https://www.bms.com/assets/bms/us/en-us/pdf/Disease-State-Info/LUNG-CANCER-BY-THE-NUMBERS.pdf> (accessed on 26 July 2020).
6. Antonia, S.J.; Villegas, A.; Daniel, D.; Vicente, D.; Murakami, S.; Hui, R.; Kurata, T.; Chiappori, A.; Lee, K.H.; de Wit, M.; et al. Overall survival with durvalumab after chemoradiotherapy in stage III NSCLC. *N. Engl. J. Med.* **2018**, *379*, 2342–2350. [CrossRef]
7. ESMO. ESMO Interactive Guidelines, Lung & Chest Tumours 2018. Available online: <http://interactiveguidelines.esmo.org/esmo-web-app/home/> (accessed on 19 July 2020).
8. Antonia, S.J.; Villegas, A.; Daniel, D.; Vicente, D.; Murakami, S.; Hui, R.; Yokoi, T.; Chiappori, A.; Lee, K.H.; de Wit, M.; et al. Durvalumab after chemoradiotherapy in stage III non-small-cell lung cancer. *N. Engl. J. Med.* **2017**, *377*, 1919–1929. [CrossRef] [PubMed]
9. Curran, W.J., Jr.; Paulus, R.; Langer, C.J.; Komaki, R.; Lee, J.S.; Hauser, S.; Movsas, B.; Wasserman, T.; Rosenthal, S.A.; Gore, E.; et al. Sequential vs concurrent chemoradiation for stage III non-small cell lung cancer: Randomized phase III trial RTOG 9410. *J. Natl. Cancer Inst.* **2011**, *103*, 1452–1460. [CrossRef] [PubMed]

10. Kim, T.Y.; Yang, S.H.; Lee, S.H.; Park, Y.-S.; Im, Y.H.; Kang, W.-K.; Ha, S.H.; Park, C.I.; Heo, D. S.; Bang, Y.-J.; et al. A phase III randomized trial of combined chemoradiotherapy versus radiotherapy alone in locally advanced non-small-cell lung cancer. *Am. J. Clin. Oncol.* **2002**, *25*, 238–243. [[CrossRef](#)]
11. Ito, H.; Matsuo, Y.; Ohtsu, S.; Nishimura, T.; Terada, Y.; Sakamoto, T.; Mizowaki, T. Impact of histology on patterns of failure and clinical outcomes in patients treated with definitive chemoradiotherapy for locally advanced non-small cell lung cancer. *Int. J. Clin. Oncol.* **2020**, *25*, 274–281. [[CrossRef](#)] [[PubMed](#)]
12. Vasilakos, A.V.; Tang, Y.; Yao, Y.; Yao, Y. Neural networks for computer-aided diagnosis in medicine: A review. *Neurocomputing* **2016**, *216*, 700–708.
13. Castaneda, C.; Nalley, K.; Mannion, C.; Bhattacharyya, P.; Blake, P.; Pecora, A.; Goy, A.; Suh, K.S. Clinical decision support systems for improving diagnostic accuracy and achieving precision medicine. *J. Clin. Bioinform.* **2015**, *5*, 4. [[CrossRef](#)]
14. Hu, W.; Chen, Y.; Wang, S.; Chen, H.; Liu, Z.; Zeng, C.; Zhang, H.; Liu, Z. Clinical-morphological features and outcomes of lupus podocytopathy. *Clin. J. Am. Soc. Nephrol.* **2016**, *11*, 585–592. [[CrossRef](#)]
15. Silvestri, L.; Pacisopi, M.; Soda, P.; Biamonte, F.; Iannello, G.; Frasconi, P.; Pavone, F.S. Quantitative neuroanatomy of all Purkinje cells with light sheet microscopy and high-throughput image analysis. *Front. Neuroanat.* **2015**, *9*, 68. [[CrossRef](#)] [[PubMed](#)]
16. Percannella, G.; Soda, P.; Vento, M. Mitotic HEp-2 cells recognition under class skew. In *International Conference on Image Analysis and Processing*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 353–362.
17. Lambin, P.; Rios-Velazquez, E.; Leijenaar, R.; Carvalho, S.; Van Stiphout, R.G.P.M.; Granton, P.; Zegers, C.M.L.; Gillies, R.; Boellard, R.; Dekker, A.; et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur. J. Cancer* **2012**, *48*, 441–446. [[CrossRef](#)] [[PubMed](#)]
18. Kumar, V.; Gu, Y.; Basu, S.; Berglund, A.; Eschrich, S.A.; Schabath, M.B.; Forster, K.; Aerts, H.J.; Dekker, A.; Fenstermacher, D.; et al. Radiomics: The process and the challenges. *Magn. Reson. Imaging* **2012**, *30*, 1234–1248. [[CrossRef](#)] [[PubMed](#)]
19. Aerts, H.J.; Velazquez, E.R.; Leijenaar, R.T.H.; Parmar, C.; Grossmann, P.; Carvalho, S.; Bussink, J.; Monshouwer, R.; Haibe-Kains, B.; Rietveld, D.; et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* **2014**, *5*, 1–9. [[CrossRef](#)]
20. Parmar, C.; Grossmann, P.; Bussink, J.; Lambin, P.; Aerts, H.J. Machine learning methods for quantitative radiomic biomarkers. *Sci. Rep.* **2015**, *5*, 13087. [[CrossRef](#)]
21. Li, H.; Zhu, Y.; Burnside, E.S.; Huang, E.; Drukker, K.; Hoadley, K.A.; Fan, C.; Conzen, S.D.; Zuley, M.; Net, J.M. et al. Quantitative MRI radiomics in the prediction of molecular classifications of breast cancer subtypes in the TCGA/TCIA data set. *NPJ Breast Cancer* **2016**, *2*, 16012. [[CrossRef](#)]
22. D’Amico, N.C.; Merone, M.; Sicilia, R.; Cordelli, E.; D’Antoni, F.; Bossi Zanetti, I.; Valbusa, G.; Grossi, E.; Beltramo, G.; Fazzini, D.; et al. Tackling imbalance radiomics in acoustic neuroma. *Int. J. Data Min. Bioinform.* **2019**, *22*, 365–388.
23. Sicilia, R.; Cordelli, E.; Merone, M.; Luperto, E.; Papalia, R.; Iannello, G.; Soda, P. Early radiomic experiences in classifying prostate cancer aggressiveness using 3D local binary patterns. In *Proceedings of the 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*, Cordoba, Spain, 5–7 June 2019; pp. 355–360.
24. Li, Q.; Bai, H.; Chen, Y.; Sun, Q.; Liu, L.; Zhou, S.; Wang, G.; Liang, C.; Li, Z.-C. A fully-automatic multiparametric radiomics model: Towards reproducible and prognostic imaging signature for prediction of overall survival in glioblastoma multiforme. *Sci. Rep.* **2017**, *7*, 1–9. [[CrossRef](#)]
25. Kickingeder, P.; Burth, S.; Wick, A.; Götz, M.; Eidel, O.; Schlemmer, H.-P.; Maier-Hein, K.H.; Wick, W.; Bendszus, M.; Radbruch, A.; et al. Radiomic profiling of glioblastoma: Identifying an imaging predictor of patient survival with improved performance over established clinical and radiologic risk models. *Radiology* **2016**, *280*, 880–889. [[CrossRef](#)]
26. Lao, J.; Chen, Y.; Li, Z.-C.; Li, Q.; Zhang, J.; Liu, J.; Zhai, G. A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. *Sci. Rep.* **2017**, *7*, 1–8. [[CrossRef](#)]
27. Larue, R.T.; Klaassen, R.; Jochems, A.; Leijenaar, R.T.; Hulshof, M.C.; van Berge Henegouwen, M.I.; Schreurs, W.M.; Sosef, M.N.; van Elmpt, W.; van Laarhoven, H.W.; et al. Pre-treatment CT radiomics to predict 3-year overall survival following chemoradiotherapy of esophageal cancer. *Acta Oncol.* **2018**, *57*, 1475–1481. [[CrossRef](#)]

28. Jiang, Y.; Chen, C.; Xie, J.; Wang, W.; Zha, X.; Lv, W.; Chen, H.; Hu, Y.; Li, T.; Yu, J.; et al. Radiomics signature of computed tomography imaging for prediction of survival and chemotherapeutic benefits in gastric cancer. *EBioMedicine* **2018**, *36*, 171–182. [[CrossRef](#)] [[PubMed](#)]
29. Spraker, M.B.; Wootton, L.S.; Hippe, D.S.; Ball, K.C.; Peeken, J.C.; Macomber, M.W.; Chapman, T.R.; Hoff, M.N.; Kim, E.Y.; Pollack, S.M.; et al. MRI radiomic features are independently associated with overall survival in soft tissue sarcoma. *Adv. Radiat. Oncol.* **2019**, *4*, 413–421. [[CrossRef](#)] [[PubMed](#)]
30. Huynh, E.; Coroller, T.P.; Narayan, V.; Agrawal, V.; Hou, Y.; Romano, J.; Franco, I.; Mak, R.H.; Aerts, H.J. CT-based radiomic analysis of stereotactic body radiation therapy patients with lung cancer. *Radiother. Oncol.* **2016**, *120*, 258–266. [[CrossRef](#)]
31. Ramella, S.; Fiore, M.; Greco, C.; Cordelli, E.; Sicilia, R.; Merone, M.; Molfese, E.; Miele, M.; Cornacchione, P.; Ippolito, E.; et al. A radiomic approach for adaptive radiotherapy in non-small cell lung cancer patients. *PLoS ONE* **2018**, *13*, e0207455. [[CrossRef](#)]
32. Homayounieh, F.; Ebrahimian, S.; Babaei, R.; Karimi Mobin, H.; Zhang, E.; Bizzo, B.C.; Mohseni, I.; Digumarthy, S.R.; Kalra, M.K. CT Radiomics, Radiologists and Clinical Information in Predicting Outcome of Patients with COVID-19 Pneumonia. *Radiol. Cardiothorac. Imaging* **2020**, *2*, e200322. [[CrossRef](#)]
33. Fang, X.; Li, X.; Bian, Y.; Ji, X.; Lu, J. Radiomics nomogram for the prediction of 2019 novel coronavirus pneumonia caused by SARS-CoV-2. *Eur. Radiol.* **2020**, 1–14. [[CrossRef](#)] [[PubMed](#)]
34. Di Noto, T.; von Spiczak, J.; Mannil, M.; Gantert, E.; Soda, P.; Manka, R.; Alkadhi, H. Radiomics for Distinguishing Myocardial Infarction from Myocarditis at Late Gadolinium Enhancement at MRI: Comparison with Subjective Visual Analysis. *Radiol. Cardiothorac. Imaging* **2019**, *1*, e180026. [[CrossRef](#)]
35. Niu, L.; Zhou, X.; Duan, C.; Zhao, J.; Sui, Q.; Liu, X.; Zhang, X. Differentiation researches on the meningioma subtypes by radiomics from contrast-enhanced magnetic resonance imaging: A preliminary study. *World Neurosurg.* **2019**, *126*, e646–e652. [[CrossRef](#)]
36. Yanling, W.; Duo, G.; Zuojun, G.; Zhongqiang, S.; Yankai, W.; Shan, L.; Hongying, C. Radiomics Nomogram Analyses for Differentiating Pneumonia and Acute Paraquat Lung Injury. *Sci. Rep.* **2019**, *9*, 1–9. [[CrossRef](#)]
37. Rahmim, A.; Huang, P.; Shenkov, N.; Fotouhi, S.; Davoodi-Bojd, E.; Lu, L.; Mari, Z.; Soltanian-Zadeh, H.; Sossi, V. Improved prediction of outcome in Parkinson’s disease using radiomics analysis of longitudinal DAT SPECT images. *NeuroImage Clin.* **2017**, *16*, 539–544. [[CrossRef](#)] [[PubMed](#)]
38. Fave, X.; Zhang, L.; Yang, J.; Mackin, D.; Balter, P.; Gomez, D.; Followill, D.; Jones, A.K.; Stingo, F.; Liao, Z.; et al. Delta-radiomics features for the prediction of patient outcomes in non-small cell lung cancer. *Sci. Rep.* **2017**, *7*, 1–11. [[CrossRef](#)] [[PubMed](#)]
39. Li, H.; Galperin-Aizenberg, M.; Pryma, D.; Simone, C.B., II; Fan, Y. Unsupervised machine learning of radiomic features for predicting treatment response and overall survival of early stage non-small cell lung cancer patients treated with stereotactic body radiation therapy. *Radiother. Oncol.* **2018**, *129*, 218–226. [[CrossRef](#)] [[PubMed](#)]
40. Scrivener, M.; De Jong, E.E.C.; van Timmeren, J.E.; Pieters, T.; Ghaye, B.; Geets, X. Radiomics applied to lung cancer: A review. *Transl. Cancer Res.* **2016**, *5*, 398–409. [[CrossRef](#)]
41. Thawani, R.; McLane, M.; Beig, N.; Ghose, S.; Prasanna, P.; Velcheti, V.; Madabhushi, A. Radiomics and radiogenomics in lung cancer: A review for the clinician. *Lung Cancer* **2018**, *115*, 34–41. [[CrossRef](#)]
42. Zhao, G.; Pietikainen, M. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 915–928. [[CrossRef](#)]
43. Burnet, N.G.; Thomas, S.J.; Burton, K.E.; Jefferies, S.J. Defining the tumour and target volumes for radiotherapy. *Cancer Imaging* **2004**, *4*, 153. [[CrossRef](#)]
44. Ramella, S.; Fiore, M.; Silipigni, S.; Zappa, M.C.; Jaus, M.; Alberti, A.M.; Matteucci, P.; Molfese, E.; Cornacchione, P.; Greco, C.; et al. Local control and toxicity of adaptive radiotherapy using weekly CT imaging: Results from the LARTIA trial in stage III NSCLC. *J. Thorac. Oncol.* **2017**, *12*, 1122–1130. [[CrossRef](#)]
45. Schimek-Jasch, T.; Troost, E.G.; Rücker, G.; Prokic, V.; Avlar, M.; Duncker-Rohr, V.; Mix, M.; Doll, C.; Grosu, A.L.; Nestle, U. A teaching intervention in a contouring dummy run improved target volume delineation in locally advanced non-small cell lung cancer. *Strahlenther. Und Onkol.* **2015**, *191*, 525–533. [[CrossRef](#)]

46. Giraud, P.; Elles, S.; Helfre, S.; De Rycke, Y.; Servois, V.; Carette, M.F.; Alzieu, C.; Bondiau, P.Y.; Dubray, B.; Touboul, E. Conformal radiotherapy for lung cancer: Different delineation of the gross tumor volume (GTV) by radiologists and radiation oncologists. *Radiother. Oncol.* **2002**, *62*, 27–36. [[CrossRef](#)]
47. Nestle, U.; De Ruyscher, D.; Ricardi, U.; Geets, X.; Belderbos, J.; Pöttgen, C.; Dziadiuszko, R.; Peeters, S.; Lievens, Y.; Hurkmans, C.; et al. ESTRO ACROP guidelines for target volume definition in the treatment of locally advanced non-small cell lung cancer. *Radiother. Oncol.* **2018**, *127*, 1–5. [[CrossRef](#)] [[PubMed](#)]
48. Sebastian V, B.; Unnikrishnan, A.; Balakrishnan, K. Gray level co-occurrence matrices: Generalisation and some new features. *arXiv* **2012**, arXiv:1205.4831.
49. Haralick, R.M.; Shanmugam, K.; Dinstein, I. Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* **1973**, *6*, 610–621. [[CrossRef](#)]
50. Ojala, T.; Pietikainen, M.; Maenpaa, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 971–987. [[CrossRef](#)]
51. Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene selection for cancer classification using support vector machines. *Mach. Learn.* **2002**, *46*, 389–422. [[CrossRef](#)]
52. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
53. Van Griethuysen, J.J.M.; Fedorov, A.; Parmar, C.; Hosny, A.; Aucoin, N.; Narayan, V.; Beets-Tan, R.G.H.; Fillion-Robin, J.-C.; Pieper, S.; Aerts, H.J. Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* **2017**, *77*, e104–e107. [[CrossRef](#)]
54. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet Classification with Deep Convolutional Neural Networks. In Proceedings of the 2012 Annual Conference on Neural Information Processing Systems (NIPS), Lake Tahoe, NV, USA, 3–8 December 2012; pp. 1097–1105.
55. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
56. Kwan, J.Y.Y.; Su, J.; Huang, S.H.; Ghoraie, L.S.; Xu, W.; Chan, B.; Yip, K.W.; Giuliani, M.; Bayley, A.; Kim, J.; et al. Radiomic biomarkers to refine risk models for distant metastasis in HPV-related oropharyngeal carcinoma. *Int. J. Radiat. Oncol. Biol. Phys.* **2018**, *102*, 1107–1116. [[CrossRef](#)] [[PubMed](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).