

Review

# Machine Learning and Radiomics Applications in Esophageal Cancers Using Non-Invasive Imaging Methods—A Critical Review of Literature

Chen-Yi Xie <sup>1</sup>, Chun-Lap Pang <sup>2,3</sup>, Benjamin Chan <sup>4</sup>, Emily Yuen-Yuen Wong <sup>4</sup>, Qi Dou <sup>5</sup>  
and Varut Vardhanabhuti <sup>1,\*</sup>

- <sup>1</sup> Department of Diagnostic Radiology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China; cheniyx@hku.hk
- <sup>2</sup> Department of Radiology, The Christies' Hospital, Manchester M20 4BX, UK; c.pang@doctors.org.uk
- <sup>3</sup> Division of Dentistry, School of Medical Sciences, University of Manchester, Manchester M15 6FH, UK
- <sup>4</sup> Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China; chanbenj@connect.hku.hk (B.C.); emily372@connect.hku.hk (E.Y.-Y.W.)
- <sup>5</sup> Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, China; qidou@cuhk.edu.hk
- \* Correspondence: varv@hku.hk; Tel.: +(852)-22554524



**Citation:** Xie, C.-Y.; Pang, C.-L.; Chan, B.; Wong, E.Y.-Y.; Dou, Q.; Vardhanabhuti, V. Machine Learning and Radiomics Applications in Esophageal Cancers Using Non-Invasive Imaging Methods—A Critical Review of Literature. *Cancers* **2021**, *13*, 2469. <https://doi.org/10.3390/cancers13102469>

Academic Editors: Claudio Fiorino and Mary Frances McMullin

Received: 13 March 2021  
Accepted: 15 May 2021  
Published: 19 May 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 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 (<https://creativecommons.org/licenses/by/4.0/>).

**Simple Summary:** Non-invasive imaging modalities are commonly used in clinical practice. Recently, the application of machine learning (ML) techniques has provided a new scope for more detailed imaging analysis in esophageal cancer (EC) patients. Our review aims to explore the recent advances and future perspective of the ML technique in the disease management of EC patients. ML-based investigations can be used for diagnosis, treatment response evaluation, prognostication, and investigation of biological heterogeneity. The key results from the literature have demonstrated the potential of ML techniques, such as radiomic techniques and deep learning networks, to improve the decision-making process for EC patients in clinical practice. Recommendations have been made to improve study design and future applicability.

**Abstract:** Esophageal cancer (EC) is of public health significance as one of the leading causes of cancer death worldwide. Accurate staging, treatment planning and prognostication in EC patients are of vital importance. Recent advances in machine learning (ML) techniques demonstrate their potential to provide novel quantitative imaging markers in medical imaging. Radiomics approaches that could quantify medical images into high-dimensional data have been shown to improve the imaging-based classification system in characterizing the heterogeneity of primary tumors and lymph nodes in EC patients. In this review, we aim to provide a comprehensive summary of the evidence of the most recent developments in ML application in imaging pertinent to EC patient care. According to the published results, ML models evaluating treatment response and lymph node metastasis achieve reliable predictions, ranging from acceptable to outstanding in their validation groups. Patients stratified by ML models in different risk groups have a significant or borderline significant difference in survival outcomes. Prospective large multi-center studies are suggested to improve the generalizability of ML techniques with standardized imaging protocols and harmonization between different centers.

**Keywords:** machine learning; esophageal neoplasms; radiology

## 1. Introduction

Esophageal cancer (EC) is a malignancy affecting more than 500,000 people worldwide, ranking as the sixth leading cause of cancer death [1,2]. ECs are classified based on cell types on histopathology, most commonly identified as either adenocarcinoma or squamous cell carcinoma (SCC). Late presentation is a frequent occurrence which means a higher

proportion of EC patients have a worse prognosis [2]. The standard treatments for EC patients include surgery, radiotherapy, and chemotherapy. It is reported that neoadjuvant chemoradiotherapy (nCRT) plus surgery can improve patient prognosis for late-stage EC [3,4]. Pathologic complete response (pCR) is related to the improvement of patients' survival. However, due to tumor heterogeneity, a proportion of patients receiving nCRT may still have disease recurrence after treatment [5]. Unfortunately, some patients may fail to achieve sufficient tumor reduction prior to surgery as well as suffer from chemoradiation-related toxicity. For locally advanced disease, chemoradiotherapy is considered an option to improve survival and reduce recurrence [6]. Despite recent advances in therapeutics, the 5-year overall survival rate remains poor, ranging from 15% to 25% (3).

Accurate staging, treatment planning, and prognostication in EC patients is vital. Most recently, investigators have looked at novel applications such as radiomics using non-invasive imaging methods to improve the patient pathway. Previously hidden information can be found among different imaging modalities which may reflect the pathogenesis of EC. Computed tomography (CT), positron emission tomography (PET), mostly combined with CT (PET-CT), magnetic resonance imaging (MRI), and endoscopic ultrasonography (EUS) are commonly used for staging and follow-up [7]. The two most commonly used modalities for EC patients are CT and PET. However, their abilities to find small-sized lesions are limited, which affects the sensitivity and specificity of disease detection [8].

Artificial intelligence (AI) is a general term that refers to a wide range of algorithms capable of identifying features amongst a large amount of data and performing inferences. Machine learning (ML) is a subset of AI and refers to algorithms able to learn and predict without explicit instructions. The development of ML has provided a new scope in imaging analysis. Radiomics can convert images into high-dimensional features by calculations based on shape and textural features as well as higher-order spatial features, incorporating relationships at a pixel level [9]. It has been reported that using a handcrafted radiomics approach improved the imaging-based classification system in characterizing the heterogeneity of primary tumors and lymph nodes in EC patients. More recently, deep neural networks, a branch of ML, have been regarded as promising alternatives for extracting features and conducting medical image analysis in an end-to-end scheme [10–12]. Remarkable success with various analysis tasks have been witnessed, including diagnosis of skin cancer [13], prediction of cardiovascular risk factors [14], classification of lung cancer histopathology images [15], detection of lung abnormalities in COVID-19 CT images [16], and automated segmentation for robust quantitative measurements [17]. Unlike the conventional handcrafted radiomics approaches, deep networks are composed of multiple nonlinear transformation layers to perform automatic feature extraction in a data-driven way, thus is powerful in capturing more high-level and generalizable features for intelligent analysis of complex medical images. Nevertheless, there is ongoing debate, particularly about the interpretability and clinical usefulness of ML studies. We set out to provide a summary of the evidence of the most recent developments in ML application in imaging pertinent to EC diagnosis and follow-up.

## 2. Machine Learning and Radiomics Workflow for Oncology Imaging

### 2.1. Radiomics

Advancements in pattern recognition, medical image analysis, and artificial intelligence have laid the foundation for the rapid maturation of radiomics. Radiomics is the process of quantifying medical images into high-dimensional data with respect to prediction targets and subsequently mining the data for more informed clinical decision-making [18]. Contrary to conventional application of medical images, mainly for qualitative diagnosis and visual representation, radiomics exploits image analysis to identify features, often from standard-of-care images, for improving prediction accuracy in diagnosis, prognosis, and determination of optimal treatment regimes. Inclusion of the suffix -omics in radiomics indicates its ability in producing high-dimensional data from limited samples, as -omics fields are characterized by their data-mining potential [19]. The idea that studying

information embedded in medical images could aid in improving prediction and reflect more accurate clinical endpoints is the driving force of radiomics. More specifically for cancer imaging, extracting image features based on tumor texture, phenotype, and habitat, radiomics' potential in quantifying intratumoral and intertumoral heterogeneity brings hope for more personalized and effective treatments for cancer patients, particularly where contemporary treatments may fail due to intratumor heterogeneity [20]. Varying methods of feature extraction exist between different radiomics pipelines [21]. The image biomarker standardization initiative (IBSI) is an independent international collaboration, which aims to improve the reproducibility and standardization of radiomics studies [22]. The IBSI reference manual has made recommendations for a set of standardized radiomics feature calculation and processing guidelines to practice radiomic studies. It is suggested to check radiomics software's compliance with the IBSI standard before application for better clinical translation.

## 2.2. Radiomics Workflow

The workflow of a radiomic study can be categorized into six stages: problem identification, data curation, feature extraction, feature reduction, modelling, and model validation [23,24].

### 2.2.1. Problem Identification

The first step is to identify the clinical problem and what researchers are aiming to predict. A prediction endpoint (PE) usually refers to a clinical endpoint relevant to the patient and attending clinician. PEs could be categorical (for example treatment response patterns) or scalar (for example disease-free survival in months).

### 2.2.2. Data Curation

Information about imaging modality that would be applicable to the clinical problem is needed. Image acquisition follows and involves obtaining the medical images from all patients in full accordance with the protocol to ensure consistency. Radiomics features are typically extracted from the regions of interests (ROIs). Delineation of ROI is conducted manually, semi-automatically or fully automatically. Manual segmentation requires experienced radiologists to contour the full tumor regions, which is still regarded as the gold standard in many applications [25–27]. Semi-automatic segmentation is usually realized by thresholding and region-growing techniques, which is more efficient as it requires less manual intervention [28,29]. The manual or semi-automatic segmentation could be conducted using open-source or commercial software, such as ITK-SNAP [30], LIFEX [31], MITK [32], ImageJ [33], and 3D Slicer [34]. Bias caused by inter-algorithm and inter-reader variability [35] could be limited by multiple segmentation, such as manual segmentation by multiple clinicians or using a diverse combination of segmentation algorithms. A recent systematic review by Traverso et al. [36] pointed out that radiomics textural and shape features could be sensitive to manual segmentation differences. Semiautomated segmentation methods are found to provide a better alternative to manual slice-by-slice delineations in terms of feature reproducibility [37,38]. Automatic segmentation methods are suggested to be broadly applied in EC studies to optimize the clinical applicability. Auto-segmentation using deep learning is now a rapidly developing technique and has been proved to be potentially a reliable and reproducible tool for tumor delineation [39,40]. U-Net is a commonly used network architecture for auto-segmentation in medical images with exceptional performance demonstrated on many applications, which is a convolutional neural network consisting of a contracting path and a symmetric expanding path [41]. More complex deep neural network architectures with residual connections, skip connections, dense connections, or attention mechanisms have also been investigated for boosting automated segmentation accuracy. In addition to medical imaging data, researchers will also need to collect related clinical data (for example survival outcomes, demographics, clinical categorization).

### 2.2.3. Feature Extraction

Feature extraction with high throughput is the essence of radiomics and involves quantitative descriptions of characteristics of ROI. There are several features that can be extracted, including shape-based features, first order, and higher-order features. Currently, there are several open-source packages and software available that can be utilized to create up to thousands of features. Readers are advised to refer to the following excellent articles for further insight [22,31,42]. It is important to note that features can be computed based on pixel spacing, quantization method, and the number of gray levels but care must be taken to ensure a consistent approach is taken when generating these features. Moreover, radiomics features could be both 2-dimensional (2D), depicting the lesion in a representative layer (usually the layer with the maximal diameter of largest tumor cross-section) or 3-dimensional (3D), contouring the whole lesion volume ROIs. Controversy still exists over the discriminative ability of 2D and 3D radiomics features. The 3D ROIs are considered to contain more comprehensive information and have shown good classification ability [43]. However, Shen et al. has demonstrated that 2D radiomics features showed better prognostic performance than 3D features of non-small cell lung cancer, probably due to inconsistent spatial resolution in CT images [44]. Yang et al. has proved that a combination of 2D and 3D features could optimize the prediction performance [45].

Radiomics features are sensitive to CT acquisition and reconstruction parameters. It is impractical to standardize platforms and parameters from all institutions in clinical settings. To reduce the batch effect, the ComBat harmonization has been recently shown to correct the difference in extracted radiomics feature values resulting from different image acquisition protocols [46,47]. This method has been used to properly correct for the scanner effect while preserving the specific characteristics of each texture pattern.

### 2.2.4. Feature Reduction

Increasing the number of features does not necessarily mean better performance. In fact, it has been shown that a high number of features often leads to high redundancy and poor repeatability necessitating feature reduction [48]. The purpose of feature reduction is to optimize the feature space (either to selected few from the original list, engineered, or by reducing the dimension) before inputting into the model.

Feature selection is a common method used to remove irrelevant and redundant features. Traditionally, they are divided into filter, embedded, and wrapper methods. Filter methods have been most favored, and these are independent of the machine learning classifier. A subset class of the filter method known as ranker filters are techniques that have been most commonly employed in the early machine learning application in radiomics. In practical terms, the process usually begins with the removal of correlated features, usually based on a correlation matrix. For example, a Pearson's correlation matrix can be computed on the features set, and the highly correlated features (set for example as  $p > 0.95$ ) can then be excluded. Association scores can further be created based on univariate or multivariate analyses and these can be ranked according to the association scores. A common method is to then choose the top-ranked features (either in number or proportion). Additionally, features can also be subsequently ranked based on their predictive power, for example using Maximum Relevance Minimum Redundancy (mRMR) and recursive feature elimination (RFE) [49,50]. Care must be taken when using other feature selection methods such as wrapper selection bias could be introduced, although this could be mitigated by ML approaches such as Borda method or the bootstrap method [51,52]. Investigators have also used regularization techniques. Regularization techniques, such as least absolute shrinkage and selection operator (LASSO) and ridge regression in prediction models, aim to reduce (or shrink or smooth) the number of input features by removing redundant or highly correlated features [53]. Finally, investigators have also used unsupervised dimension reduction methods that could reduce feature number without the introduction of PE information, for example principal component (PCA) or linear discriminant analysis (LDA). These have been shown to reduce overfitting and improve prediction performance

in ML models [54,55]. Reducing the number of features helps with interpretation, making a more plausible linkage between feature and outcome easier to identify.

#### 2.2.5. Modelling

Model development aims to classify patients into different risk groups in relation to PE. ML methods explore the correlation between selected features and the targeted PE in the training set [25]. *Logistic regression* is a classic binary classification tool that fits a logistic function to classify data into one of two dependent variables and can be extended to classify multiple groups of events [56]. *Support vector machine* is one of the most commonly used ML models for radiomics studies [49,57,58]. This technique identifies a hyperplane or set of hyperplanes in the high-dimensional feature space for the classification of new data [59]. *K-nearest neighbor* classifies an unknown datum by comparing it with labelled data [60]. *Decision trees* are a series of classification branches that help visualize the process of making a prediction about the PE [61]. *Random forests* are a combination of multiple *decision trees* developed from the same training set to correct for potential overfitting of using a single decision tree [62]. Previous studies comparing the predictive performances across different classifiers have shown the optimal performance achieved by *Random forests* [63,64].

Overfitting may result from complex model architectures which have the issue of over-parameterization. The model may perform well in training data, but poorly in validation data, even split from the same cohort source. This can be reduced by acquiring sufficient data for training or making the training data more diverse so that it can be representative of the population. If the rate of the expected outcome is low making the dataset imbalanced, there are resampling techniques that may help to boost the minority class and improve predictive performance [63,65]. While it is ideal to keep the number of features as low as possible, the number of features is not a major issue if the model is validated externally with reasonable performance [18]. The selection of the optimal model for one specific dataset requires an understanding of the classifier mechanism and characterization of the dataset or exploratory data experiments. For better clinical application, a nomogram incorporating the radiomics signature and clinical risk factors could be built. This graphical calculation instrument is a user-friendly tool for patient-clinician communication [66]. Akaike information criterion (AIC), which measures the goodness of fit of the model, is commonly used for selection of important clinical variables for the final nomogram model [67].

#### 2.2.6. Model Development

In machine learning model development, it is preferred for the input dataset to be divided into the training set and the test set. A “training set” is only used to train the model. A “test set” could be further divided into internal or external, and is only used to evaluate the model. If the test set is from an independent center, this is often termed “external validation” set in clinical studies. This means that there is no overlap with the training data, and it is also from a different center. Nonetheless, it may be difficult to recruit significant amounts of independent data in preliminary or pilot radiomics studies. If this is not possible, then internal validation method of the test set is used. Note that the term “internal validation has different meanings in clinical and machine learning literatures. For the purpose of this article “internal validation” refers to validation with the dataset from one single center.

In some cases, data can be used for model hyper parameters optimization requiring “cross-validation”, in which part of the training dataset is used for training and part for validation. There are several methods to do this. Firstly, in k-fold cross-validation [13], the dataset is divided into k (usually k = 10) groups (called folds). (k-1) groups will be used as the training set whilst the remaining group is the testing set. A small proportion of data from the training set is used to validate the model during training time, in order to select hyper-parameters of ML models. This process is repeated k times in the test set once and training set k-1 times, with each group being in the training set k-1 times

and testing set once. This ensures that all data is used to test the model for performance evaluation. Secondly, in leave-one-out cross-validation [68], which is similar to k-fold cross-validation, the data is divided such that each patient is treated as one fold. In each cycle, one patient will be the testing set whilst all others will be the training set; this process repeats until all patients have been used in the testing set once. This scheme is used only when the dataset is small, in order to keep more data for model training. Thirdly, hold-out validation [69], which is the simplest form of internal validation in one center, involves randomly separating the data into a training set (again including part of data for model hyper-parameter validation) and a testing set [70]. Besides these, a more ideal way is to validate the model in a prospective dataset.

Several performance metrics have been used in the evaluation of ML performance, and a few commonly used methods will be discussed here. The most commonly used method typically involves an analysis of the area under curve (AUC) of the receiver operating characteristic (ROC) curve. Comparison between AI algorithms can be done by evaluating the AUC and individual sensitivity, specificity values of each ROC. The optimal AI algorithm depends on the context in which the model will be implemented. For comparison between humans and AI performance, a summary ROC curve rather than the convention of using one point on an ROC graph can more accurately quantify human performance and allow for robust comparison between ML models and humans [71]. Other commonly used metrics such as Matthew's correlation coefficient (MCC) have also been used which incorporates true and false positives, negatives, and equals [72]. A coefficient of +1 represents a perfect prediction, 0 no better than chance, and -1 total disagreement between prediction and the actual outcome. Precision-Recall (PR) curve has also been used, which plots precision against recall to show the trade-off between them for the different feature or parameter settings. For classification tasks involving datasets with very few positive cases, the area under the PR curve (AUPRC) is a better measure of accuracy than AUROC [73].

Validation reflects the discrimination and calibration capacity of the model, which respectively quantifies model sensitivity and specificity, and the agreement between predicted and observed outcomes. Relying purely on the performance metrics outlined above may have limitations. For example, risk prediction models may be highly discriminatory but poorly calibrated [74] or have sensitivity/specificity cut-points that fail to maximize clinical utility [75].

### 2.3. Deep Learning

Different from the conventional low-level radiomics features with hand-crafted descriptors, deep learning [76] has recently emerged to automatically learn high-level features in a data-driven way, which is capable of capturing the complicated characteristics of medical images. Among different classes of deep learning algorithms, convolutional neural network is one of the most widely-used solutions in analyzing image data to extract hierarchical feature representations. A convolutional neural network (CNN) is composed of multiple convolutional layers with a set of filters (also called kernels), which is the core component, combining with pooling operations and non-linear activation functions. The weight-sharing of convolution kernels over the entire input image significantly reduces the number of trainable parameters. CNNs with 2D kernels (namely 2D CNNs) are generally applicable to different types of imaging modalities, but for volumetric images, 3D CNNs have been demonstrated to be more effective in capturing important contextual information along the third dimension [77,78].

For recent deep learning models, the classifier learning can be performed in a unified network with the feature extraction and trained in an end-to-end process. As previously mentioned, CNNs are the most popular deep networks in medical image analysis and have shown exceptional performance. Alexnet [79] is one of the commonly used CNN architectures in cancer imaging [80], which is an early proposed shallow network with 11 layers. Inception network [81] is another popular architecture [13], which has a more

complex design with deeper architecture and multiple filters of different sizes operating on the same level. ResNet [82] is another recent CNN architecture with skip connections to benefit gradient backpropagation [83]. Besides using standard network architectures, some studies have designed specific CNNs for the targeted problem. For example, Bizzego et al. [72] use two identical and parallel CNN streams for CT and PET data respectively, that each stream consisting of five convolutional layers combining with normalization, dropout, and pooling layers, and the outputs of the two streams are merged at the fully connected layer. Effective training of a CNN usually requires large amounts of data to sufficiently optimize the large number of parameters in a network and to reduce the overfitting problem. If the size of available dataset is large enough, training a deep network from scratch can better optimize the model for the specific problem and dataset. However, obtaining annotations from medical experts is costly and typically limits the amount of available data for network training. When data is limited, data augmentation techniques, such as generative or associative modelling can generate additional, artificial training data based on real training sets [84,85].

Transfer learning is commonly adopted to mitigate the problem of inadequate training data. With transfer learning, a deep network is firstly trained with available large-scale natural images or medical images of different problems, the pre-trained network is then fine-tuned with the desired data and task. The effectiveness of transfer learning in reducing the overfitting problem caused by a small number of training samples and improving the model's performance has been well studied in previous works on mammographic tumor classification [86] and thoraco-abdominal lymph node detection [87]. Deep features can be extracted from medical images by training CNNs with pre-defined prediction tasks such as disease classification, tumor stage diagnosis, or survival prediction. For deep learning-based radiomics with end-to-end training, the extracted features can be directly used by the classifier layer of the rest of the deep networks for the targeted analysis tasks. There are also studies separating the learned features from the deep networks to be combined with the hand-crafted radiomics features for subsequent modelling [72]. The major advantage of deep learning-based feature extraction is that no specific domain knowledge is required for feature engineering, and the representative and high-level features can be learned in a completely automatic manner. Recent works [88,89] have shown that the automatically learned deep features of neural networks can outperform the hand-crafted ones in some applications. One key challenge of applying deep networks in clinical decision making is that deep networks are black box models with multilayer nonlinear operations, thus the reasoning behind the results from deep networks are very difficult to interpret clinically. Explainable AI is an emerging field of active research in trying to address this challenge [90,91].

Recent studies [72,80] have shown that the conventional handcrafted radiomics and the deep learning-based radiomics can be fused to improve analysis accuracy. The fusion operation can be performed at the decision level or at the feature level. For decision-level fusion, the hand-crafted radiomics and deep radiomics are first independently trained to produce prediction respectively, then the outputs are then combined together with a certain voting strategy to achieve the final decision. For feature-level fusion, the hand-crafted features and deep features are extracted separately, and then the two types of features are combined to go through a classifier for the final prediction. On the other hand, the combination of handcrafted radiomics and deep learning features did not enhance the prediction performance in Yun's previous study [92]. This is an area of active research, and further investigations are needed to properly address the value of radiomics in combination with deep learning methods [93]. The use of deep learning methods, however, is advantageous in potentially being more automated than radiomics features extraction, as the latter requires segmentation, which depending on the disease entity may not be automated.

#### 2.4. Automation of Machine Learning Pipeline in Clinical Workflows

For the identification of optimal pattern for a specific dataset, the discovered machine learning pipeline could be quite variable, as the selection of the ML approach and setting of parameters are dependent on individual researchers. In recent years, there has been a trend for automation of machine learning pipeline in clinical workflows. Cai et al. [94] provided an online calculator based on the radiomics model for the prediction of treatment response of bevacizumab in brain necrosis after radiotherapy. Such a user-friendly tool could help the facilitation of personalized and precise treatment of patients. Su et al. [95] applied the Tree-based Pipeline Optimization Tool for the construction of the optimal radiomics model. Without human annotation, the automatically optimized machine learning pipeline showed good prediction accuracy for H3 K27M mutation status in patients with midline gliomas. More recently, automation of machine learning pipeline has found promising applications for fast diagnosis and risk stratification for patients with infectious diseases [96–98]. Wang et al. [96] developed a novel fully automatic deep learning system using CT imaging for fast screening of COVID-19 to identify potential high-risk patients, which could be fast and more repeatable as it required no time-consuming human involvement. Automation of machine learning pipeline in clinical workflows could significantly improve the optimization process of medical resource.

### 3. A review of Literature Using Machine Learning and Radiomics Applications in EC

#### 3.1. Eligible Studies

In this paper, a comprehensive review of studies based on ML methods for the diagnosis of any aspect in EC patients using non-invasive medical imaging was performed. According to the scope of the review, we consider ML-based investigations aimed at relevant objectives in clinical practice: treatment response evaluation, prognostication prediction, diagnosis, and biological characterization. Non-invasive imaging modalities considered are CT, PET, PET-CT, and MRI. Studies using invasive endoscopy or endoscopic ultrasound were not included. We searched for articles in 3 database resources including PubMed, EMBASE and Cochrane Library. All the English publications from 1 January 2000 until 16 October 2020 were searched. The reference articles of the selected papers were also checked. The complete search strategy is shown in Supplementary Material. We included any study design, with a minimum of 10 patients, except for letters to the author, comments, and case reports. Studies with only correlation analyses between individual imaging features and outcomes, without utilizing ML approaches were excluded.

#### 3.2. Data Analysis

The measurements obtained in the validation groups were used as the main results. For papers with multiple clinical outcomes, the primary aim was chosen. The AUC was preferred to summarize the predictive value of proposed models. If unavailable, other accuracy metrics were recorded (for example sensitivity, specificity, C-index, etc.). We rated AUCs by the following range: 0.60–0.70 as poor, 0.70–0.80 as acceptable, 0.80–0.90 as excellent, and 0.90–1.00 as outstanding [99]. For prognostic groups, cox-proportional hazards ratios were used as outcome measures for risk stratification. For articles with multiple test models for the same cohort of subjects, the final model with the best performance was recorded. Confidence intervals of these accuracy metrics were retrieved if available.

The results from the included studies are summarized in Table 1. The studies for imaging machine learning applications in EC commonly followed certain steps, which are shown in Figure 1.



**Table 1.** Characteristics of studies for non-invasive imaging machine learning applications in esophageal cancers.

Studies	Year	Type	Treatment Regime	Approach	Modality	Sample Size (Training + Testing)	ML Techniques	Classifiers for The Final Model	Specific Predicted Clinical Outcome	Type of Validation	Main Results (in Test Set)	Reference Standard
<b>Treatment response</b>												
Cao et al. [100]	2020	All SCC	CRT	Radiomics	PET	159 (93 + 66)	LASSO	LASSO	pCR	External validation	AUC = 0.835	CT
Hu et al. [101]	2020a	All SCC	nCRT followed by surgery	Radiomics	CT	231 (161 + 70)	Decision tree, recursive feature addition, LR, SVM, K-nearest neighbors, naive bayes, decision tree, RF, and extreme gradient boosting	SVM	pCR	External validation	AUC = 0.852 (95% CI, 0.753–0.951), accuracy = 84.3%, Se = 90.3%, Sp = 79.5%	Histology
Hu et al. [83]	2020b	All SCC	nCRT followed by surgery	Radiomics and deep learning	CT	231 (161 + 70)	Xception, VGG16, VGG19, ResNet50, InceptionV3, InceptionResNetV2, recursive feature addition, SVM	ResNet50-SVM	pCR	External validation	AUC = 0.805 (95% CI, 0.696–0.913), accuracy = 77.1%, Se = 83.9%, Sp = 71.8%	Histology
Yang et al. [102]	2019a	All SCC	nCRT followed by surgery	Radiomics	CT	55 (44 + 11)	LASSO	LR	pCR	Training + testing set (randomly separated)	AUC = 0.79 (95% CI, 0.48 to 1.00)	Histology
Hou et al. [103]	2018	All SCC	CRT	Radiomics	MRI	68 (43 + 25)	SVM and ANN	ANN	pCR	Training + testing set (randomly separated)	AUC = 0.843, accuracy = 84.3%, Sp = 100%	CT/MRI
Beukinga et al. [104]	2018	Adenocarcinoma 89.0%, SCC 11.0%	nCRT followed by surgery	Radiomics	PET	73	LASSO	LR	pCR	No validation	AUC = 0.81	Histology
Hou et al. [105]	2017	All SCC	CRT	Radiomics	CT	49 (37 + 12)	SVM and ANN	ANN	pCR	Training + testing set (randomly separated)	AUC = 0.800, accuracy = 91.7%	CT

Table 1. Cont.

Studies	Year	Type	Treatment Regime	Approach	Modality	Sample Size (Training + Testing)	ML Techniques	Classifiers for The Final Model	Specific Predicted Clinical Outcome	Type of Validation	Main Results (in Test Set)	Reference Standard
Van Rossum et al. [106]	2016	All adenocarcinoma	nCRT followed by surgery	Radiomics	PET	217	LR	LR	pCR	Training + testing set (randomly separated, bootstrap method, repeated 1000 time)	C-index = 0.77 (95%, 0.70–0.83), Se = 0.78	Histology
Beukinga et al. [107]	2016	Adenocarcinoma 90.7%, SCC 9.3%	nCRT followed by surgery	Radiomics	PET-CT	97	LASSO	LR	pCR	No validation	AUC = 0.74	Histology
Desbordes et al. [108]	2016	Adenocarcinoma 12%, SCC 88%	nCRT followed by surgery or CRT	Radiomics	PET	65	Hierarchical forward selection method, RF, SVM	RF	pCR	Training + testing set (randomly separated, repeated 10 times)	AUC = 0.836 ± 0.105 (mean ± SD), Se = 82 ± 9%, Sp = 91 ± 12%	Follow-up based on clinical examination, endoscopy with biopsies and PET/CT
Ypsilantis et al. [109]	2015	Adenocarcinoma 81.1%, SCC 18.9%	nCRT followed by surgery	Radiomics and deep learning	PET	107 (96 + 11)	LR, gradient boosting, RF, SVM, 1S-CNN, 3S-CNN	3S-CNN	pCR	10-fold cross validation	Averaged Se = 80.7%, Sp = 81.6%	Histology
Zhang et al. [110]	2013	Adenocarcinoma 85%, SCC 15%	nCRT followed by surgery	Radiomics	PET	20	SVM and LR	SVM	pCR	10-fold cross validation	Averaged AUC = 1.00, Se = 100%, Sp = 100%	Histology

Table 1. Cont.

Studies	Year	Type	Treatment Regime	Approach	Modality	Sample Size (Training + Testing)	ML Techniques	Classifiers for The Final Model	Specific Predicted Clinical Outcome	Type of Validation	Main Results (in Test Set)	Reference Standard
<b>Prognosis</b>												
Qiu et al. [111]	2020	All SCC	nCRT followed by surgery	Radiomics	CT	206 (146 + 60)	LASSO	Cox proportional hazards model	RFS	Training + testing (temporally separated)	Radiomics signature was significantly associated with RFS (log-rank test, $p < 0.0001$ ; HR, 3.606; 95% CI, 1.742–7.464). Radiomics nomogram C-index 0.724 (log-rank test, $p < 0.001$ ; 95% CI, 0.696–0.752)	Follow-up
Chen et al. [112]	2019	All SCC	nCRT followed by surgery	Radiomics	PET	44 (22 + 22)	LR	Cox proportional hazards model	OS and DFS	Training + testing set (randomly separated)	Significant risk stratification for DFS (log-rank test, $p = 0.001$ ) and OS (log-rank test, $p < 0.001$ )	Follow-up
Yang et al. [113]	2019b	All SCC	Not specified to one kind of treatments	Deep learning	PET	model 1: 1107 (798 + 309), model 2: 548	3D-CNN based on ResNet	3D-CNN based on ResNet	OS	5-fold cross validation	The prediction result remained an independent prognostic factor (multivariable overall survival analysis, hazard ratio: 2.83, $p < 0.001$ ).	Follow-up
Xie et al. [114]	2019	All SCC	CRT	Radiomics	CT	133 (87 + 46)	The K-means method, LASSO	Cox proportional hazards model	OS	External validation	Prediction model AUC, 0.805 (95% CI: 0.638–0.973). Significant risk stratification (log-rank test, $p < 0.001$ )	Follow-up

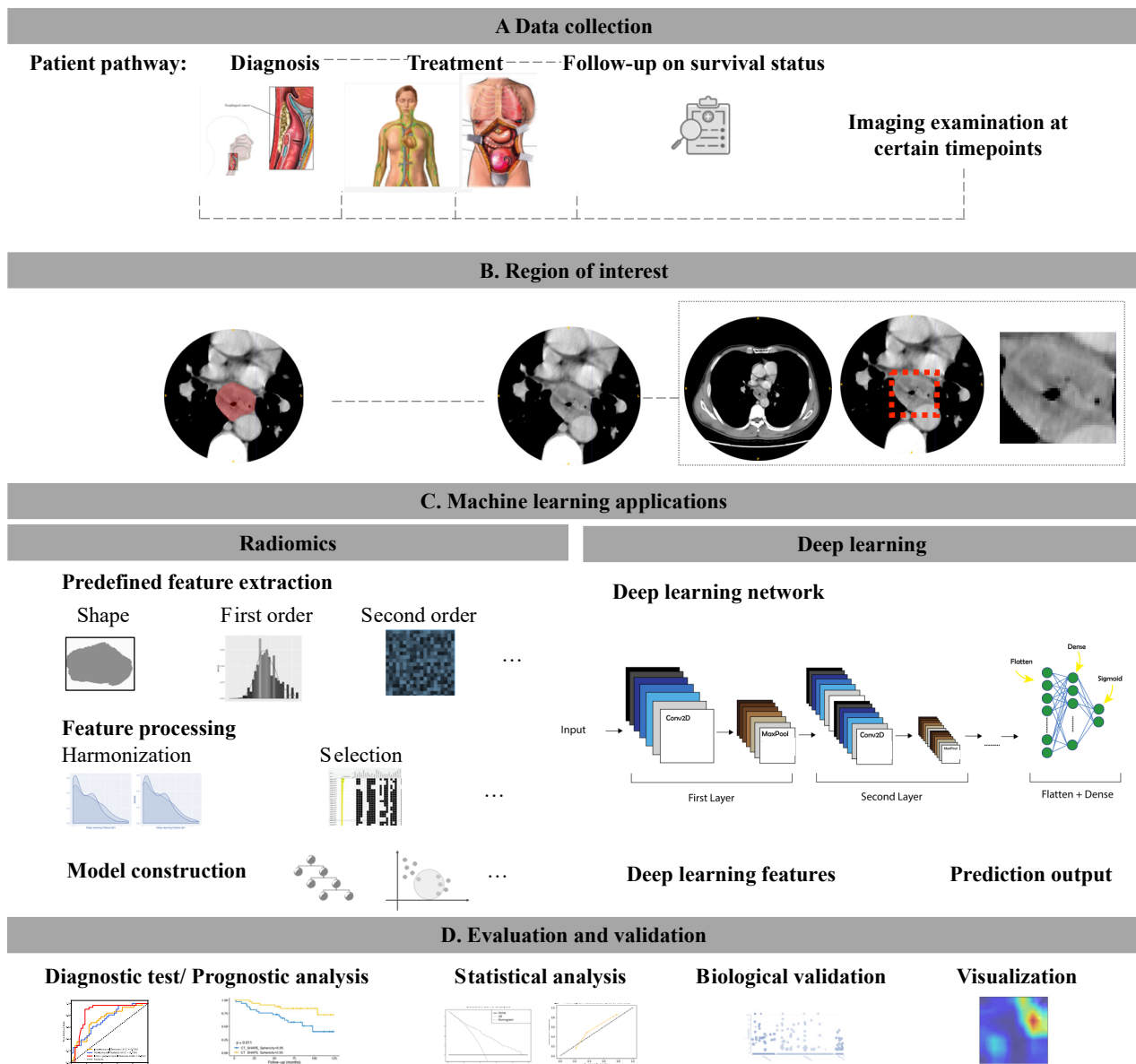
Table 1. Cont.

Studies	Year	Type	Treatment Regime	Approach	Modality	Sample Size (Training + Testing)	ML Techniques	Classifiers for The Final Model	Specific Predicted Clinical Outcome	Type of Validation	Main Results (in Test Set)	Reference Standard
Larue et al. [115]	2018	Adenocarcinoma 81%, SCC 19%	nCRT followed by surgery	Radiomics	CT	239 (165 + 74)	Recursive feature elimination, RF	RF	OS	External validation	Prediction model AUC: 0.61 (95% CI: 0.47–0.75). Borderline significant risk stratification (log-rank test, $p = 0.053$ )	Follow-up
Foley et al. [116]	2017	Adenocarcinoma 78.4%, SCC 21.6%	Not specified to one kind of treatments	Radiomics	PET	403 (302 + 101)	Automatic Decision Tree Learning Algorithm for Advanced Segmentation	Cox Regression Model	OS	Training + testing (temporally separated)	Significant risk stratification (log-rank test, $p < 0.001$ )	Follow-up
Xiong et al. [117]	2018	SCC	CRT	Radiomics	PET	30	RF, SVM, LR and extreme learning machine	RF	PFS	Leave-one-out cross validation	Prediction model accuracy = 93.3%, Sp = 95.7%, Se = 85.7%. Significant risk stratification: (log-rank test, $p < 0.001$ )	Follow up
<b>Lymph Node Metastasis</b>												
Wu et al. [88]	2020	All SCC	Surgery alone	Radiomics, computer vision, and deep learning	CT	411 (321 + 90)	Random Forest-Recursive Feature Elimination algorithm Elastic net approach (a combination of the LASSO and the ridge regression approaches)	LR	LN-positive versus LN-negative	External validation	AUC = 0.840	Histology
Qu et al. [118]	2018	Not stated	Surgery alone	Radiomics	MRI	181 (90 + 91)		LR	LN-positive versus LN-negative	Training + testing (temporally separated)	AUC = 0.762 (95% CI: 0.713–0.812).	Histology

Table 1. Cont.

Studies	Year	Type	Treatment Regime	Approach	Modality	Sample Size (Training + Testing)	ML Techniques	Classifiers for The Final Model	Specific Predicted Clinical Outcome	Type of Validation	Main Results (in Test Set)	Reference Standard
Tan et al. [119]	2018	All SCC	Surgery alone	Radiomics	CT	230(154 + 76)	LASSO	LR	LN-positive versus LN-negative	Training + testing set (randomly separated)	AUC = 0.773 (95% CI: 0.666–0.880)	Histology
Shen et al. [120]	2018	Not stated	Surgery alone	Radiomics	CT	197 (140 + 57)	Elastic net approach (a combination of the LASSO and the ridge regression approaches)	LR	LN-positive versus LN-negative	Training + testing (temporally separated)	AUC = 0.771 (95% CI: 0.632–0.910)	Histology
Li et al. [121]	2020	All SCC	Surgery alone (T3 cases) or no treatment (non-disease controls)	Radiomics	CT	57	Unspecified	LR	Malignant versus normal esophageal wall	No validation	AUC = 0.80	Histology
Ou et al. [122]	2019	All SCC	Not specified to one kind of treatments	Radiomics	CT	591 (413 + 178)	LASSO, LR, decision tree, random forest, SVM, and X-Gradient boost	LR	Resectability	Training + testing set (randomly separated)	AUC = 0.87 ± 0.02; accuracy = 0.86, and F-1score = 0.86	NCCN guidelines
Hoshino et al. [123]	2020	All SCC	Not specified to one kind of treatments	Radiomics	CT	92	LR	LR	Expression of microRNA-1246	No validation	AUC = 0.754, Se = 71.29%, Sp = 73.91%	Follow-up

SCC: squamous cell carcinoma; CT: computed tomography; PET: positron emission tomography; MRI: magnetic resonance imaging; LASSO: least absolute shrinkage and selection operator; LR: logistic regression; SVM: support vector machine; RF: random forest; pCR pathologic complete response; AUC: area under the curve; CI: confidence interval; SD: standard deviation; ANN: artificial neural network; CNN: convolutional neural network; RFS: recurrence-free survival; PFS: progression-free survival; CRT: chemoradiation therapy; nCRT: neoadjuvant chemoradiation therapy; DFS: disease-free survival; OS: overall survival; Se: sensitivity; Sp: specificity; LN: lymph node; NCCN: National Comprehensive Cancer Network.



**Figure 1.** An analysis workflow summarization of studies for imaging machine learning applications in esophageal cancers. (A) Patient pathway in clinical practice (B) Radiological features extracted from the handcrafted radiomics and deep learning method using different regions of interest. (C) Machine learning models constructed with the selected features (D) Model evaluation and evaluation.

### 3.3. Main Findings

#### 3.3.1. ML and Treatment Response Evaluation in ECs

Most published studies ( $n = 12$ ) focused on the prediction of treatment response for patients receiving chemoradiotherapy or nCRT [83,100–110]. The ML algorithms achieved an AUC of 0.78–1.00. There were six studies using PET, four studies using CT, one study using MRI, and one study using PET-CT. Zhang et al. [110], which constructed a support vector machine model based on PET radiomics features combined with conventional imaging and clinical variables. They achieved the best prediction result (AUC = 1.00) with 20 patients although with no validation on a test set. Ypsilantis et al. [109] provided initial evidence for the potential predictive power of CNNs with an averaged sensitivity and specificity of 80.7% and 81.6%, respectively. Desbordes et al. [108] applied a random forest

classifier based on conventional and textural features for treatment response prediction by using baseline PET images. Both individual features and the constructed model with combined features were of predictive value, with an AUC of 0.810 and 0.836. Beukinga et al. published two studies evaluating radiomics for the prediction of response to nCRT. The first study [107] selected only pre-treatment clinical and radiomics features from PET/CT scans. For the latter study [104], significant features were derived from PET images of two phases (both baseline and restaging). The discrimination accuracy was improved from acceptable (AUC = 0.78) to excellent (AUC = 0.81). Van Rossum et al. [106] also showed the added value of posttreatment radiomics features for disease evaluation. Hou et al. studied two modalities for the evaluation of CRT response. Radiomics features extracted from both the CT [105] and MRI [103] showed predictive capabilities with AUC of 0.972 for CT and AUC of 0.929 for MRI, using imaging interpretation as a reference standard. Yang et al. [102] developed three predictive models for treatment response after nCRT and noted that overfitting was a problem for small sample size studies. Hu et al. [101] found that a combination of peritumoral radiomics features appeared to improve the predictive performance of intratumoral radiomics for pre-treatment prediction of pCR (AUC = 0.85, 95% CI, 0.75–0.95) for nCRT using CT radiomics features, with a cohort of 231 patients and external validation of the results. The same cohort of patients was used for the exploration of the transfer learning approach and the results showed that ResNet50-based deep learning features had the predictive ability for treatment response in esophageal SCC [83]. Cao et al. [100] used the limma method, which is commonly used for genetic analysis, for the identification of significant radiomics features. The selected features were then fitted into the least absolute shrinkage and selection operator (LASSO) logistic regression model and achieved an AUC of 0.84 in the test set to predict response for concurrent CRT. The majority of studies [83,101,102,104,106,107,109,110] evaluated pCR for nCRT before surgery and used histology results as the reference standard. Other studies [100,103,105,108] collected patients receiving chemoradiotherapy and adopted PET/CT/MRI imaging and follow-up information for pCR evaluation because surgical resection was not performed and, therefore, pathological reference was not available.

### 3.3.2. ML and Prognosis Prediction in ECs

ML methods have shown the potential to act as prognostic tools for risk stratification in ECs. There are seven studies in the prognostic group [111–117], including three PET, three CT, and one PET-CT studies. Patients in different risk groups had a significant or borderline significant difference in survival outcomes. The primary outcomes were overall survival (OS), recurrence-free survival (RFS), and disease-free survival (DFS) predictions. Xiong et al. [117] extracted PET radiomics features from different treatment time points and concluded that mid-treatment features could be more informative for predicting local control. Foley et al. [116] demonstrated that a clinical prognostic model incorporating PET radiomics features could bring additional benefit for patients' risk stratification. Larue et al. [115] validated their radiomics-based model for EC patients receiving nCRT in an independent external cohort, which split risk groups for survival rates with borderline significance. Xie et al. [114] found that sub-regional radiomics features were of prognostic prediction value. The authors further correlated the imaging traits with the clinical variables and gene data to verify the biological significance. Yang et al. [113] proved that deep learning-based prediction could be an independent survival prognosticator for EC patients. Chen et al. [112] constructed a scoring system based on both clinical and PET radiomics features and enabled better stratification of patients into different long-term prognosis. Qiu et al. [111] developed three prognostic models, the nomogram based on both radiomics and clinical features achieved optimal performance with a C-index of 0.72 in the validation set.

### 3.3.3. ML and Lymph Node Metastasis Status in ECs

There were four studies [88,118–120] focusing on the lymph node metastasis status of EC, with three studies using CT and one study using MRI. Shen et al. [120] built a nomogram incorporating radiomics features, CT-reported suspicious lymph node number and tumor location, which showed good discrimination of lymph node status with a C-index of 0.75 in the validation set. Tan et al. [119] demonstrated that radiomics nomogram provided a good estimation of lymph node metastasis (AUC = 0.77) and outperformed size criteria. Most studies delineated whole tumor volume in all slices as the ROIs for feature selection. Qu et al. [118] provided evidence that MRI-based radiomics features were of predictive value for lymph node involvement. Wu et al. [88] showed that two-dimensional (2D) ROIs based on the largest cross-sectional area of the tumor lesion were also of prediction value (AUC = 0.84).

### 3.3.4. ML and Other Clinically Significant Outcomes in ECs

There were two studies focusing on the diagnosis of EC and one study focusing on gene expression [121–123]. CT imaging was used. Li et al. [121] found differences in the same parts of the normal esophageal wall and EC lesions using a multivariate regression classifier based on radiomics features. Ou et al. [122] investigated the value of radiomic models related to the resectability of ECs. Hoshino et al. [123] found that the gene expression level of miR-1246 could be inferred by imaging radiomics features, which was predictive of the prognosis of EC patients.

### 3.3.5. Study Characteristics

In general, all studies were retrospective design and most studies adopted radiomics approaches. Four studies [88,109,113,115] applied deep learning networks in the model construction. Incorporation of clinical features into data-based ML models could improve the prediction accuracy in 11 studies [88,104,106,110–113,116,119,120]. The majority of studies [83,88,100–103,105,111–114,119,121–123] evaluated SCC patients and some [104,106–110,115,116] mainly focused on adenocarcinoma. Many studies ( $n = 11$ ) did not provide sufficient sample size with less than 100 patients in the study cohort [102–105,107,108,110,112,117,121,123]. Four studies [104,107,121,123] lacked independent validation by using the same datasets as both training and test sets.

Feature harmonization was conducted in only two studies to remove inter-site technical variability [83,101]. Nine studies [88,103,104,106,111,116,119,120,122] applied additional statistical analysis on the prediction outcomes to evaluate model fitting, calibration, and clinical usefulness. Three studies [83,101,114] explored the correlation between imaging features extracted by ML and the genomic profiles for biological interpretation. Two studies [83,109] used visualization techniques to highlight important regions in the medical images. The visualized pictures generated from specific layers of the deep learning model highlighted the regions of interest for patient characterization. Visualizing the hot zone in the feature map could help to evaluate the interpretability of the models.

## 4. Summary and Perspectives

This study summarizes the main results and basic characteristics of ML techniques relevant to the clinical practice in EC patients. Surgery is one of the most frequently used treatments for resectable EC, but the long-term survival remains unsatisfactory even after curative operation [124,125]. The optimal treatment for EC is still unclear. Neoadjuvant or adjuvant therapies administrated as chemotherapy, radiotherapy, or simultaneous chemoradiotherapy have been adopted in clinical practice. The adjuvant chemotherapy or radiation therapy has not been shown to have additional survival benefit to patients compared to surgery alone [126]. More recently, as shown by recent landmark trials, nCRT plus surgery could be the most effective strategy for the improvement of resectability and maximization of long-term survival for locally advanced EC patients [3,4,127,128]. The updated CROSS study has shown that the median overall survival was 48.6 months



in patients receiving nCRT plus surgery cohort and 24.0 months in the surgery cohort ( $p = 0.030$ ) [128]. The recent NEOCRTEC5010 trial also showed that EC patients receiving nCRT followed by surgery have a significantly increased median overall survival than those receiving surgery alone (100.1 vs. 66.5 months,  $p = 0.025$ ) [4]. But patients have different responses to nCRT treatment, which significantly affects the survival outcomes. According to a recent review by Eyck et al. [129], the ability of commonly used imaging modalities to detect pCR after nCRT of EC patients was insufficient, indicated by the pooled sensitivities and specificities. An accurate estimation of residual disease could improve patient management.

For this purpose, there has been an increasing number of studies in the literature that utilized new methods for more accurate prediction. In our review article, we assessed the potential of ML approaches for a more accurate evaluation of pCR. This showed that ML models evaluating treatment response achieved reliable predictions ranging from acceptable (AUC = 0.70 to 0.80,  $n = 3$ ) to excellent (AUC = 0.80 to 0.90,  $n = 7$ ), and outstanding (AUC > 0.90,  $n = 1$ ) in the validation group. The reviewed studies adopted a variety of ML approaches. Advanced imaging features were extracted using radiomics calculation algorithm or transfer learning technique from various pre-trained CNNs. According to Hu's study characterizing relatively large sample sizes with extrapolation of developed models by external validation, the model using deep learning features extracted from Resnet had a better performance than the handcrafted radiomics model [83]. It is suggested that non-invasive ML-based imaging applications, such as radiomic techniques and deep learning networks are potentially useful for individualized patient's tumor characterization.

Patients' risk stratification mainly relies on the American Joint Committee on Cancer (AJCC) tumor, node, and metastasis (TNM) staging classification of epithelial cancers of the esophagus. Improvement of the staging method will aid in the current clinical practice. Involvement of lymph node invasion has been proved to be an important prognostic factor for both esophageal SCC and adenocarcinoma [130,131]. Recent refinement of the N descriptors (regional lymph node invasion) subcategory in the eighth edition has been shown to provide a more accurate and reliable risk stratification of EC patients [132]. Compared with the seventh edition, this new N subcategory excludes some regional lymph node stations common in the staging system of lung cancers [133]. Determination of clinical N (regional lymph node invasion) mainly relies on current imaging techniques. CT, one of the most commonly used non-invasive imaging techniques is often initially used for the evaluation of tumor growth and other structures. However, it is not optimal for lymph node determination [134,135]. Besides, the efficacy of MRI for the accurate estimation of N staging is still uncertain [136].

Improvement in imaging assessment of regional lymph nodes by using ML techniques is of clinical importance for prognosis and treatment decisions for EC patients. Radiomics, as an emerging tool, has shown potential values in predicting LN metastasis by extracting high-throughput quantitative features from medical images. Less attention has been paid to the prediction of LN metastasis in previous years. With the first study published in 2018, specific attention is paid to the evaluation of LN metastasis in EC patients. The reviewed studies [118–120] were mainly focused on the application of radiomics and Wu et al. [88] showed the prediction value of deep learning features for LN metastasis. According to our review, the new ML approach significantly improved the evaluation of N status using non-invasive CT and MRI modalities (AUC = 0.762–0.840 in the validation set), even outperforming the size criteria [88,118–120]. However, current published studies were limited to binary discrimination of LN status. The specific N staging information (N0/N1/N2/3) determined by evidence acquired before treatment affects the therapy scheme. Moreover, the drainage of the lymph nodes in EC is more complicated than in some cancers like lung cancers and breast cancers. The regional lymph node stations for EC patients extended from cervical to celiac regions, and the lymphatic drainages could be different at different anatomic sites. An optimal lymphadenectomy is necessary to maximize patients' survival. The extent of lymphadenectomy is variable in current clinical

practice [137,138]. Greater extent of lymphadenectomy was associated with patients' long-term survival and its effect on postoperative complications [139,140]. The evaluation of individual LN requires laborious and time-consuming work from experienced radiologists. More accurate evaluation of more precise N staging or individual lymph node status could lead to a more precise assessment of the ideal range of lymph node dissection.

In our review, we have uncovered several shortcomings of existing clinical applications using machine learning. We will discuss these issues more broadly and then more specifically in the following paragraphs. The topic of explainability has been shown to be an important aspect of ML research [141]. Doctors may find it difficult to apply ML models routinely. CNNs, for example, could contain millions of trainable parameters without clearly understandable biological patterns for human readers. It is crucial that clinicians and patients have the ability to understand the reasonings of predictions of these models for better informed decision-making. The accountability of ML approaches is of vital importance if we introduce computerized ML systems into clinical practice. Currently, there is a paucity in the understanding of the relationship between radiomics and underlying tumor biology. The exact relationship of a combination of radiomics features used in the predictive modelling remains unclear. Further exploration has been attempted by, for example, doing radiogenomics studies, which were done in a few reviewed studies, providing proof of potential biological explanation. The reliability of the established prediction models could also be validated by statistical methods to reduce the concerns regarding applicability. To this end, some models were validated by statistical analysis using model fitting, calibration, and clinical usefulness. Simple data-driven correlation may not be robust and links between imaging, clinical, and genetic features should be built. Note that the precise relationship between imaging features and tumor heterogeneity is not simply straightforward. Further investigation of the advanced statistical relationship should be established. Visualization is suggested to be employed to improve the interpretability of ML models. More work should be dedicated to making ML models more interpretable or explainable for decision-making.

A few more specific limitations are noted from our review. Firstly, ML algorithms utilized in some studies were trained and validated on the same dataset, which is not an accurate estimation of ML models. Secondly, the sample size is one of the vital factors for clinical studies, which could significantly affect the repeatability, reproducibility, and statistical power. Adequate sample size is needed to train and test the ML methods to minimize overfitting and improve the estimation performance. The sample size criteria for some ML studies for EC patients were established with less than 100 patients, which limits the reliability of the proposed models. Clinically meaningful size is required for significant discrimination of patients in different risk groups. Thirdly, the reference standard for each outcome in question may be different, reflecting the heterogeneity of the study design impacting comparison. Most studies evaluating treatment response, for example, compared their model with histology findings by pathologists, which is the current gold standard. However, some studies [100,103,105,108] used pre-treatment and post-treatment CT images to assess the response based on the Response Evaluation Criteria in Solid Tumors (RECIST) [142], when resected specimens were not available. RECIST is commonly used for residual disease evaluation for solid tumors but has limited ability in EC for its obscure boundaries or scarred tissues after chemoradiation therapy [143]. Fourthly, standardization in imaging protocols (for example, CT with or without intravenous contrast), radiomics extraction methods, and attempts to harmonize imaging data prior to predictive modelling were scantily performed. These have been known to affect the extracted features and remain major obstacles to generalizability.

We have a few recommendations for future investigations. Firstly, prospective large multi-center studies should be performed to improve the ML techniques and generalizability with standardized imaging protocols and harmonization between different centers. In our review, only six studies [83,88,100,101,114,115] included multi-center datasets. Secondly, the techniques of machine learning analysis should be standardized. More strict

adherence to the standard working pipeline as well as openly available source code is suggested to increase reproducibility and generalizability, particularly for radiomics features pipelines [144]. Thirdly, the application of ML methods should be expanded from restricted data-processing computers to portable machines accessible by cloud services or to intraoperative decisions in real-time. There are some challenges to overcome such as automated segmentation and real-time inference processing which currently limits its wider utility in the clinical environment. If this could be done, it would allow for prospective study design, and if randomized, it could add significant value to the assessment of the utility of these techniques in real-world clinical practices. Finally, the discrimination impact using these models needs to be clinically meaningful. Several studies based their performance assessment on AUC. Clinically meaningful size effects like decision curve analysis should be included more routinely.

## 5. Conclusions

Our review has summarized the predictive performance of non-invasive imaging ML applications in EC patients. Recent advances and future perspectives of the ML technique demonstrate its potential to provide novel quantitative imaging markers in medical imaging. A few recommendations are made to improve study design and future applicability.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/cancers13102469/s1>, Table S1: Systematic search.

**Author Contributions:** C.-Y.X., C.-L.P., and V.V. contributed to literature search, figures, study design, data collection, data analysis, data interpretation, manuscript editing, and writing; B.C. and E.Y.-Y.W. contributed to data collection, data analysis, data interpretation, manuscript editing, and writing; Q.D. contributed to data collection, data analysis, data interpretation, manuscript editing, and writing. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Primary data cited in this review are openly available in PubMed, Embase and Cochrane database.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Jemal, A.; Bray, F.; Center, M.M.; Ferlay, J.; Ward, E.; Forman, D. Global Cancer Statistics. *CA Cancer J. Clin.* **2011**, *61*, 69–90. [[CrossRef](#)]
2. Pennathur, A.; Gibson, M.K.; A Jobe, B.; Luketich, J.D. Oesophageal carcinoma. *Lancet* **2013**, *381*, 400–412. [[CrossRef](#)]
3. Van Hagen, P.; Hulshof, M.; Van Lanschot, J.; Steyerberg, E.; Henegouwen, M.V.B.; Wijnhoven, B.; Richel, D.; Nieuwenhuijzen, G.A.; Hospers, G.A.P.; Bonenkamp, J.; et al. Preoperative Chemoradiotherapy for Esophageal or Junctional Cancer. *N. Engl. J. Med.* **2012**, *366*, 2074–2084. [[CrossRef](#)] [[PubMed](#)]
4. Yang, H.; Liu, H.; Chen, Y.; Zhu, C.; Fang, W.; Yu, Z.; Mao, W.; Xiang, J.; Han, Y.; Chen, Z.; et al. Neoadjuvant Chemoradiotherapy Followed by Surgery Versus Surgery Alone for Locally Advanced Squamous Cell Carcinoma of the Esophagus (NEOCRTEC5010): A Phase III Multicenter, Randomized, Open-Label Clinical Trial. *J. Clin. Oncol.* **2018**, *36*, 2796–2803. [[CrossRef](#)]
5. Barbetta, A.; Sihag, S.; Nobel, T.; Hsu, M.; Tan, K.S.; Bains, M.; Jones, D.R.; Molena, D. Patterns and risk of recurrence in patients with esophageal cancer with a pathologic complete response after chemoradiotherapy followed by surgery. *J. Thorac. Cardiovasc. Surg.* **2019**, *157*, 1249–1259.e5. [[CrossRef](#)]
6. Gwynne, S.; Wijnhoven, B.; Hulshof, M.; Bateman, A. Role of Chemoradiotherapy in Oesophageal Cancer—Adjuvant and Neoadjuvant Therapy. *Clin. Oncol.* **2014**, *26*, 522–532. [[CrossRef](#)] [[PubMed](#)]
7. Lin, J.; Kligerman, S.; Goel, R.; Sajedi, P.; Suntharalingam, M.; Chuong, M.D. State-of-the-art molecular imaging in esophageal cancer management: Implications for diagnosis, prognosis, and treatment. *J. Gastrointest. Oncol.* **2015**, *6*, 3–19.
8. Li, B.; Li, N.; Liu, S.; Li, Y.; Qian, B.; Zhang, Y.; He, H.; Chen, X.; Sun, Y.; Xiang, J.; et al. Does [18F] fluorodeoxyglucose-positron emission tomography/computed tomography have a role in cervical nodal staging for esophageal squamous cell carcinoma? *J. Thorac. Cardiovasc. Surg.* **2020**, *160*, 544–550. [[CrossRef](#)]

9. 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](#)]
10. Shimizu, H.; Nakayama, K.I. Artificial intelligence in oncology. *Cancer Sci.* **2020**, *111*, 1452–1460. [[CrossRef](#)]
11. Kermany, D.S.; Goldbaum, M.; Cai, W.; Valentim, C.C.; Liang, H.; Baxter, S.L.; McKeown, A.; Yang, G.; Wu, X.; Yan, F.; et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell* **2018**, *172*, 1122–1131.e9. [[CrossRef](#)] [[PubMed](#)]
12. Zhu, W.; Xie, L.; Han, J.; Guo, X. The Application of Deep Learning in Cancer Prognosis Prediction. *Cancers* **2020**, *12*, 603. [[CrossRef](#)]
13. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **2017**, *542*, 115–118. [[CrossRef](#)]
14. Poplin, R.; Varadarajan, A.V.; Blumer, K.; Liu, Y.; McConnell, M.V.; Corrado, G.S.; Peng, L.; Webster, D.R. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat. Biomed. Eng.* **2018**, *2*, 158–164. [[CrossRef](#)] [[PubMed](#)]
15. Coudray, N.; Ocampo, P.S.; Sakellaropoulos, T.; Narula, N.; Snuderl, M.; Fenyö, D.; Moreira, A.L.; Razavian, N.; Tsirigos, A. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat. Med.* **2018**, *24*, 1559–1567. [[CrossRef](#)]
16. Dou, Q.; So, T.Y.; Jiang, M.; Liu, Q.; Vardhanabhuti, V.; Kaissis, G.; Li, Z.; Si, W.; Lee, H.H.C.; Yu, K.; et al. Federated deep learning for detecting COVID-19 lung abnormalities in CT: A privacy-preserving multinational validation study. *npj Digit. Med.* **2021**, *4*, 1–11. [[CrossRef](#)] [[PubMed](#)]
17. Ding, J.; Cao, P.; Chang, H.-C.; Gao, Y.; Chan, S.H.S.; Vardhanabhuti, V. Deep learning-based thigh muscle segmentation for reproducible fat fraction quantification using fat–water decomposition MRI. *Insights Imaging* **2020**, *11*, 1–11. [[CrossRef](#)]
18. Kocak, B.; Durmaz, E.S.; Ates, E.; Kilickesmez, O. Radiomics with artificial intelligence: A practical guide for beginners. *Diagn. Interv. Radiol.* **2019**, *25*, 485–495. [[CrossRef](#)]
19. Committee on the Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials; Board on Health Care, Services; Board on Health Sciences Policy; Institute of Medicine Evolution of Translational Omics; Micheel, C.M.; Nass, S.J.; Omenn, G.S. *Evolution of Translational Omics: Lessons Learned and the Path Forward*; National Academies Press: Washington, DC, USA, 2012. [[CrossRef](#)]
20. Gerlinger, M.; Rowan, A.J.; Horswell, S.; Larkin, J.; Endesfelder, D.; Gronroos, E.; Martinez, P.; Matthews, N.; Stewart, A.; Tarpey, P.; et al. Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *N. Engl. J. Med.* **2012**, *366*, 883–892. [[CrossRef](#)]
21. Liang, Z.-G.; Tan, H.Q.; Zhang, F.; Tan, L.K.R.; Lin, L.; Lenkiewicz, J.; Wang, H.; Ong, E.H.W.; Kusumawidjaja, G.; Phua, J.H.; et al. Comparison of radiomics tools for image analyses and clinical prediction in nasopharyngeal carcinoma. *Br. J. Radiol.* **2019**, *92*, 20190271. [[CrossRef](#)]
22. Zwanenburg, A.; Vallières, M.; Abdalah, M.A.; Aerts, H.J.W.L.; Andrearczyk, V.; Apte, A.; Ashrafinia, S.; Bakas, S.; Beukinga, R.J.; Boellaard, R.; et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. *Radiology* **2020**, *295*, 328–338. [[CrossRef](#)] [[PubMed](#)]
23. Gillies, R.J.; Kinahan, P.E.; Hricak, H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology* **2016**, *278*, 563–577. [[CrossRef](#)]
24. Lambin, P.; Leijenaar, R.T.; Deist, T.M.; Peerlings, J.; De Jong, E.E.; Van Timmeren, J.; Sanduleanu, S.; LaRue, R.T.; Even, A.J.; Jochems, A.; et al. Radiomics: The bridge between medical imaging and personalized medicine. *Nat. Rev. Clin. Oncol.* **2017**, *14*, 749–762. [[CrossRef](#)] [[PubMed](#)]
25. Mayerhoefer, M.E.; Materka, A.; Langs, G.; Häggström, I.; Szczypiński, P.; Gibbs, P.; Cook, G. Introduction to Radiomics. *J. Nucl. Med.* **2020**, *61*, 488–495. [[CrossRef](#)] [[PubMed](#)]
26. Yang, L.; Gu, D.; Wei, J.; Yang, C.; Rao, S.; Wang, W.; Chen, C.; Ding, Y.; Tian, J.; Zeng, M. A Radiomics Nomogram for Preoperative Prediction of Microvascular Invasion in Hepatocellular Carcinoma. *Liver Cancer* **2019**, *8*, 373–386. [[CrossRef](#)]
27. Peeken, J.C.; Shouman, M.A.; Kroenke, M.; Rauscher, I.; Maurer, T.; Gschwend, J.E.; Eiber, M.; Combs, S.E. A CT-based radiomics model to detect prostate cancer lymph node metastases in PSMA radioguided surgery patients. *Eur. J. Nucl. Med. Mol. Imaging* **2020**, *47*, 2968–2977. [[CrossRef](#)]
28. Sun, R.; Limkin, E.J.; Vakalopoulou, M.; Derclé, L.; Champiat, S.; Han, S.R.; Verlingue, L.; Brandao, D.; Lancia, A.; Ammari, S.; et al. A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: An imaging biomarker, retrospective multicohort study. *Lancet Oncol.* **2018**, *19*, 1180–1191. [[CrossRef](#)]
29. Collarino, A.; Garganese, G.; Fragomeni, S.M.; Arias-Bouda, L.M.P.; Ieria, F.P.; Boellaard, R.; Rufini, V.; De Geus-Oei, L.-F.; Scambia, G.; Olmos, R.A.V.; et al. Radiomics in Vulvar Cancer: First Clinical Experience Using 18F-FDG PET/CT Images. *J. Nucl. Med.* **2018**, *60*, 199–206. [[CrossRef](#)]
30. Yushkevich, P.A.; Gao, Y.; Gerig, G. ITK-SNAP: An interactive tool for semi-automatic segmentation of multi-modality biomedical images. In Proceedings of the 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL, USA, 16–20 August 2016; Institute of Electrical and Electronics Engineers (IEEE): Orlando, FL, USA, 2016; Volume 2016, pp. 3342–3345.

31. Nioche, C.; Orlhac, F.; Boughdad, S.; Reuzé, S.; Goya-Outi, J.; Robert, C.; Pellot-Barakat, C.; Soussan, M.; Frouin, F.; Buvat, I. LIFEX: A Freeware for Radiomic Feature Calculation in Multimodality Imaging to Accelerate Advances in the Characterization of Tumor Heterogeneity. *Cancer Res.* **2018**, *78*, 4786–4789. [[CrossRef](#)]
32. Wolf, I.; Vetter, M.; Wegner, I.; Nolden, M.; Böttger, T.; Hastenteufel, M.; Schöbinger, M.; Kunert, T.; Meinzer, H.-P. The medical imaging interaction toolkit (MITK): A toolkit facilitating the creation of interactive software by extending VTK and ITK. *Medical Imaging* **2004**, 5367, 16–27. [[CrossRef](#)]
33. Abramoff, M.D.; Magalhães, P.J.; Ram, S.J. Image processing with ImageJ. *Biophotonics Int.* **2004**, *11*, 36–42.
34. Fedorov, A.; Beichel, R.; Kalpathy-Cramer, J.; Finet, J.; Fillion-Robin, J.-C.; Pujol, S.; Bauer, C.; Jennings, M.; Fennessy, F.; Sonka, M.; et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn. Reson. Imaging* **2012**, *30*, 1323–1341. [[CrossRef](#)]
35. Velazquez, E.R.; Aerts, H.J.; Gu, Y.; Goldgof, D.B.; De Ruyscher, D.; Dekker, A.; Korn, R.; Gillies, R.J.; Lambin, P. A semiautomatic CT-based ensemble segmentation of lung tumors: Comparison with oncologists' delineations and with the surgical specimen. *Radiother. Oncol.* **2012**, *105*, 167–173. [[CrossRef](#)]
36. Traverso, A.; Wee, L.; Dekker, A.; Gillies, R. Repeatability and Reproducibility of Radiomic Features: A Systematic Review. *Int. J. Radiat. Oncol.* **2018**, *102*, 1143–1158. [[CrossRef](#)]
37. Parmar, C.; Velazquez, E.R.; Leijenaar, R.; Jermoumi, M.; Carvalho, S.; Mak, R.H.; Mitra, S.; Shankar, B.U.; Kikinis, R.; Haibe-Kains, B.; et al. Robust Radiomics Feature Quantification Using Semiautomatic Volumetric Segmentation. *PLoS ONE* **2014**, *9*, e102107. [[CrossRef](#)] [[PubMed](#)]
38. Nestle, U.; Kremp, S.; Schaefer-Schuler, A.; Sebastian-Welsch, C.; Hellwig, D.; Rube, C.; Kirsch, C.-M. Comparison of different methods for delineation of 18F-FDG PET-positive tissue for target volume definition in radiotherapy of patients with non-small cell lung cancer. *J. Nucl. Med.* **2005**, *46*, 1342–1348.
39. Schelb, P.; Kohl, S.; Radtke, J.P.; Wiesenfarth, M.; Kickingeder, P.V.Ñ.; Bickelhaupt, S.; Kuder, T.A.; Stenzinger, A.; Hohenfellner, M.; Schlemmer, H.-P.; et al. Classification of Cancer at Prostate MRI: Deep Learning versus Clinical PI-RADS Assessment. *Radiol.* **2019**, *293*, 607–617. [[CrossRef](#)]
40. Jin, J.; Zhu, H.; Zhang, J.; Ai, Y.; Zhang, J.; Teng, Y.; Xie, C.; Jin, X. Multiple U-Net-Based Automatic Segmentations and Radiomics Feature Stability on Ultrasound Images for Patients with Ovarian Cancer. *Front. Oncol.* **2021**, *10*, 614201. [[CrossRef](#)]
41. Ronneberger, O.; Fischer, P.; Brox, T.; Navab, N.; Hornegger, J.; Wells, W.; Frangi, A. *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Springer-Verlag: Berlin/Heidelberg, Germany, 2015.
42. Van Griethuysen, J.J.; Fedorov, A.; Parmar, C.; Hosny, A.; Aucoin, N.; Narayan, V.; Beets-Tan, R.G.; Fillion-Robin, J.-C.; Pieper, S.; Aerts, H.J. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res.* **2017**, *77*, e104–e107. [[CrossRef](#)]
43. Ortiz-Ramon, R.; Larroza, A.; Arana, E.; Moratal, D. A radiomics evaluation of 2D and 3D MRI texture features to classify brain metastases from lung cancer and melanoma. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **2017**, 2017, 493–496. [[CrossRef](#)]
44. Shen, C.; Liu, Z.; Guan, M.; Song, J.; Lian, Y.; Wang, S.; Tang, Z.; Dong, D.; Kong, L.; Wang, M.; et al. 2D and 3D CT Radiomics Features Prognostic Performance Comparison in Non-Small Cell Lung Cancer. *Transl. Oncol.* **2017**, *10*, 886–894. [[CrossRef](#)]
45. Yang, L.; Yang, J.; Zhou, X.; Huang, L.; Zhao, W.; Wang, T.; Zhuang, J.; Tian, J. Development of a radiomics nomogram based on the 2D and 3D CT features to predict the survival of non-small cell lung cancer patients. *Eur. Radiol.* **2018**, *29*, 2196–2206. [[CrossRef](#)]
46. Orlhac, F.; Frouin, F.; Nioche, C.; Ayache, N.; Buvat, I. Validation of A Method to Compensate Multicenter Effects Affecting CT Radiomics. *Radiology* **2019**, *291*, 53–59. [[CrossRef](#)] [[PubMed](#)]
47. Da-Ano, R.; Masson, I.; Lucia, F.; Doré, M.; Robin, P.; Alfieri, J.; Rousseau, C.; Mervoyer, A.; Reinhold, C.; Castelli, J.; et al. Performance comparison of modified ComBat for harmonization of radiomic features for multicenter studies. *Sci. Rep.* **2020**, *10*, 1–12. [[CrossRef](#)] [[PubMed](#)]
48. Brunzell, H.; Eriksson, J. Feature reduction for classification of multidimensional data. *Pattern Recognit.* **2000**, *33*, 1741–1748. [[CrossRef](#)]
49. Xu, L.; Yang, P.; Liang, W.; Liu, W.; Wang, W.; Luo, C.; Wang, J.; Peng, Z.; Xing, L.; Huang, M.; et al. A radiomics approach based on support vector machine using MR images for preoperative lymph node status evaluation in intrahepatic cholangiocarcinoma. *Theranostics* **2019**, *9*, 5374–5385. [[CrossRef](#)]
50. Wang, H.; Chen, H.; Duan, S.; Hao, D.; Liu, J. Radiomics and Machine Learning With Multiparametric Preoperative MRI May Accurately Predict the Histopathological Grades of Soft Tissue Sarcomas. *J. Magn. Reson. Imaging* **2020**, *51*, 791–797. [[CrossRef](#)]
51. Bizzego, A.; Bussola, N.; Chierici, M.; Maggio, V.; Francescatto, M.; Cima, L.; Cristoforetti, M.; Jurman, G.; Furlanello, C. Evaluating reproducibility of AI algorithms in digital pathology with DAPPER. *PLoS Comput. Biol.* **2019**, *15*, e1006269. [[CrossRef](#)]
52. Furlanello, C.; Serafini, M.; Merler, S.; Jurman, G. Entropy-based gene ranking without selection bias for the predictive classification of microarray data. *BMC Bioinform.* **2003**, *4*, 54. [[CrossRef](#)]
53. Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **1996**, *58*, 267–288. [[CrossRef](#)]
54. Ringnér, M. What is principal component analysis? *Nat. Biotechnol.* **2008**, *26*, 303–304. [[CrossRef](#)] [[PubMed](#)]
55. Balakrishnama, S.; Ganapathiraju, A. *Linear Discriminant Analysis—A Brief Tutorial*; Mississippi State University: Mississippi State, MS, USA, 1998; Volume 18, pp. 1–8.
56. Kleinbaum, D.G.; Dietz, K.; Gail, M.; Klein, M.; Klein, M. *Logistic Regression*, 2nd ed.; Springer: New York, NY, USA, 2002.

57. Xie, C.; Ng, M.-Y.; Ding, J.; Leung, S.T.; Lo, C.S.Y.; Wong, H.Y.F.; Vardhanabhuti, V. Discrimination of pulmonary ground-glass opacity changes in COVID-19 and non-COVID-19 patients using CT radiomics analysis. *Eur. J. Radiol. Open* **2020**, *7*, 100271. [[CrossRef](#)] [[PubMed](#)]
58. Zhang, B.; He, X.; Ouyang, F.; Gu, D.; Dong, Y.; Zhang, L.; Mo, X.; Huang, W.; Tian, J.; Zhang, S. Radiomic machine-learning classifiers for prognostic biomarkers of advanced nasopharyngeal carcinoma. *Cancer Lett.* **2017**, *403*, 21–27. [[CrossRef](#)] [[PubMed](#)]
59. St, J.P.O.E.D.; Dumais, S.; Osuna, E.; Platt, J.; Scholkopf, B. Support vector machines. *IEEE Intell. Syst.* **1998**, *13*, 18–28. [[CrossRef](#)]
60. Peterson, L.E. K-nearest neighbor. *Scholarpedia* **2009**, *4*, 1883. [[CrossRef](#)]
61. Rokach, L.; Maimon, O. Decision Trees. In *Data Mining and Knowledge Discovery Handbook*; Springer: Boston, MA, USA, 2005; pp. 165–192.
62. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
63. Zhang, Y.; Oikonomou, A.; Wong, A.; Haider, M.A.; Khalvati, F. Radiomics-based Prognosis Analysis for Non-Small Cell Lung Cancer. *Sci. Rep.* **2017**, *7*, srep46349. [[CrossRef](#)]
64. Parmar, C.; Grossmann, P.; Bussink, J.; Lambin, P.; Aerts, H.J.W.L. Machine Learning methods for Quantitative Radiomic Biomarkers. *Sci. Rep.* **2015**, *5*, 13087. [[CrossRef](#)]
65. Xie, C.; Du, R.; Ho, J.W.; Pang, H.H.; Chiu, K.W.; Lee, E.Y.; Vardhanabhuti, V. Effect of machine learning re-sampling techniques for imbalanced datasets in 18F-FDG PET-based radiomics model on prognostication performance in cohorts of head and neck cancer patients. *Eur. J. Nucl. Med. Mol. Imaging* **2020**, *47*, 2826–2835. [[CrossRef](#)]
66. Iasonos, A.; Schrag, D.; Raj, G.V.; Panageas, K.S. How to Build and Interpret a Nomogram for Cancer Prognosis. *J. Clin. Oncol.* **2008**, *26*, 1364–1370. [[CrossRef](#)]
67. Yamaoka, K.; Nakagawa, T.; Uno, T. Application of Akaike's information criterion (AIC) in the evaluation of linear pharmacokinetic equations. *J. Pharmacokin. Biopharm.* **1978**, *6*, 165–175. [[CrossRef](#)]
68. Wels, M.; Carneiro, G.; Aplas, A.; Huber, M.; Hornegger, J.; Comaniciu, D. A Discriminative Model-Constrained Graph Cuts Approach to Fully Automated Pediatric Brain Tumor Segmentation in 3-D MRI. In *Proceedings of the Computer Vision*; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 2008; Volume 11, pp. 67–75.
69. Shi, Z.; Miao, C.; Schoepf, U.J.; Savage, R.H.; Dargis, D.M.; Pan, C.; Chai, X.; Li, X.L.; Xia, S.; Zhang, X.; et al. A clinically applicable deep-learning model for detecting intracranial aneurysm in computed tomography angiography images. *Nat. Commun.* **2020**, *11*, 1–11. [[CrossRef](#)]
70. Zheng, X.; Yao, Z.; Huang, Y.; Yu, Y.; Wang, Y.; Liu, Y.; Mao, R.; Li, F.; Xiao, Y.; Wang, Y.; et al. Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer. *Nat. Commun.* **2020**, *11*, 1236. [[CrossRef](#)] [[PubMed](#)]
71. Jones, C.M.; Athanasiou, T. Summary Receiver Operating Characteristic Curve Analysis Techniques in the Evaluation of Diagnostic Tests. *Ann. Thorac. Surg.* **2005**, *79*, 16–20. [[CrossRef](#)]
72. Bizzego, A.; Bussola, N.; Salvalai, D.; Chierici, M.; Maggio, V.; Jurman, G.; Furlanello, C. Integrating deep and radiomics features in cancer bioimaging T2. In *Proceedings of the 2019 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, Siena, Italy, 9–11 July 2019. [[CrossRef](#)]
73. Ozenne, B.; Subtil, F.; Maucourt-Boulch, D. The precision–recall curve overcame the optimism of the receiver operating characteristic curve in rare diseases. *J. Clin. Epidemiol.* **2015**, *68*, 855–859. [[CrossRef](#)] [[PubMed](#)]
74. Shah, N.D.; Steyerberg, E.W.; Kent, D.M. Big Data and Predictive Analytics. *JAMA* **2018**, *320*, 27–28. [[CrossRef](#)]
75. Shah, N.H.; Milstein, A.; Bagley, S.C. Making Machine Learning Models Clinically Useful. *JAMA* **2019**, *322*, 1351. [[CrossRef](#)]
76. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)]
77. Chen, S.; Ma, K.; Zheng, Y. Med3d: Transfer learning for 3d medical image analysis. *arXiv* **2019**, arXiv:1904.00625.
78. Kamnitsas, K.; Ledig, C.; Newcombe, V.; Simpson, J.P.; Kane, A.D.; Menon, D.K.; Rueckert, D.; Glocker, B. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med. Image Anal.* **2017**, *36*, 61–78. [[CrossRef](#)]
79. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **2012**, *60*, 1097–1105. [[CrossRef](#)]
80. Kim, Y.J.; Bae, J.P.; Chung, J.W.; Park, D.K.; Kim, K.G.; Kim, Y.J. New polyp image classification technique using transfer learning of network-in-network structure in endoscopic images. *Sci. Rep.* **2021**, *11*, 3605. [[CrossRef](#)] [[PubMed](#)]
81. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. *Conf. Proc.* **2016**, 2818–2826. [[CrossRef](#)]
82. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition T2. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 27–30 June 2016. [[CrossRef](#)]
83. Hu, Y.; Xie, C.; Yang, H.; Ho, J.W.; Wen, J.; Han, L.; Lam, K.-O.; Wong, I.Y.; Law, S.Y.; Chiu, K.W.; et al. Computed tomography-based deep-learning prediction of neoadjuvant chemoradiotherapy treatment response in esophageal squamous cell carcinoma. *Radiother. Oncol.* **2021**, *154*, 6–13. [[CrossRef](#)] [[PubMed](#)]
84. Shorten, C.; Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60. [[CrossRef](#)]
85. Perez, L.; Wang, J. The effectiveness of data augmentation in image classification using deep learning. *arXiv* **2017**, arXiv:1712.04621.
86. Huynh, B.Q.; Li, H.; Giger, M.L. Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. *J. Med. Imaging* **2016**, *3*, 034501. [[CrossRef](#)] [[PubMed](#)]

87. Shin, H.-C.; Roth, H.R.; Gao, M.; Lu, L.; Xu, Z.; Nogues, I.; Yao, J.; Mollura, D.; Summers, R.M. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Trans. Med. Imaging* **2016**, *35*, 1285–1298. [[CrossRef](#)]
88. Wu, L.; Yang, X.; Cao, W.; Zhao, K.; Li, W.; Ye, W.; Chen, X.; Zhou, Z.; Liu, Z.; Liang, C. Multiple Level CT Radiomics Features Preoperatively Predict Lymph Node Metastasis in Esophageal Cancer: A Multicentre Retrospective Study. *Front. Oncol.* **2020**, *9*, 1548. [[CrossRef](#)]
89. Kontos, D.; Summers, R.M.; Giger, M.L. Special Section Guest Editorial: Radiomics and Deep Learning. *J. Med. Imaging* **2018**, *4*, 041301. [[CrossRef](#)]
90. Du, R.; Lee, V.H.; Yuan, H.; Lam, K.-O.; Pang, H.H.; Chen, Y.; Lam, E.Y.; Khong, P.-L.; Lee, A.W.; Kwong, D.L.; et al. Radiomics Model to Predict Early Progression of Nonmetastatic Nasopharyngeal Carcinoma after Intensity Modulation Radiation Therapy: A Multicenter Study. *Radiol. Artif. Intell.* **2019**, *1*, e180075. [[CrossRef](#)] [[PubMed](#)]
91. Papadimitroulas, P.; Brocki, L.; Chung, N.C.; Marchadour, W.; Vermet, F.; Gaubert, L.; Eleftheriadis, V.; Plachouris, D.; Visvikis, D.; Kagadis, G.C.; et al. Artificial intelligence: Deep learning in oncological radiomics and challenges of interpretability and data harmonization. *Phys. Medica* **2021**, *83*, 108–121. [[CrossRef](#)] [[PubMed](#)]
92. Yun, J.; Park, J.E.; Lee, H.; Ham, S.; Kim, N.; Kim, H.S. Radiomic features and multilayer perceptron network classifier: A robust MRI classification strategy for distinguishing glioblastoma from primary central nervous system lymphoma. *Sci. Rep.* **2019**, *9*, 1–10. [[CrossRef](#)]
93. Hosny, A.; Aerts, H.J.; Mak, R.H. Handcrafted versus deep learning radiomics for prediction of cancer therapy response. *Lancet Digit. Health* **2019**, *1*, e106–e107. [[CrossRef](#)]
94. Cai, J.; Zheng, J.; Shen, J.; Yuan, Z.; Xie, M.; Gao, M.; Tan, H.; Liang, Z.-G.; Rong, X.; Li, Y.; et al. A Radiomics Model for Predicting the Response to Bevacizumab in Brain Necrosis after Radiotherapy. *Clin. Cancer Res.* **2020**, *26*, 5438–5447. [[CrossRef](#)]
95. Su, X.; Chen, N.; Sun, H.; Liu, Y.; Yang, X.; Wang, W.; Zhang, S.; Tan, Q.; Su, J.; Gong, Q.; et al. Automated machine learning based on radiomics features predicts H3 K27M mutation in midline gliomas of the brain. *Neuro-Oncol.* **2019**, *22*, 393–401. [[CrossRef](#)]
96. Wang, S.; Zha, Y.; Li, W.; Wu, Q.; Li, X.; Niu, M.; Wang, M.; Qiu, X.; Li, H.; Yu, H.; et al. A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis. *Eur. Respir. J.* **2020**, *56*, 2000775. [[CrossRef](#)] [[PubMed](#)]
97. Zhang, H.-T.; Zhang, J.-S.; Nan, Y.-D.; Zhao, Y.; Fu, E.-Q.; Xie, Y.-H.; Liu, W.; Li, W.-P.; Zhang, H.-J.; Jiang, H.; et al. Automated detection and quantification of COVID-19 pneumonia: CT imaging analysis by a deep learning-based software. *Eur. J. Nucl. Med. Mol. Imaging* **2020**, *47*, 1–8. [[CrossRef](#)] [[PubMed](#)]
98. Tan, H.-B.; Xiong, F.; Jiang, Y.-L.; Huang, W.-C.; Wang, Y.; Li, H.-H.; You, T.; Fu, T.-T.; Lu, R.; Peng, B.-W. The study of automatic machine learning base on radiomics of non-focus area in the first chest CT of different clinical types of COVID-19 pneumonia. *Sci. Rep.* **2020**, *10*, 1–10. [[CrossRef](#)]
99. Mandrekar, J.N. Receiver Operating Characteristic Curve in Diagnostic Test Assessment. *J. Thorac. Oncol.* **2010**, *5*, 1315–1316. [[CrossRef](#)]
100. Cao, Q.; Li, Y.; Li, Z.; An, D.; Li, B.; Lin, Q. Development and validation of a radiomics signature on differentially expressed features of 18F-FDG PET to predict treatment response of concurrent chemoradiotherapy in thoracic esophagus squamous cell carcinoma. *Radiother. Oncol.* **2020**, *146*, 9–15. [[CrossRef](#)] [[PubMed](#)]
101. Hu, Y.; Xie, C.; Yang, H.; Ho, J.W.K.; Wen, J.; Han, L.; Chiu, K.W.H.; Fu, J.; Vardhanabhuti, V. Assessment of Intratumoral and Peritumoral Computed Tomography Radiomics for Predicting Pathological Complete Response to Neoadjuvant Chemoradiation in Patients With Esophageal Squamous Cell Carcinoma. *JAMA Netw. Open* **2020**, *3*, e2015927. [[CrossRef](#)]
102. Yang, Z.; He, B.; Zhuang, X.; Gao, X.; Wang, D.; Li, M.; Lin, Z.; Luo, R. CT-based radiomic signatures for prediction of pathologic complete response in esophageal squamous cell carcinoma after neoadjuvant chemoradiotherapy. *J. Radiat. Res.* **2019**, *60*, 538–545. [[CrossRef](#)] [[PubMed](#)]
103. Hou, Z.; Li, S.; Ren, W.; Liu, J.; Yan, J.; Wan, S. Radiomic analysis in T2W and SPAIR T2W MRI: Predict treatment response to chemoradiotherapy in esophageal squamous cell carcinoma. *J. Thorac. Dis.* **2018**, *10*, 2256–2267. [[CrossRef](#)] [[PubMed](#)]
104. Beukinga, R.J.; Hulshoff, J.B.; Mul, V.E.M.; Noordzij, W.; Kats-Ugurlu, G.; Slart, R.H.J.A.; Plukker, J.T.M. Prediction of Response to Neoadjuvant Chemotherapy and Radiation Therapy with Baseline and Restaging 18F-FDG PET Imaging Biomarkers in Patients with Esophageal Cancer. *Radiology* **2018**, *287*, 983–992. [[CrossRef](#)] [[PubMed](#)]
105. Hou, Z.; Ren, W.; Li, S.; Liu, J.; Sun, Y.; Yan, J.; Wan, S. Radiomic analysis in contrast-enhanced CT: Predict treatment response to chemoradiotherapy in esophageal carcinoma. *Oncotarget* **2017**, *8*, 104444–104454. [[CrossRef](#)] [[PubMed](#)]
106. Van Rossum, P.S.; Fried, D.V.; Zhang, L.; Hofstetter, W.L.; Van Vulpen, M.; Meijer, G.J.; Court, L.E.; Lin, S.H. The Incremental Value of Subjective and Quantitative Assessment of 18F-FDG PET for the Prediction of Pathologic Complete Response to Preoperative Chemoradiotherapy in Esophageal Cancer. *J. Nucl. Med.* **2016**, *57*, 691–700. [[CrossRef](#)]
107. Beukinga, R.J.; Hulshoff, J.B.; Van Dijk, L.V.; Muijs, C.T.; Burgerhof, J.G.M.; Kats-Ugurlu, G.; Slart, R.H.J.A.; Slump, C.H.; Mul, V.E.M.; Plukker, J.T.M. Predicting Response to Neoadjuvant Chemoradiotherapy in Esophageal Cancer with Textural Features Derived from Pretreatment 18F-FDG PET/CT Imaging. *J. Nucl. Med.* **2017**, *58*, 723–729. [[CrossRef](#)]
108. Desbordes, P.; Ruan, S.; Modzelewski, R.; Pineau, P.; Vauclin, S.; Gouel, P.; Michel, P.; Di Fiore, F.; Vera, P.; Gardin, I. Predictive value of initial FDG-PET features for treatment response and survival in esophageal cancer patients treated with chemo-radiation therapy using a random forest classifier. *PLoS ONE* **2017**, *12*, e0173208. [[CrossRef](#)]

109. Ypsilantis, P.-P.; Siddique, M.; Sohn, H.-M.; Davies, A.; Cook, G.; Goh, V.; Montana, G. Predicting Response to Neoadjuvant Chemotherapy with PET Imaging Using Convolutional Neural Networks. *PLoS ONE* **2015**, *10*, e0137036. [[CrossRef](#)]
110. Zhang, H.; Tan, S.; Chen, W.; Kligerman, S.; Kim, G.; D'Souza, W.D.; Suntharalingam, M.; Lu, W. Modeling Pathologic Response of Esophageal Cancer to Chemoradiation Therapy Using Spatial-Temporal 18F-FDG PET Features, Clinical Parameters, and Demographics. *Int. J. Radiat. Oncol.* **2014**, *88*, 195–203. [[CrossRef](#)] [[PubMed](#)]
111. Qiu, Q.; Duan, J.; Deng, H.; Han, Z.; Gu, J.; Yue, N.J.; Yin, Y. Development and Validation of a Radiomics Nomogram Model for Predicting Postoperative Recurrence in Patients With Esophageal Squamous Cell Cancer Who Achieved pCR After Neoadjuvant Chemoradiotherapy Followed by Surgery. *Front. Oncol.* **2020**, *10*, 1398. [[CrossRef](#)] [[PubMed](#)]
112. Chen, Y.-H.; Lue, K.-H.; Chu, S.-C.; Chang, B.-S.; Wang, L.-Y.; Liu, D.-W.; Liu, S.-H.; Chao, Y.-K.; Chan, S.-C. Combining the radiomic features and traditional parameters of 18F-FDG PET with clinical profiles to improve prognostic stratification in patients with esophageal squamous cell carcinoma treated with neoadjuvant chemoradiotherapy and surgery. *Ann. Nucl. Med.* **2019**, *33*, 657–670. [[CrossRef](#)] [[PubMed](#)]
113. Yang, C.-K.; Yeh, J.C.-Y.; Yu, W.-H.; Chien, L.-I.; Lin, K.-H.; Huang, W.-S.; Hsu, P.-K. Deep Convolutional Neural Network-Based Positron Emission Tomography Analysis Predicts Esophageal Cancer Outcome. *J. Clin. Med.* **2019**, *8*, 844. [[CrossRef](#)]
114. Xie, C.; Yang, P.; Zhang, X.; Xu, L.; Wang, X.; Li, X.; Zhang, L.; Xie, R.; Yang, L.; Jing, Z.; et al. Sub-region based radiomics analysis for survival prediction in esophageal tumours treated by definitive concurrent chemoradiotherapy. *EBioMedicine* **2019**, *44*, 289–297. [[CrossRef](#)] [[PubMed](#)]
115. LaRue, R.T.H.M.; Klaassen, R.; Jochems, A.; Leijenaar, R.T.H.; Hulshof, M.C.C.M.; Henegouwen, M.I.V.B.; Schreurs, W.M.J.; Sosef, M.N.; Van Elmpt, W.; Van Laarhoven, H.W.M.; et al. Pre-treatment CT radiomics to predict 3-year overall survival following chemoradiotherapy of esophageal cancer. *Acta Oncol.* **2018**, *57*, 1475–1481. [[CrossRef](#)]
116. Foley, K.G.; Hills, R.; Berthon, B.; Marshall, C.; Parkinson, C.; Lewis, W.G.; Crosby, T.D.L.; Spezi, E.; Roberts, S.A. Development and validation of a prognostic model incorporating texture analysis derived from standardised segmentation of PET in patients with oesophageal cancer. *Eur. Radiol.* **2018**, *28*, 428–436. [[CrossRef](#)]
117. Xiong, J.; Yu, W.; Ma, J.; Ren, Y.; Fu, X.; Zhao, J. The Role of PET-Based Radiomic Features in Predicting Local Control of Esophageal Cancer Treated with Concurrent Chemoradiotherapy. *Sci. Rep.* **2018**, *8*, 1–11. [[CrossRef](#)]
118. Qu, J.; Shen, C.; Qin, J.; Wang, Z.; Liu, Z.; Guo, J.; Zhang, H.; Gao, P.; Bei, T.; Wang, Y.; et al. The MR radiomic signature can predict preoperative lymph node metastasis in patients with esophageal cancer. *Eur. Radiol.* **2019**, *29*, 906–914. [[CrossRef](#)]
119. Tan, X.; Ma, Z.; Yan, L.; Ye, W.; Liu, Z.; Liang, C. Radiomics nomogram outperforms size criteria in discriminating lymph node metastasis in resectable esophageal squamous cell carcinoma. *Eur. Radiol.* **2019**, *29*, 392–400. [[CrossRef](#)]
120. Shen, C.; Liu, Z.; Wang, Z.; Guo, J.; Zhang, H.; Wang, Y.; Qin, J.; Li, H.; Fang, M.; Tang, Z.; et al. Building CT Radiomics Based Nomogram for Preoperative Esophageal Cancer Patients Lymph Node Metastasis Prediction. *Transl. Oncol.* **2018**, *11*, 815–824. [[CrossRef](#)]
121. Li, X.-F.; Wang, Q.; Duan, S.-F.; Yao, B.; Liu, C.-Y. Heterogeneity of T3 stage esophageal squamous cell carcinoma in different parts based on enhanced CT radiomics. *Medicine* **2020**, *99*, e21470. [[CrossRef](#)] [[PubMed](#)]
122. Ou, J.; Li, R.; Zeng, R.; Wu, C.-Q.; Chen, Y.; Chen, T.-W.; Zhang, X.-M.; Wu, L.; Jiang, Y.; Yang, J.-Q.; et al. CT radiomic features for predicting resectability of oesophageal squamous cell carcinoma as given by feature analysis: A case control study. *Cancer Imaging* **2019**, *19*, 66. [[CrossRef](#)] [[PubMed](#)]
123. Hoshino, I.; Yokota, H.; Ishige, F.; Iwatate, Y.; Takeshita, N.; Nagase, H.; Uno, T.; Matsubara, H. Radiogenomics predicts the expression of microRNA-1246 in the serum of esophageal cancer patients. *Sci. Rep.* **2020**, *10*, 1–8. [[CrossRef](#)]
124. Daly, J.M.; Karnell, L.H.; Menck, H.R. National Cancer Data Base report on esophageal carcinoma. *Cancer* **1996**, *78*, 1820–1828. [[CrossRef](#)]
125. Ng, T.; Vezeridis, M.P. Advances in the surgical treatment of esophageal cancer. *J. Surg. Oncol.* **2010**, *101*, 725–729. [[CrossRef](#)] [[PubMed](#)]
126. Pasquali, S.; Yim, G.; Vohra, R.S.; Mocellin, S.; Nyanhongo, D.; Marriott, P.; Geh, J.I.; Griffiths, E.A. Survival After Neoadjuvant and Adjuvant Treatments Compared to Surgery Alone for Resectable Esophageal Carcinoma. *Ann. Surg.* **2017**, *265*, 481–491. [[CrossRef](#)] [[PubMed](#)]
127. Tepper, J.; Krasna, M.J.; Niedzwiecki, D.; Hollis, D.; Reed, C.E.; Goldberg, R.; Kiel, K.; Willett, C.; Sugarbaker, D.; Mayer, R. Phase III Trial of Trimodality Therapy With Cisplatin, Fluorouracil, Radiotherapy, and Surgery Compared With Surgery Alone for Esophageal Cancer: CALGB 9781. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **2008**, *26*, 1086–1092. [[CrossRef](#)] [[PubMed](#)]
128. Shapiro, J.; van Lanschot, J.J.B.; Hulshof, M.C.; van Hagen, P.; van Berge Henegouwen, M.I.; Wijnhoven, B.P.L.; van Laarhoven, H.W.M.; Nieuwenhuijzen, G.A.P.; Hospers, G.A.P.; Bonenkamp, J.J.; et al. Neoadjuvant chemoradiotherapy plus surgery versus surgery alone for oesophageal or junctional cancer (CROSS): Long-term results of a randomised controlled trial. *Lancet Oncol.* **2015**, *16*, 1090–1098. [[CrossRef](#)]
129. Eyck, B.M.; Onstenk, B.D.; Noordman, B.J.; Nieboer, D.; Spaander, M.C.W.; Valkema, R.; Lagarde, S.M.; Wijnhoven, B.P.L.; van Lanschot, J.J.B. Accuracy of Detecting Residual Disease After Neoadjuvant Chemoradiotherapy for Esophageal Cancer. *Ann. Surg.* **2020**, *271*, 245–256. [[CrossRef](#)]
130. Sugawara, K.; Yamashita, H.; Uemura, Y.; Mitsui, T.; Yagi, K.; Nishida, M.; Aikou, S.; Mori, K.; Nomura, S.; Seto, Y. Numeric pathologic lymph node classification shows prognostic superiority to topographic pN classification in esophageal squamous cell carcinoma. *Surgery* **2017**, *162*, 846–856. [[CrossRef](#)] [[PubMed](#)]



131. Gabriel, E.; Attwood, K.; Du, W.; Tuttle, R.; Alnaji, R.M.; Nurkin, S.J.; Malhotra, U.; Hochwald, S.N.; Kukar, M. Association Between Clinically Staged Node-Negative Esophageal Adenocarcinoma and Overall Survival Benefit from Neoadjuvant Chemoradiation. *JAMA Surg.* **2016**, *151*, 234–245. [[CrossRef](#)] [[PubMed](#)]
132. Rice, T.W.; Gress, D.M.; Patil, D.T.; Hofstetter, W.L.; Kelsen, D.P.; Blackstone, E.H. Cancer of the esophagus and esophagogastric junction—Major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *CA A Cancer J. Clin.* **2017**, *67*, 304–317. [[CrossRef](#)]
133. D'Journo, X.B. Clinical implication of the innovations of the 8th edition of the TNM classification for esophageal and esophago-gastric cancer. *J. Thorac. Dis.* **2018**, *10*, S2671–S2681. [[CrossRef](#)]
134. Choi, J.; Kim, S.G.; Kim, J.S.; Jung, H.C.; Song, I.S. Comparison of endoscopic ultrasonography (EUS), positron emission tomography (PET), and computed tomography (CT) in the preoperative locoregional staging of resectable esophageal cancer. *Surg. Endosc.* **2010**, *24*, 1380–1386. [[CrossRef](#)]
135. Kato, H.; Kuwano, H.; Nakajima, M.; Miyazaki, T.; Yoshikawa, M.; Ojima, H.; Tsukada, K.; Oriuchi, N.; Inoue, T.; Endo, K. Comparison between positron emission tomography and computed tomography in the use of the assessment of esophageal carcinoma. *Cancer* **2002**, *94*, 921–928. [[CrossRef](#)] [[PubMed](#)]
136. Malik, V.; Harmon, M.; Johnston, C.; Fagan, A.J.; Claxton, Z.; Ravi, N.; O'Toole, D.; Muldoon, C.; Keogan, M.; Reynolds, J.V.; et al. Whole Body MRI in the Staging of Esophageal Cancer—A Prospective Comparison with Whole Body 18F-FDG PET-CT. *Dig. Surg.* **2015**, *32*, 397–408. [[CrossRef](#)] [[PubMed](#)]
137. Rice, T.W.; Ishwaran, H.; Hofstetter, W.L.; Schipper, P.H.; Kesler, K.A.; Law, S.; Lerut, T.; Denlinger, C.E.; Salo, J.A.; Scott, W.J.; et al. Esophageal Cancer. *Ann. Surg.* **2017**, *265*, 122–129. [[CrossRef](#)]
138. Kutup, A.; Nentwich, M.F.; Bollschweiler, E.; Bogoevski, D.; Izbicki, J.R.; Hölscher, A.H. What Should Be the Gold Standard for the Surgical Component in the Treatment of Locally Advanced Esophageal Cancer. *Ann. Surg.* **2014**, *260*, 1016–1022. [[CrossRef](#)] [[PubMed](#)]
139. Rizk, N.P.; Ishwaran, H.; Rice, T.W.; Chen, L.-Q.; Schipper, P.H.; Kesler, K.A.; Law, S.; Lerut, T.E.M.R.; Reed, C.E.; Salo, J.A.; et al. Optimum Lymphadenectomy for Esophageal Cancer. *Ann. Surg.* **2010**, *251*, 46–50. [[CrossRef](#)]
140. Ye, T.; Sun, Y.; Zhang, Y.; Zhang, Y.; Chen, H. Three-Field or Two-Field Resection for Thoracic Esophageal Cancer: A Meta-Analysis. *Ann. Thorac. Surg.* **2013**, *96*, 1933–1941. [[CrossRef](#)] [[PubMed](#)]
141. Choy, G.; Khalilzadeh, O.; Michalski, M.; Synho, D.; Samir, A.E.; Pianykh, O.S.; Geis, J.R.; Pandharipande, P.V.; Brink, J.A.; Dreyer, K.J. Current Applications and Future Impact of Machine Learning in Radiology. *Radiology* **2018**, *288*, 318–328. [[CrossRef](#)] [[PubMed](#)]
142. Eisenhauer, E.A.; Therasse, P.; Bogaerts, J.; Schwartz, L.H.; Sargent, D.; Ford, R.; Dancey, J.; Arbuck, S.; Gwyther, S.; Mooney, M.; et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *Eur. J. Cancer* **2009**, *45*, 228–247. [[CrossRef](#)] [[PubMed](#)]
143. Yanagawa, M.; Tatsumi, M.; Miyata, H.; Morii, E.; Tomiyama, N.; Watabe, T.; Isohashi, K.; Kato, H.; Shimosegawa, E.; Yamasaki, M.; et al. Evaluation of Response to Neoadjuvant Chemotherapy for Esophageal Cancer: PET Response Criteria in Solid Tumors Versus Response Evaluation Criteria in Solid Tumors. *J. Nucl. Med.* **2012**, *53*, 872–880. [[CrossRef](#)]
144. Hatt, M.; Vallieres, M.; Visvikis, D.; Zwanenburg, A. IBSI: An international community radiomics standardization initiative. *J Nucl. Med.* **2018**, *59*, 287.