IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 14, 2019, accepted July 7, 2019, date of publication July 25, 2019, date of current version August 12, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2930891

Computer Vision in Esophageal Cancer: A Literature Review

INÊS DOMINGUES[1], INÊS LUCENA SAMPAIO[1,2], HUGO DUARTE[1,2], JOÃO A. M. SANTOS[1,3,4], AND PEDRO H. ABREU[5]

[1]Medical Physics, Radiobiology and Radiation Protection Group, IPO Porto Research Centre (CI-IPOP), 4200-072 Porto, Portugal
[2]Nuclear Medicine Department, Portuguese Institute of Oncology of Porto (IPO-Porto), 4200-072 Porto, Portugal
[3]Medical Physics Department, Portuguese Institute of Oncology of Porto (IPO-Porto), 4200-072 Porto, Portugal
[4]Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto, 4099-002 Porto, Portugal
[5]CISUC, University of Coimbra, 3030-789 Coimbra, Portugal

Corresponding author: Inês Domingues (ines.domingues@isec.pt)

This work was supported by the Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF) under Project NORTE-01-0145-FEDER-000027.

ABSTRACT Esophageal cancer is a disease with a high prevalence that can be evaluated by a variety of imaging modalities, including endoscopy, computed tomography, and positron emission tomography. Computer-aided techniques could provide a valuable help in the analysis of these images, decreasing the medical workflow time and human errors. The goal of this paper is to review the existing literature on the application of computer vision techniques in the domain of esophageal cancer. After an initial phase where a set of keywords was chosen, the selected terms were used to retrieve papers from four well-known databases: Web of Science, Scopus, PubMed, and Springer. The results were scanned by merging identical entries, and eliminating the out of scope works, resulting in 47 selected papers. These were organized according to the image modality. Major results were then summarized and compared, and main shortcomings were identified. It could be concluded that, even though the scientific community has already paid attention to the esophageal cancer problem, there are still several open issues. Two major findings of this review are the nonexistence of works on MRI data and the under-exploration of recent techniques using deep learning strategies, showing the need for further investigation.

INDEX TERMS Computed tomography, computer vision, computer aided analysis, endoscopy, esophageal cancer, positron emission tomography.

I. INTRODUCTION

Esophageal Cancer (EC) is, globally, one of the most frequently reported malignancies [1], [2]. This disease ranks seventh in terms of incidence (572,000 new cases) and sixth in general mortality (509,000 deaths), the latter meaning that esophageal cancer will account for an estimated one in every 20 deaths from cancer in 2018 [3].

As for other types of cancer, a multimodality approach is presently used to treat EC (including different combinations of endoscopic therapies, chemotherapy, radiotherapy, and surgery) [4]. The Tumour-Node-Metastasis (TNM) staging system for EC provided by the American Joint Committee on Cancer (AJCC) [5] and Union for International Cancer Control (UICC) [6] is generally used, considering the characteristics of the primary tumor, regional nodal metastases, and distant metastases (Table 1) [7].

Radiotherapy (RT) is an important component in the multimodal treatment of EC. For its planning, it is crucial not only to accurately delineate the tumor in order to ensure adequate coverage of the target, but also to identify critical organs such as the lungs and heart in order to limit the dose applied to these organs.

As with other diseases of the upper GastroIntestinal (GI) tract, EC can be evaluated by a variety of imaging modalities [9], including:

- Endoscopy
  - Application: endoscopy with biopsy is used to diagnose EC [10], [11].
  - Main advantages: permits direct inspection and biopsy of the esophageal mucosa for histologic diagnosis.

The associate editor coordinating the review of this manuscript and approving it for publication was Li Zhang.

TABLE 1. TNM staging for esophageal cancer [5], [8].

Type	Description
Tis	High-grade dysplasia
T1a	Tumour invades lamina propria or muscularis mucosae
T1b	Tumour invades submucosa
T2	Tumour invades muscularis propria
T3	Tumour invades adventitia
T4a	Resectable tumour invading pleura, pericardium, diaphragm or adjacent peritoneum
T4b	Unresectable tumour invading other adjacent structures, such as aorta, trachea, vertebral body
N0	No regional lymph node metastases
N1	1-2 regional lymph node metastases
N2	3-6 regional lymph node metastases
N3	$\geq$7 regional lymph node metastases
M0	No distant metastases
M1	Distant metastases

  - Main disadvantages: it is an invasive technique [11] and operator dependent [9], [11], [12].
- Computed Tomography (CT)
  - Application: useful in distinguishing between patients with early cancer (T1a/T1b and T2) who need further evaluation with Endoscopy and those with T3 and T4 disease [4]; used for tumor delineation during radiotherapy planning [11].
  - Main advantages: reliable in determining resectability [9], [11].
  - Main disadvantages: CT is unable to distinguish the esophageal wall layers to determine the depth of tumor infiltration [11].
- Positron Emission Tomography (PET)
  - Application: Fluorodesoxiglicose (FDG)-PET is a well-established imaging technique for staging EC [13], being the most important role of this modality the detection of distant metastases [11].
  - Main advantages: assessing of metabolic function, high sensibility, high reproducibility (when complying with the acquisition standards), and existence of quantitative measurements such as SUV [7].
  - Main disadvantages: low spatial resolution when compared with other imaging techniques, low specificity of 18F-FDG-PET, and still a lack of commercial availability of other pharmaceuticals beyond F-18:FDG [14].

A summary of the different imaging techniques in EC used in clinical practice for diagnosis, TNM-staging, tumor delineation for RT, and treatment response assessment, is given in Table 2.

The role of MRI in the management of EC is still unclear and poorly studied. Although recent technical advances suggest that MRI may become a powerful technique not only for the initial evaluation of esophageal cancer but also for the evaluation of the response to neoadjuvant treatment prior to surgery [15], no work of computational vision for this modality was found. The clinical trial with the identifier NCT03347630[1] aims to evaluate the accuracy of the MRI to visualize esophageal tumors, identify tumor burden and potential contact with adjacent structures, as well as the associated lymph nodes, and also if the MRI helps in the better evaluation of the response to the treatment. We anticipate that if this clinical trial (and possibly others) prove the importance of MRI, computer vision techniques will emerge for this modality.

TABLE 2. Imaging techniques in esophageal cancer used for diagnosis, TNM-staging, tumor delineation for RT, and treatment response assessment.

	Endoscopy	CT	PET
Diagnosis	✓		
T-Staging	✓	✓	
N-Staging	✓	✓	✓
M-Staging		✓	✓
Tumour delineation for RT		✓	✓
Evaluation of response	✓	✓	✓

To achieve the main goal of this paper, which is to draw a global picture of the computer vision state of the art in Esophageal Cancer, a systematic literature search was first performed. Four well established search engines were used, and the results were merged and pruned, resulting in a total of 47 papers. The search engine results were then organized accordingly to the image modality studied. A first observation, as stated above, is that no works using MRI have yet been performed. Endoscopy and PET have been gathering most of the attention of the community. In relation to the computer vision techniques used, there is still a high prevalence of traditional techniques based on hand-crafted features. Deep learning is still in the early stages of adoption and further work and evaluation needs to be performed.

The present work reviews computer vision works with application in EC and is organized by image modality. In this way, we start in Section II by explaining the methodology used to select the reviewed works. Next, the main body of the paper is given in Section III and is organized by image modality used, namely, Endoscopy (Section III-A), CT (Section III-B), and PET (Section III-C). In each section, an introductory paragraph is given, followed by a summary table with all of the works found for that modality and then a paragraph per work, briefly describing their purpose. Section IV presents a discussion organized into a section on Deep Learning (DL) methods and a section on the publicly available databases. The paper finishes with some conclusions in Section V.

II. METHODOLOGY

Studies were identified through a systematic electronic search of several databases on May 9th, 2018, with keywords chosen by the multidisciplinary authors (specialized in areas from applied mathematics, medical doctors, physics, and

[1] https://clinicaltrials.gov/ct2/show/NCT03347630

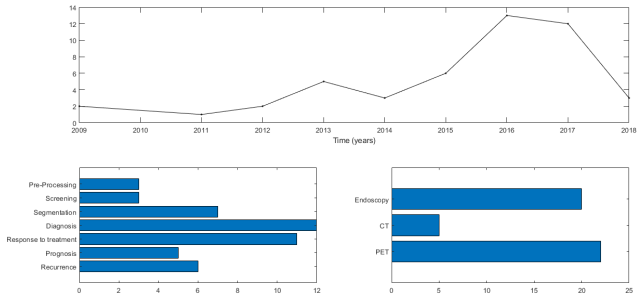


FIGURE 1. Number of publications per year (top), per topic (bottom-left) and per data type (bottom-right).

engineering). The following bullet point list summarizes the performed search:

- **Web of Science** with the topic ''oesophagus'' or ''esophagus'', refined by the categories Computer science artificial intelligence, Computer science information systems, Computer science interdisciplinary applications, Computer science theory methods, Engineering biomedical, Engineering electrical electronic, Nuclear science technology
- **Scopus** with the string: ABS (esophagus OR oesophagus) AND (LIMIT-TO (SUBJAREA, ''ENGI'') OR LIMIT-TO (SUBJAREA, ''COMP'') OR LIMIT-TO (SUBJAREA, ''DECI'') OR LIMIT-TO (SUBJAREA, ''Undefined'')) AND (LIMIT-TO (LANGUAGE, ''English''))
- **PubMed** with the string (''oesophagus'' [All Fields] OR ''esophagus'' [All Fields]) AND (''radiomic'' [All Fields] OR ''radiomics'' [All Fields])
- **Springer** with the keyword string ''esophageal cancer'' and selecting the Discipline ''Computer Science''

The returned results were merged and screened first with basis on the titles and then the abstracts. The selection criteria was to retrieve works that used computer vision techniques to analyze esophagus image data. When pertinent, citations within the articles were added to the pool of papers. Works that cite the ones already in the pool were also verified.

Only works written in English were considered. This resulted in 47 papers. Distributions by year, subject, and type of exam are given in Fig. 1. It is clear that the interest in the field has been growing (we note that the search was performed in early 2018 and thus the apparent decay in 2018, which we believe will not be true once the year ends). Moreover, the topic that raised most interest was diagnosis, followed by treatment response evaluation, while other applications are still under-studied. Concerning the type of exam, PET and Endoscopy are used in most computer vision works, while CT has been the subject of less studies.

III. LITERATURE REVIEW

Rossum, in the fourth chapter of his Ph.D. thesis in 2016 [11], made a literature review of radiomic applications in patients with EC. The author concludes that since the first publication on analysis of image texture characteristics in the year 2011 [16],[2] the quantity of work in this scenario has been growing, suggesting incremental value in the automatic assessment of staging, prediction of response to treatment and survival forecast.

Here, we update and complement Rossum's review [11], the only existing survey on this subject that we are aware of. The works are organized, in the next subsections, by imaging technique. In each subsection, a brief description of the reviewed papers is provided along with a summary Table.

A. ENDOSCOPY

Munzer *et al.* [17] survey the state of the art on content-based processing and analysis of endoscopic images and videos, pointing out that ''digital endoscopy has established as key technology for medical screenings and minimally invasive surgery''. When zooming on EC works, as summarized in Table 3, the same pattern emerges. The interest in the field is also shown by the emergence of challenges in conferences such as ''EndoVis''[3] in MICCAI 2015, ''AIDA-E''[4] in ISBI 2016, and ''EAD''[5] in ISBI 2019. Techniques employed range from Mosaicing [18], supervised classification with hand-crafted features [12], [19]–[30] or DL [31]–[34], tracking [35]. These are described in the next paragraphs.

Saraf *et al.* [19], propose an Artificial Neural Network for the classification of esophagitis into four classes. The network receives as input characteristics based on the colour and texture of images captured during endoscopy. The Hue, Saturation and Intensity models are adapted. Statistics drawn from Hue and Saturation are used as color characteristics and the coefficients of the Discrete Cosine Transform are used as texture characteristics. The classification efficiency in a balanced data set of 30 images for each category ranges from 80% to 96%, with the sensitivity ranging from 0.8 to 0.97 and the specificity between 0.94 and 0.98.

Carroll and Seitz [18] try to capture an entire scene in a single image, through a process called Mosaicing. Having as input a video captured by a camera moving through a tube or other surface, the method returns an image representing the texture of the unfolded surface. One of the motivations of this work is to create a screening tool and in the paper an example made from 220 frames of esophageal endoscopy is presented.

Sommen *et al.* [20] propose an algorithm to identify irregularities in the esophagus. In order for the system to both detect if an endoscopic image contains early cancer, and also locate it, tile based processing is employed. Performance is evaluated in the RGB, HSI and YCbCr color spaces using as features color Histogram and Gabor filters and as classifier Support Vector Machines (SVM). In the experiments with images from 66 patients (the total number of images used

[2]Note that in the present review we consider earlier works, possible unknown to Rossum

[3]Available at endovis.grand-challenge.org

[4]Available at isbi-aida.grand-challenge.org

[5]Available at ead2019.grand-challenge.org

TABLE 3. Works on endoscopic images. [p] indicates private databases, while [o] indicates open ones.

Paper	Goal	Methods	Database
Saraf et al. (2009) [19]	Diagnosis of Esophagitis	Feed forward neural network with Hue Saturation and an adapted Intensity colour model	[p] 120 images (30 images of each class, Normal, Grade I, Grade II, Grade III)
Carroll and Seitz (2009) [18]	To capture the appearance of an entire screen in a single image	Mosaicing: The 6 DOF camera pose is first estimated. Next, an inverse projection of each frame onto the surface is performed. The 2D mosaic is formed by flattening the surface onto a plane.	[p] 220 frames of an esophageal endoscopy
Sommen et al. (2013) [20]	Automatically identify irregularities in the esophagus	The identification is based on the following steps: (1) preprocessing, (2) feature extraction with dimensionality reduction with Principal Component Analysis - PCA, (3) classification with SVM	[p] Endoscopic images, with a resolution of 1600×1200 pixels, of 66 patients
Sommen et al. (2014) [21]	Automated detection and annotation of early cancer in BE	SVM with local color and texture features based on the original and on the Gabor-filtered image	[p] 32 HD (1600x1200 pixels) endoscopic images of 7 patients with histologically proven EAC
Sommen et al. (2016) [22]	Assess the feasibility of a computer system to detect early NPL in BE	Gabor filtering, followed by image features and classification with SVM	[p] 100 images from 44 patients with BE
Sommen et al. (2016) [23]	Detection of early neoplastic lesions in BE	SST	[p] 100 endoscopic images (resolution of 1200×1600 pixels) of 39 patients
Zhang et al. (2016) [36]	Prediction of the types of esophageal lesions	Graph representation (expressing pixels as nodes and similarity between the grey-level or local features as edges) followed by quantification of the properties of the texture features by similarity measurements (high-order graph matching kernel)	[p] EUS of 66 patients with early EC + 91 control patients
Matsunaga et al. (2016) [24]	Early stage cancer detection	DWT	[p] 5 FICE mode images from 3 patients
Janse et al. (2016) [25]	Early stage cancer detection	RF	[p] 39-patient dataset, containing 100 static HD endoscopic images
Ohura et al. (2016) [12]	Early EC diagnosis	DYWT and the fractal dimension	[p] Endoscopic images of 23 early EC patients
Shin et al. (2016) [26]	Discriminate BE with NPL from benign esophageal tissue including normal squamous, BE without dysplasia, and BE with low-grade dysplasia	Sequential binary classification using LDA with quantitative image features	[p] HRME images from 230 sites in 58 patients acquired during a standard endoscopic procedure
Xue et al. (2016) [31]	Classification of microvascular morphological types to aid EC detection	CNN-SVM model where feature extraction is performed with CNN and classification is made using SVMs	[p] NBI-ME dataset with 261 images from 67 patients
Vemuri (2016) [35]	Navigation system for endoluminal surgery	tracking and relocalization	[p] Synthetic and real (porcine) endoscopy
Ghatwary et al. (2017) [27]	Grade classification of BE	Multi-stage classification (3 cascade SVMs)	[p] 32 patients with 262 CLE images
Swager et al. (2017) [28]	Detection of early BE NPL	4 generic features and 3 new ones in combination with 8 supervised machine learning methods	[p] 60 ex vivo VLE images
Klomp et al. (2017) [29]	Predict the presence of dysplastic tissue in VLE BE images	SVM using modified Haralick features and optimal image cropping	[p] 30 dysplastic and 30 non-dysplastic VLE images
Souza et al. (2017) [30]	Automatic identification of BE	OPF with SURF and SIFT	[o] 100 endoscopic "EndoVis" pictures of the lower esophagus captured from 39 individuals, 17 of them being diagnosed with early stage Barrett's, and 22 displaying signs of EAC
Hong et al. (2017) [32]	Distinguish IM, GMP, and NPL	Data augmentation and CNN	[o] 262 endomicroscopy "AIDA-E" imaging data of BE
Riel et al. (2018) [33]	Early detection of EAC in the BE	Transfer information from a non-medical domain, in the form of pre-trained CNNs trained on ImageNet, and apply them as features for a conventional classifier	[o] Early Barrett's Cancer Detection Challenge dataset (subset of "EndoVis")
Garcia-Peraza-Herrera et al. (2018) [34]	Interpretability of classification results coming from a CNN	Deep supervision and eCAM	[o] 17 monocular videos (7,046 frames from 17 patients - one video per patient)

is not mentioned), the system reaches an accuracy of 95.9% with an area under the ROC curve of $AUC = 0.990$.

The same team continues this line of work and develops an algorithm based on the original and filtered image [21], by using custom filters. After classification with an SVM, post-processing techniques are applied to annotate the region of the image containing the cancer. For 7 patients, 32 annotations made by the algorithm are compared with the corresponding delineations done by a gastroenterologist. The system was able to detect 36 of the 38 lesions, with a recall of 0.95 and an accuracy of 0.75. The same system was then evaluated [22] in 100 images from 44 patients identifying early neoplastic lesions with a sensitivity and specificity of 0.83 at the image level and sensitivity and specificity of 0.86 and 0.87 respectively, at the patient level.

Continuing with the tasks of detection and segmentation, [23] argue in [23] that the detection of cancerous tissue in the gastrointestinal tract should not be approached as a binary problem, but that several specialists should segment the lesions. A ''sweet spot'' metric is proposed for the training phase (Sweet Spot Training - SST) and the Jaccard Golden Standard (JIGS) a metric able to handle multiple annotations is also proposed. Results on 100 endoscopic images from 39 patients show that, in this way, the performance of a detection algorithm can be increased by up to 10% of F1.

Graph theory is used to analyze Endoscopic Ultrasonography (EUS) by Zhang et al. [36]. Points of interest calculated by Scale-Invariant Feature Transform (SIFT) are used as nodes and the similarity between the grey level or other local characteristics of the image is used to derive the weights of the edges. A refined SVM model, with a new kernel based on graphics matching for EUS images, is designed to predict three classes, normal, leiomyoma of the esophagus and early EC. The overall accuracy was 93%. For the diagnosis of early EC, the accuracy, sensitivity, specificity, positive predictive value and negative predictive value were 89%, 94%, 95%, 89 and 97%, respectively. The authors note that the false alarm rate is small; however, failure to detect (patient with early EC signs confused with healthy) poses a risk.

Matsunaga *et al.* [24] propose a method to detect early EC from Flexible spectral Imaging Color Enhancement (FICE) images. The image is first converted to the CIEL*a*b* color space, and then only the a* component is used. The Daubechies Wavelet Transform (DWT) is calculated in non-overlapping blocks with a size of 64×64 pixels. Detection is then performed by applying a threshold to the histogram. The results are illustrated in five images from three patients with EC, but no quantification of the results is performed.

A computer-aided detection system for the identification of early cancer is improved in [25] by including a Random Forest (RF), and thereby introducing a measure of confidence for the detected regions. The method consists of the following steps: Region of Interest (ROI) detection, feature extraction, classification, and annotation. The system is validated on 100 static High Definition (HD) endoscopic images of 39 patients, manually annotated by 5 gastroenterologists. A 75% accuracy and 90% recall were achieved, improving the original system results by 11 and 6 percentage points, respectively.

Ohura et al. [12] propose a method to diagnose early EC from endoscopic images using the DYadic Wavelet Transform (DYWT) and the fractal dimension. Endoscopic images of 23 patients with EC were used. The authors claim that the method is capable of providing diagnostic information that may assist medical specialists, but no quantitative data is provided.

Barrett's esophagus (BE) is a precancerous condition that may progress to become Esophageal AdenoCarcinoma (EAC). Shin *et al.* [26] describe an approach for the analysis of microendoscopic images with the aim of identifying neoplastic lesions in patients with BE. Features extracted from High-Resolution MicroEndoscope (HRME) images are used to develop a classification algorithm between neoplastic (HGD or cancer) or non-neoplastic tissue (normal squamous mucosa, gastric cardia, Barrett's metaplasia or LGD). Linear discriminant analysis of two classes is used, where the image characteristics are added one at a time until the classification performance stops improving. Images were acquired from 230 sites in 58 patients, achieving a sensitivity and specificity of 88% and 85% for the 5-class problem.

The work in [31] classifies microvascular morphological types to aid EC detection. They develop a model in Caffe where patches are greedily generated, feature extraction is performed with a Convolution Neural Network (CNN), and classification is made with an SVM. They use a NarrowBand Imaging with Magnified Endoscopy (NBI-ME) dataset composed of 261 images from 67 patients, achieving a 92.74% recognition rate.

Esophageal surveillance usually involves obtaining biopsies in different regions along the oesophagus. Identification of these biopsy sites between operations presents a challenge. The doctoral thesis of Vemuri [35] describes a navigational system for endoluminal surgery, not resorting to the use of a preoperative model. An intervention is recorded and later used to provide guidance to the specialist, assisting in the re-positioning of the endoscope at previously studied sites. This is achieved by using video synchronization between operations. The quantitative evaluation of the system is performed with synthetic and real data (porcine). The results show an improvement in the relocation rate from 47.5% to 94.0%.

In [27], an automatic BE classification system is introduced. Several features are extracted from an image of Confocal Laser Endomicroscopy (CLE), filtered at different levels. SVMs are then applied in a multi-stage fashion in order to obtain a classification into three classes. The approach is validated on a dataset of 32 patients with 262 images of different histological grades. The three target classes are: Gastric Metaplasia (GMP), Intestinal Metaplasia (IM) and NeoPlasia (NPL). A general accuracy of 90.5% was achieved, with binary results reaching 98.8% to discriminate between

NPL and other images and 96.7% to separate IM from GMP. For a survey on automated detection of BE Using Endoscopic Images, we refer to the review by the same authors in [37].

Swager *et al.* [28] also focus on the detection of early Barrett's NPL using Volumetric Laser Endomicroscopy (VLE). The authors use 4 generic features, (1) GLCMs, (2) Local Binary Patterns, (3) Histogram of Oriented Gradients, and (4) Wavelet transform, and also propose 3 new features that they name (1) Layering, (2) Signal Intensity Distribution, (3) Layering and Signal Decay Statistics. Each set of features is tested in combination with several supervised machine learning methods, namely SVM, Discriminant Analysis, AdaBoost, RF, K-Nearest Neighbor classifier (k-NN), Naive Bayes, Linear Regression, and Logistic Regression. In a database of 60 ex vivo VLE images, an AUC of 0.91 was achieved when using Layering and Signal Decay Statistics with the AdaBoost classifier.

VLE is also used in [29], where three contributions are made, (1) benchmarking of machine learning and feature extraction techniques, (2) proposal of three new features, (3) evaluation through automated adjustment of parameters by applying feature selection methods and grid searching. The results are evaluated in clinically validated data with 30 dysplastic and 30 non-dysplastic images. The best classification accuracy is obtained by applying an SVM as classifier and using modified Haralick characteristics and optimized image cropping, obtaining an AUC of 0.95.

Souza et al. [30] introduce the Optimum-Path Forest (OPF) classifier for the BE identification task. Descriptors based on key-points, such as Speeded Up Robust Features (SURF) and SIFT, are extracted in order to create a bag-of-visual-words. The OPF was compared with an SVM in a set of 100 endoscopic images of the lower esophagus captured from 39 patients, 17 of whom were diagnosed in the initial stage of Barrett and 22 showed signs of EAC, provided by the EndoVis MICCAI 2015 Challenge. The best results were obtained by OPF, with values of sensitivity, specificity, and accuracy of 0.732, 0.782 and 0.738 when using SURF and 0.735, 0.806 and 0.732 when using SIFT.

Hong *et al.* [32] built a CNN to distinguish between IM, GMP and NPL. The architecture is composed of four convolutional layers, two max-pooling layers and two Fully Connected layers (FC). Image distortion was employed for data augmentation. 262 BE endomicroscopy images were obtained from the ISBI 2016 "AIDA-E" challenge, achieving a classification accuracy of 80.77%.

Motivated by the existence of few annotated data, Riel *et al.* [33] present an approach for the early detection of EC using Transfer Learning of CNNs. Intermediate layers of the network are used as features in traditional classifiers. Sliding windows are applied to obtain the approximate location of possible cancerous regions. The experiments are performed with the Early Barrett's Cancer Detection Challenge dataset, an EndoVis sub-challenge. The approach achieves an AUC of 0.92, allowing detection and annotation at two frames per second (fps).

The work of Garcia-Peraza-Herrera *et al.* [34] deals with the interpretability of classification results coming from a CNN. They present a new dataset composed of 7,046 frames from 17 patients, and a novel DL method with deep supervision and an embedded Class Activation Map (eCAM). The approach aims at directing the attention to those areas of the esophageal tissue that lead the network to conclude that the images belong to a particular class, in particular for the case of IntraPapillary Capillary Loops (IPCL). In comparison with a method that does not do feature deep supervision, but provides attention by grafting Class Activation Maps, the F1-score is improved from 87.3% to 92.7%. The authors provide more detailed attention maps.

B. COMPUTED TOMOGRAPHY

CT volumes (Table 4) have been analyzed mostly with traditional hand-crafted features [38]–[40]. Two exceptions are [41], [42] where DL techniques are used. One particular feature of the work [42] is the use of convolutions in 3D, fully using the volumetric information, while most methods use 2D or 2.5D techniques, even when the input is 3D [43]. Another interesting aspect of this work is the use of a publicly accessible database, made available in 2015 at the Workshop and Challenge of MICCAI - Multi-Atlas Labelling Beyond the Cranial Vault.

The authors of [38] analyzed the association between tumor heterogeneity, morphological tumor response, and Overall Survival (OS) in primary EC treated with chemotherapy and RT. The dataset consisted of 36 patients with stage T2 or worse, who underwent contrast-enhanced CT before and after ChemoRadioTherapy (CRT). Entropy, uniformity, mean intensity, kurtosis, histogram standard deviation and skewness were extracted after the application of filters 1.0-2.5. It is concluded that post-treatment texture features are associated with survival time, and the combination of pre-treatment texture features with maximum wall thickness showed better performance in survival models than when only using the morphological tumor response.

A different goal is the focus of the work in [39], where the authors evaluate the relationship between radiation dose and evolution in patients who received RT for EC to identify patients who develop Radiation Pneumonitis (RP). Pre and Post RT CT scans were acquired, and a CT treatment plan with an associated dose map was designed for 106 patients. Non-overlapping ROIs of size 32×32 pixels were randomly drawn from the lungs region in the pre-treatment scans. The ROIs were then aligned to both the scan after the treatment and the planning scan using deformable image registration. Texture and intensity features, including first order, fractal, Laws' filter and Gray-Level Co-Occurrence matrix (GLCM) were extracted. The absolute value of the difference between pre and post treatment images, ΔVF, was then calculated. Logistic Regression (LR) was used to study if the combination of features improves the identification of patients on stage $\geq$ 2. ΔVF was significant for all 20 features, with increasing radiation dose; while for patients with RP, only

TABLE 4. Works on CT images. [p] indicates private databases, while [o] indicates open ones.

Paper	Goal	Methods	Database
Yip et al. (2014) [38]	Determine the association between tumour heterogeneity, morphologic tumour response, and OS	Entropy, uniformity, mean grey-level intensity, kurtosis, standard deviation of the histogram, and skewness for fine to coarse textures	[p] CT studies of 36 patients
Cunliffe et al. (2015) [39]	Identify patients who develop RP	Deformable image registration; 20 texture and intensity features; LR	[p] CTs before, planning and after after RT for 106 patients
Yip et al. (2015) [40]	To assess the changes in tumour heterogeneity following neoadjuvant chemotherapy	Texture parameters (mean grey-level intensity, entropy, uniformity, kurtosis, skewness, standard deviation of the histogram)	[p] Staging and post-chemotherapy CT scans of 31 patients
Trullo et al. (2017)	Esophagus segmentation	Fully Convolutional Networks	[p] 30 scans
Fechter et al. (2017) [42]	Esophagus segmentation	3D CNN and random walk	[p,o] 20 clinical and 30 publicly available CTs

12 features were statistically significant. Discrimination of patients with and without RP when using a single feature achieves AUCs between 0.49 and 0.78, while when combining features, the same metric increases from 0.59 to 0.84.

Yip et al. [40] assess the impact of neoadjuvant chemotherapy on tumor heterogeneity. Image texture parameters (mean grey level, entropy, uniformity, kurtosis, skewness, histogram standard deviation) are extracted from both staging and post-chemotherapy exams for thirty-one patients. Proportional changes in each parameter were then calculated. It was been found that the texture of the tumor becomes more homogeneous after treatment with a significant decrease in entropy and increase in uniformity. The standard deviation of the histogram showed a borderline association with the tumor response. A proportional change in skewness smaller than 0.39 was associated with improved survival.

The paper [41] concentrates on the segmentation of the oesophagus. This is a challenging problem given that the walls of the oesophagus have a low CT contrast. The authors propose a two-step approach: (1) gross segmentation where multiple organs are identified; (2) finer segmentation, focused on a smaller region identified in step (1), but large enough to contain context information. The Sharpmask architecture is used in both steps. Performance is assessed in 30 CTs of patients which have lung cancer or Hodgkin's lymphoma, obtaining a Dice score of 0.72 ± 0.07.

3D CNNs are used in [42] to locate the oesophagus in CT images. The results are refined through a random walker approach in order to achieve the segmentation of the oesophagus. More specifically, the pipeline is as follows: (1) a 3D CNN is used to generate a smooth probability map, (2) an Active Contour Model - ACM - is fitted to the probability map in order to obtain a rough estimate of the location of the oesophagus, (3) a random walker is applied on the combination of a probability model based on Hounsfield CT units with the outputs of the previous steps. The training and assessment were done on 50 CTs from two different data sets, one private with 20 scans, and one public with 30 scans from the ''Multi-Atlas Labelling Beyond the Cranial Vault - Workshop and Challenge'' (synapse dataset). Evaluation using Dice coefficient and symmetric square distance achieved average values of of 0.76 and 1.36 mm, respectively.

C. POSITRON EMISSION TOMOGRAPHY

PET data has been automatically processed since 2011, as shown in Table 5. The focus has been on the selection of the best hand-crafted features [16], [44]–[63], with the exception of the work in [64] where response to treatment is predicted using a three-slice CNN (3S-CNN). Although having the advantage of not using hand-crafted features, the proposed 3S-CNN, is not volumetric, and thus not having the potential advantages described in the beginning of Section III-B for the work of Fechter *et al.* [42]. More details on each work are given next.

The predictive value of FDG uptake heterogeneity, quantified using texture analysis, in 41 patients with locally advanced EC treated by concomitant CRT is studied in [16]. SUV_{max}, SUV_{peak}, SUV_{mean}, and 38 texture features were extracted from tumor regions of pre-treatment PET scans. Patients were divided into NR, Partial-Respondent (PR) or Complete-Respondent (CR) according to RECIST criteria. The Kruskal-Wallis test (p-value < 0.05) was used to verify the ability of each parameter to classify patients in relation to response to therapy. It was observed that the relationships between pairs of voxels are able to characterize the non-uniformities of metabolism of the local tumor and thus to significantly differentiate all three groups of patients. Regional measures, such as the size of non-uniform metabolic regions and non-uniformity of intensity within these regions, were also significant factors. ROC analysis showed that textural analysis of the tumor can provide identification of NR, PR and CR patients with greater sensitivity ($76\% - 92\%$) than any of the measures of SUV.

Tixier *et al.* [44] evaluate the reproducibility of texture features. In their method, 3D segmentation is made by a Fuzzy Locally Adaptive Bayesian (FLAB) algorithm. Extracted parameters are: 8 features based on intensity histogram, 11 based on Grey-Level Size Zone Matrix (GLSZM), and 6 based on the GLCM. On double baseline PET scans for 16 patients, they concluded that local homogeneity, entropy, dissimilarity, Intensity Variability (IV) and Size-Zone Variability (SZV) are parameters with high reproducibility.

New methods of risk stratification can lead to the optimization of management strategies leading to better results. In this way, in [45] the potential of CT texture analysis to

TABLE 5. Works on PET images. [p] indicates private databases, while [o] indicates open ones.

Paper	Goal	Methods	Database
Tixier et al. (2011) [16]	To study the predictive value of FDG uptake heterogeneity quantified using texture analysis	38 texture features (such as entropy, size, and magnitude of local and global heterogeneous and homogeneous tumour regions)	[p] Pre-treatment PET tumour images for 41 patients with locally advanced EC treated by concomitant CRT
Tixier et al. (2012) [44]	To evaluate the reproducibility of texture features	3D FLAB segmentation followed by feature extraction (8 features based on intensity histogram, 11 based on GLSZM, and 6 based on the GLCM)	[p] Double baseline PET scans for 16 patients
Ganeshan et al. (2012) [45]	Risk stratification	Texture analysis performed at different anatomical scales	[p] Baseline PET/CT for 21 patients
Hatt et al. (2013) [46]	Study of the impact of pre-processing steps on the quantification of intra-tumour uptake heterogeneity	Tumour volumes delimitation was performed by FT and AT, and the FLAB algorithm. Heterogeneity was quantified using local and regional textural features.	[p] PET scans of 50 patients with EC
Dong et al. (2013) [47]	To explore the relationship of PET texture parameters with SUV_{max} and tumour TNM staging	Entropy and energy of the volumes	[p] Preoperative 18F-FDG PET/CT of 40 patients
Tan et al. (2013) [48]	To predict pathologic tumour response to CRT	Image features defined by histogram distances extracted from PETs acquired before and after CRT	[p] 20 patients with EC treated with CRT plus surgery
Tan et al. (2013) [49]	To predict pathologic tumour response to CRT	A comprehensive set of features were extracted to characterise the SUV intensity distribution, spatial patterns (texture), tumour geometry, and associated changes	[p] 20 patients with EC treated with CRT plus surgery (FDG-PET/CT for pre-CRT and post-CRT)
Zhang et al. (2014) [50]	Prediction of pathologic tumour response	Recursive feature selection; SVM and LR	[p] PET/CT of 20 patients before and after CRT
Hatt et al. (2015) [51]	To investigate the complementary nature of MATV and texture heterogeneity	Four textural features	[p] 555 pre-treatment 18F-FDG PET images (158 breast, 45 cervix, 112 esophageal, 139 head&neck and 101 lung cancer tumours)
Ypsilantis et al. (2015) [64]	Predict response to neoadjuvant chemotherapy	3S-CNN	[p] PET scans for 107 patients
Lian et al. (2015) [52]	To find discriminant features from both PET images and clinical characteristics, so as to predict the outcome of a treatment	EFS and mEK-NN	[p] baseline PET/CT for 36 patients
Lian et al. (2016) [53]	To find discriminant features from both PET images and clinical characteristics, so as to predict the outcome of a treatment	iEFS and EK-NN	[p] baseline PET/CT for 36 patients
Yip et al. (2016) [54]	Predict response to treatment	rigid and deformable contour propagation	[p] pre and post treatment PET/CT scans for 45 patients
Yip et al. (2016) [55]	To compare texture features with SUV measures for pathologic response and OS	SUV_{max} and SUV_{mean}, two GLCM, two GLRL, and two GLSZM textures	[p] pre and post treatment PET/CT scans for 45 patients
Rossum et al. (2016) [56]	Prediction of a pathCR to CRT before surgery	Step-wise improvement of a clinical prediction model	[p] baseline and postchemoradiation 18F-FDG PET for 217 patients
Desbordes et al. (2017) [57]	Predictive and prognostic studies	RF with texture features	[p] pre-treatment whole-body FDG-PET for 65 patients
Nakajo et al. (2017) [58]	Differentiating between FDG-avid Benign Adrenal Tumour (BAT) and MAT	SUV_{max}, MTV, TLG, and 4 textural parameters (entropy, homogeneity, IV, and SZV)	[p] PET/CT for 13 BATs and 22 MATs of each 3 are EC
Foley et al. (2017) [59]	OS prediction	Semi-automatic segmentation and regression model that includes texture features	[p] PET/CT for 403 biopsy-proven patients
Nakajo et al. (2017) [60]	Examine whether the heterogeneity in primary tumour (18)F-FDG distribution can predict tumour response and prognosis	SUV_{max}, SUV_{mean}, MTV, TLG and 6 heterogeneity parameters	[p] (18)F-FDG-PET/CT before CRT for 52 patients
Beukinga et al. (2016) [61]	Predict complete response to nCRT	LR prediction model with clinical, geometrical, and pre-treatment textural features	[p] 18F-FDG PET/CT for 97 patients
Doumou et al. (2015) [62]	To test the effects of image smoothing, segmentation, and quantization on the precision of heterogeneity measurements	Concordance evaluation of homogeneity parameters for the three processing variables	[p] 64 PET/CT
Anthony et al. (2017) [63]	Determine whether the addition of SUV from PET scans to CT lung texture features could improve a radiomics-based model of RP diagnosis in patients undergoing RT	20 texture features and LR classifiers	[p] pre-therapy PET/CT scans, pre-/post-therapy diagnostic CT scans and RP status of 96 patients

contribute to risk stratification is investigated. The texture features are extracted from un-enhanced CT and compared with SUV (from PET/CT) and patient survival. The experimental database consisted of baseline PET/CT data from 21 patients. Tumor heterogeneity was found to correlate with the adverse biological characteristics of high tumor metabolism, advanced stage, and patient survival.

In [46], it was noted that although the heterogeneity of intra-tumor uptake quantified with textural characteristics has been thoroughly investigated, an evaluation of the impact of the pre-processing steps in the resulting quantification had not yet been performed. They thus perform that study with respect to the functional volume segmentation and Partial Volume Effect (PVE) Correction (PVC). Fixed Threshold (FT), Adaptive Threshold (AT) and FLAB are used for tumor identification. Local and regional textural features are then extracted. In a population of 50 patients, it was observed that, although there are differences in the absolute values of the parameters, these differences do not always translate into a significant impact. Features such as local entropy, local homogeneity, and regional Zone Percentage (ZP) should be preferred as they achieve not only a high differentiation power in terms of predicting patient response, but are also robust in relation to the segmentation method and PVE.

Dong et al. [47], propose a new texture PET image feature, named uptake heterogeneity. The heterogeneity of intra-tumoural 18F-FDG uptake is assessed by the entropy and energy of the volumes and their correlation with SUV_{max}, histological grade, tumor location, and the TNM stage is analyzed. They observe that: (1) tumors with higher SUV_{max} are more heterogeneous; (2) significant correlations exist between T stage and both entropy and energy; (3) correlations were also found between N stage and both entropy and energy. Thresholding the entropy at 4.699, detection of tumors above stage IIb achieve an AUC of 0.789 ($p < 0.001$). They conclude that the proposed uptake heterogeneity may complement the SUV and tumor stage in the staging and prognosis of squamous cell carcinoma of the oesophagus.

Tan et al. [48] propose a family of characteristics based on distances from the histogram to predict tumor pathological response to neoadjuvant CRT (nCRT). These characteristics describe the longitudinal alteration of the distribution of FDG in the tumor. The tumor is manually delineated in the pre-treatment scans. Subsequently, the pre and post treatment exams are rigidly registered and two histograms are calculated in the delimited region, one from the pre and the other from the post scans. A total of 19 histograms are examined and compared with traditional measures such as Haralick's texture features. The results obtained with data from 20 EC patients treated with CRT and surgery showed that 7 bin-bin and 7 crossbin distances exceeded traditional measures, such as SUV_{max} and Total Lesion Glycolysis (TLG). In addition, the 7 crossbin distances showed higher accuracy of prediction than the texture characteristics on post-treatment scans.

The same group, using the same database, analyze a set of features in the characterization of the SUV intensity distribution, spatial patterns (texture), tumour geometry, and changes resulting from CRT [49]. The same rigid registration method was applied, but two VOIs are now semi-automatically delineated, the original tumor in the pre-CRT scans and the residual tumor in the post-CRT scans. The authors conclude that the best traditional measure is the decline in SUV_{max}. In addition, two intensity characteristics (decline in SUV_{mean} and skewness) and three texture characteristics (inertia, correlation, and prominence of the cluster) were considered to be significant predictors. Regarding SUV intensity and tumor size, the changes in these parameters were found to be more predictive than when using either pre or post measures by themselves.

Zhang et al. [50] construct predictive models for the evaluation of tumor response to nCRT. The study included PET/CT scans before and after CRT of 20 patients submitted to trimodality therapy (CRT + surgery). Three groups of tumor characteristics were examined: (1) conventional response measures (SUV_{max}, tumor diameter, etc); (2) clinical parameters (TNM stage, histology, etc) and demographic information; (3) spatial-temporal features from [49]. Each group of characteristics was analyzed individually and also in combination. Feature selection was performed recursively, and as classifiers SVM and LR were used. The SVM model with all features combined achieved an AUC of 1.00 (with no misclassifications).

Hatt *et al.* [51] assemble a database of 555 pre-treatment scans (158 breast, 45 cervix, 112 esophageal, 139 head and neck and 101 lung tumors). Four textural features are then extracted: (1) entropy of GLCM, (2) dissimilarity of GLCM, (3) High-Intensity Large Area Emphasis, and (4) ZP. The correlation between Metabolically Active Tumour Volume (MATV) and textural characteristics varied, with correlation inversely proportional to volume. In the esophageal cohort, only the volumetric dissimilarity of GLCM was a prognostic factor, probably due to small global volumes. General results indicate that heterogeneity and volume may provide supplemental information for volumes $> 10cm^3$.

In [64], the response to chemotherapy is predicted from a single PET scan made prior to treatment. The authors compare the performance of two different strategies: an approach based on statistical classifiers (LR, gradient, RF and SVMs) using more than 100 image descriptors, and an approach based on CNNs, trained directly on the PET scans. The CNN method is named 3S-CNN since it takes sets of three adjacent intra-tumor slices. It is concluded that 3S-CNN exceeds the classical radiomic characteristics in patients with EC. In a data set of 107 patients, the sensitivity and specificity of the 3S-CNN reaches on average, 80.7% and 81.6%, respectively.

Lian *et al.* [52] propose an Evidential Feature Selection (EFS) method, a wrapper feature selection technique based on the Dempster-Shafer Theory (DST). A modified EK-NN (mEK-NN) is also designed. Thirty-six patients with esophageal squamous cell carcinomas were studied. 13 patients were labeled disease free when neither regional or distant tumor recurrence was detected, while the remaining

23 patients were labeled as positive for the disease. Among the pool of features, three were robustly selected: TLG and two clinical features (not specified). The authors conclude that EFS improves classifier performance and achieves the highest accuracy (among the performed experiments) of 89% when combined with mEK-NN.

In a subsequent work [53], the authors improve EFS by creating iEFS (improved EFS). The improvements include: (1) addition of a data balancing methodology, (2) feature selection guided by prior knowledge, and (3) reduction of complexity by modifying the oss function. A classification method based on DS, the Evidential K-NN (EK-NN) rule, is used with subsets of selected features to generate prediction results. The most stable subset of features found is: TLG, tumor staging as II, and patient gender.

In [54] the utility of contour propagation in the context of predicting pathological response is investigated. Tumor ROIs defined by a physician in pre-treatment scans were propagated to the post-treatment PET using rigid and deformable registration algorithms. It was observed that the contours propagated by the inverse consistency Horn-Schunck optical flow algorithm surround the totality of the region of high tumor FDG uptake. Then, GLCM, Grey-Level Run Length matrix (GLRL) and GLSZM textures were computed on all ROIs. The relative difference of each texture at different time points (Δ) were used as predictors. The Δ Short Zone High Gray Emphasis computed with 256 discrete values, led to a significant separation with AUC = 0.78 (q-value = 0.0005). The authors conclude that the ROIs propagated using deformable registration can lead to an accurate prediction of the pathological response, potentially accelerating the texture analysis process.

The same authors, in [55], compare the relationship between characteristics based on PET and pathological response and OS with the measures of SUV. Forty-five patients who underwent surgery were divided into CR, RP and NB to preoperative chemoradiotherapy. SUV_{max} and SUV_{mean}, two GLCM (Entropy and Homogeneity), two GLRL (high-grey-run emphasis and short-run high-grey-run emphasis), and two GLSZM (high-grey-zone emphasis and short-zone high-grey emphasis) features were extracted. The relationship between the relative difference of each measure at different time points of treatment and the pathological response and OS were evaluated. The authors note that all textures, except Homogeneity, were more related to pathological response than either SUV measure. Entropy can distinguish NRs from CRs and PRs, while median entropy and GLRL textures can distinguish patients with good from patients with poor survival.

Rossum *et al.* [56] note that a trustworthy prediction of CR, done prior to surgery, would allow the specialists to attempt to plan a strategy where the organ is preserved. In an effort to achieve this information pre and post-treatment from 217 patients with EAC was gathered, including 18F-FDG PET, subjective assessments, and quantitative parameters. A clinical prediction model was improved in a step-wise manner by adding: (a) subjective assessment of response, (b) TLG after treatment, and (c) four texture/geometry features (baseline cluster shade, ΔGLRL percentage, ΔGLCM entropy, and post-chemoradiation roundness). The authors warn, however, that the improvement obtained does not convert into a clinically relevant gain that can change decision making.

In [57], RFs were investigated to select the best texture features for predictive and prognostic studies. The dataset consisted of 65 EC patients retrospectively collected whose protocol included a pre-treatment full-body scan, treatment with a combined CRT, response assessment 1 month after termination of therapy, follow-up for 3 years after the end of treatment. Patients were divided into CR and non-CR and both medical records and PET images were used for feature extraction, followed by their reduction through Spearman's analysis. This process reduces the initial 61 features to 28. The best prediction subset identified by using RFs is composed of two characteristics: Metabolic Tumor Volume (MTV) and GLCM homogeneity; while the best subset for prognosis was composed of three characteristics: MTV and two clinical characteristics (WHO status and nutritional risk index). The predictive value ($AUC = 0.836$, $Se = 82\%$, $Sp = 91\%$) was higher than using Mann-Whitney ($AUC = 0.810$, $Se = 66\% = 88\%$) and the prognostic achieved ($AUC = 0.822$, $Se = 79\%$, $Sp = 954\%$.

Nakajo *et al.*, do a retrospective study on the use of SUV and texture features, both individually and in combination, to differentiate between benign and Metastatic Adrenal Tumor (MAT). The used database contains 13 BATs and 22 MATs of each 3 are EC. PET/CT was used to extract SUV_{max}, MTV, TLG and four texture features (entropy, homogeneity, IV and SZV). SUV_{max}, entropy and IV were significantly higher in MATs than in BATs; conversely, homogeneity was significantly lower. Accuracy for both SUV_{max} and entropy is 82.9%; slightly lower for IV at 85.7% and the lowest value of accuracy was observed for homogeneity at 71.4%. Binarizing each parameter's result, summing the results and thresholding at 2.5, the sensitivity, specificity and precision were 100%, 84.6% and 94.3%, respectively, with 0.97 of the AUC.

Foley et al. [59] develop a prognostic model that incorporates PET texture analysis. PET images segmentation is performed with the Automatic decision Tree Learning Algorithm for Advanced Segmentation (ATLAAS) tool [65] initialized with a manually defined bounding box. Age, radiological stage, treatment and 16 texture features are included in a Cox regression model designed to predict OS. Results in 403 patients, found 6 variables to be both significantly and independently associated with OS: age, radiological stage, treatment, log(TLG), log(Energy Histogram) and Kurtosis histogram. The authors conclude that PET texture analysis adds prognostic value to EC staging and that texture metrics are associated with OS.

Nakajo *et al.* [60] address the prediction of response to treatment and prognosis based on the heterogeneity of

TABLE 6. Deep learning techniques.

Imaging	Work	Highlights	Architecture	Data Augmentation	Transfer Learning
Endoscopy	[31]	The CNN is used to extract features from image patches, that are then provided to SVM classifiers	5-layer model with 2 convolutional layers, each followed by a pooling layer, and 1 FC layer at the end	Rescaling, rotation, and flipping	No
	[32]	Classification is made into 3 classes	4 convolutional layers, 2 max-pooling layers, and 2 FC layers	Random scaling, flip, rotation, brightness, and contrast	No
	[33]	Intermediate layers of a pre-trained network are used as features for conventional classifiers	AlexNet, VGG16, VGG19 and GoogLeNet trained on ImageNet	No	Transfer Learning by CNN Codes
	[34]	Approach to visualise attention	10 Convolutional and 4 Max-Pooling layers	No	No
CT	[42]	Fully 3D CNN	13 layers in total: 9 convolutional layers in each path, 3 FC layers, and the classification layer	No	No
PET	[64]	Three adjacent PET slices are treated as a three-channel input for the CNN	4 convolutional and 4 max-pooling layers	rotation by $\kappa \times 60°$, $\kappa = 1, 2, 3, 4, 5$	No

the primary tumor in patients treated with CRT. Using (18)F-FDG, several characteristics are extracted, including SUV_{max}, SUV_{mean}, MTV, TLG and 6 texture features. The authors conclude that tumor response can be predicted using two of the six texture features (IV and SZV) and two volumetric parameters (MTV and TLG). For the prognosis task, it was found that all of the studied parameters have a very limited value.

With the goal of improving the prediction of response, a model was constructed in [61] designed to predict the existence of a complete response to nCRT in patients with EC based on pre-treatment clinical parameters and textural features. 97 patients treated with nCRT followed by esophagectomy were considered. Both 18F-FDG PET and CT were used to derive clinical, geometrical and texture features. The prediction models were trained with Least Absolute Shrinkage and Selection Operator regularized LR. The pathological response was defined as Complete vs. Incomplete (Standard Tumour Regression Degree - TRG - system 1 vs. 2-5). The characteristics selected by the regularization include: histological type and clinical T-stage, long run low grey level emphasis extracted from PET, and run percentage extracted from CT. LR results with the selected variables achieved an AUC of 0.74, while an SUV_{max} model only reached 0.54.

In [62], sixty-four scans were processed using 5 different levels of Gaussian smoothing, 4 SUV_{max} segmentation thresholds and 5 different quantization levels. Heterogeneity parameters based on GLCM, GLRL, Gray Tone Neighborhood Matrix (NGTDM), GLSZM and fractal analysis methods were then extracted. Most parameters showed low agreement between different widths of quantization bins, while segmentation and smoothing have a smaller impact on precision.

Anthony *et al.* [63] examine the link between the development of symptomatic RP in patients with EC after RT using LR. Twenty texture features (first order, fractal, Laws' filter and GLCM) were extracted from diagnostic CT scans in anatomically corresponding regions of the lung. Mean change in the value of each texture feature and the standard deviation of pre-therapy SUV (SUV_{SD}) were then calculated. Although the clinical parameters (mean lung dose, smoking history, tumor location) were not significantly different between patients with and without symptomatic RP; SUV and texture parameters were significantly associated with RP status. When training the LR model with a single texture feature, AUC values ranged from 0.58 to 0.81. Using SUV_{SD} alone, achieves an AUC of 0.69. Combining SUV_{SD} with each one of the texture features increases the AUC by 0.04-0.08.

IV. DISCUSSION

As seen in the above section, independent of the type of image, the techniques used are mostly based on hand-crafted features or DL techniques. Since there are already several state of the art reviews on traditional pipelines, of which we highlight [66], [67], we will here provide a brief summary on DL techniques only (Section IV-A). This section also includes a summary of existing publicly available databases (Section IV-B).

A. DEEP LEARNING TECHNIQUES

The performance of CNNs has been shown to be close to that of humans in computer vision tasks such as on the ImageNet tests [68]. It is thus not surprising that they are the natural choice when dealing with other computer vision tasks, such as the ones on esophagus cancer. As can be seen in Table 6, Endoscopy is the imaging type with more published DL works; while CT and PET have been used in one work each.

In summary, DL works all use CNNs to process imaging data. However, the chosen CNN architecture is

TABLE 7. Publicly available databases.

Imaging	Database	Short description	Works
Endoscopy	MICCAI 2015 "EndoVis" [6]	The focus of the sub-challenge "Early Barret's cancer detection" in on such supportive algorithms for the detection of early cancerous lesions in Barrets Esophagus. The database consists on a set of 100 HD endoscopic images from 39 patients, annotated by five expert endoscopists.	[30, 33]
	ISBI 2016 "AIDA-E" [7]	The focus of the sub-challenge "Confocal Endoscopy in Barret's esophagus" is the classification of the images into one of the four histology classes (GMP, IM or proper BE, or NPL). The database consists on a set of 262 confocal images from 32 patients and 81 different bioptic sites.	[32]
	esophageal IPCL	17 monocular videos (7,046 frames from 17 patients - one video per patient)	[34]
	ISBI 2019 "EAD" [8]	The focus of this challenge is to detect multi-class artefacts in video endoscopy. As of the date of writting the dataset is not yet fully available and we thus refer to the site for more details.	-
	"GastroAtlas" [9]	At the time of writing, this atlas displays 4,836 video clips, covering almost all areas of GI pathology detectable by endoscopy.	-
CT	MICCAI 2015 "Synapse" [10]	The focus of the sub-dataset "Abdomen" subdataset is to label 13 abdominal organs, including the esophagus. The sub-dataset consists on a set of 50 abdomen CT scans.	[42]
	TCGA-ESCA [11]	This dataset is part of a larger effort to build a research community focused on connecting cancer phenotypes to genotypes by providing clinical images matched to subjects from TCGA. The database consists on a set of 17 studies from 16 patients.	-
PET	not found	-	-

[6] Available at endovis.grand-challenge.org
[7] Available at isbi-aida.grand-challenge.org
[8] Available at ead2019.grand-challenge.org
[9] Available at http://www.gastrointestinalatlas.com
[10] Available at www.synapse.org
[11] Available at https://wiki.cancerimagingarchive.net/display/Public/TCGA-ESCA

typically customised, while well-known architectures such as AlexNet, VGG, and GoogLeNet are only used in a Transfer Learning Context [33]. To cope with the well known need of DL techniques for a lot of training data [69], several works use data augmentation techniques [31], [32], [64], while [33] uses Transfer Learning.

B. PUBLICLY AVAILABLE DATABASES

A list of publicly available datasets can be found in Table 7. It is clear that endoscopy is the modality that has raised most interest in the community, followed by CT. To the best of our knowledge, there are no publicly available databases for PET. Another characteristic of the available datasets it their reduced size, with typically fewer than 50 patients.

V. CONCLUSION

Esophageal cancer is a disease with a high prevalence which can be evaluated by a variety of imaging modalities, including Endoscopy, CT, and PET. Computer vision (artificial intelligence, machine learning, classification using deep learning, among others) techniques could provide valuable help in the analysis of these images decreasing the medical workflow time and enhancing diagnostic and staging accuracy. Current guidelines for esophageal treatment typically include neoadjuvant radiochemotherapy followed by surgery. In some cases however it is known, from post surgery anatomopathologic data, that surgery could be avoided in some patients. Computer vision could present solutions aiming to find criteria for the clear separation between surgical and non-surgical

candidates, avoiding the loss of quality of life and surgery associated comorbilities. In this work, an extensive literature review on computer vision applied to esophageal cancer is presented. Major results are summarized, compared, and main shortcomings are identified. Forty seven works are reviewed in this survey, covering different image modalities. It was observed that, although with different end goals, most of the existing works use similar techniques, based on traditional machine learning with handcrafted-features. Even though the computer vision scientific community has already paid attention to the esophageal cancer problem, there are still several open issues that can be further developed and solved in the future.

REFERENCES

[1] J. Ferlay, I. Soerjomataram, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D. M. Parkin, D. Forman, and F. Bray, "Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012," *Int. J. Cancer*, vol. 136, pp. E359–E386, 2015. doi: 10.1002/ijc.29210.

[2] M. C. S. Wong, W. Hamilton, D. C. Whiteman, J. Y. Jiang, Y. Qiao, F. D. H. Fung, H. H. X. Wang, P. W. Y. Chiu, E. K. W. Ng, J. C. Y. Wu, J. Yu, F. K. L. Chan, and J. J. Y. Sung, "Global Incidence and mortality of oesophageal cancer and their correlation with socioeconomic indicators temporal patterns and trends in 41 countries," *Sci. Rep.*, vol. 8, no. 1, 2018, Art. no. 4522.

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

[4] R. M. Gore and M. S. Levine, "Diseases of the upper GI tract," in *Diseases of the Abdomen and Pelvis*, J. Hodler, R. A. Kubik-Huch, and G. K. von Schulthess, Eds. Cham, Switzerland: Springer, 2018, pp. 91–98. doi: 10.1007/978-3-319-75019-4_10.

[5] M. B. Amin, S. Edge, F. Greene, D. R. Byrd, R. K. Brookland, M. K. Washington, J. E. Gershenwald, C. C. Compton, K. R. Hess, D. C. Sullivan, J. M. Jessup, J. D. Brierley, L. E. Gaspar, R. L. Schilsky, C. M. Balch, D. P. Winchester, E. A. Asare, M. Madera, D. M. Gress, and L. R. Meyer, *AJCC Cancer Staging Manual*, 8th ed. Springer, 2017.

[6] J. D. Brierley, M. K. Gospodarowicz, and C. Wittekind, *TNM Classification of Malignant Tumours*, 8th ed. Hoboken, NJ, USA: Wiley, 2017.

[7] H. L. van Westreenen, "Positron emission tomography in staging of esophageal cancer," Ph.D. dissertation, Dept. Surg. Oncol., Univ. Groningen, Groningen, The Netherlands, 2005.

[8] T. W. Rice, V. W. Rusch, H. Ishwaran, and E. H. Blackstone, "Cancer of the esophagus and esophagogastric junction: Data-driven staging for the seventh edition of the American Joint Committee on Cancer/International Union Against Cancer Cancer Staging Manuals," *Cancer*, vol. 116, no. 16, pp. 3763–3773, Aug. 2010.

[9] K. Foley, J. Findlay, and V. Goh, "Novel imaging techniques in staging oesophageal cancer," *Best Pract. Res. Clin. Gastroenterol.*, vols. 36–37, pp. 17–25, Oct./Dec. 2018.

[10] M. Massari, U. Cioffi, M. De Simone, E. Lattuada, M. Montorsi, A. Segalin, and L. Bonavina, "Endoscopic ultrasonography for preoperative staging of esophageal carcinoma," *Surgical Laparoscopy Endoscopy*, vol. 7, no. 2, pp. 162–165, 1997.

[11] P. S. N. Van Rossum, "Towards individualized treatment for esophageal cancer," Ph.D. dissertation, Dept. Radiat. Oncol. Surg., Utrecht Univ., Utrecht, The Netherlands, 2016.

[12] R. Ohura, H. Omura, Y. Sakata, and T. Minamoto, "Computer-aided diagnosis method for detecting early esophageal cancer from endoscopic image by using dyadic wavelet transform and fractal dimension," in *Proc. 13th Int. Conf. New Gener.*, S. Latifi, Ed. Cham, Switzerland: Springer, 2016, pp. 929–938.

[13] J. D. Luketich, D. M. Friedman, T. L. Weigel, M. A. Meehan, R. J. Keenan, D. W. Townsend, and C. C. Meltzer, "Evaluation of distant metastases in esophageal cancer: 100 consecutive positron emission tomography scans," *Ann. Thoracic Surg.*, vol. 68, no. 4, pp. 1133–1137, 1999.

[14] F. Lordick, C. Mariette, K. Haustermans, R. Obermannová, and D. Arnold, "Oesophageal cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up," *Ann. Oncol.*, vol. 27, no. 5, pp. v50–v57, 2016.

[15] V. Malik, M. Harmon, C. Johnston, A. J. Fagan, Z. Claxton, N. Ravi, D. O'Toole, C. Muldoon, M. Keogan, J. V. Reynolds, and J. F. M. Meaney, "Whole body MRI in the staging of esophageal cancer—A prospective comparison with whole body 18F-FDG PET-CT," *Digestive Surg.*, vol. 32, no. 5, pp. 397–408, 2015.

[16] F. Tixier, C. C. Le Rest, M. Hatt, N. Albarghach, O. Pradier, J.-P. Metges, L. Corcos, and D. Visvikis, "Intratumor heterogeneity characterized by textural features on baseline ^{18}F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer," *J. Nucl. Med.*, vol. 52, no. 3, pp. 369–378, 2011.

[17] B. Münzer, K. Schoeffmann, and L. Böszörmenyi, "Content-based processing and analysis of endoscopic images and videos: A survey," *Multimedia Tools Appl.*, vol. 77, no. 1, pp. 1323–1362, 2018.

[18] R. E. Carroll and S. M. Seitz, "Rectified surface mosaics," *Int. J. Comput. Vis.*, vol. 85, no. 3, pp. 307–315, Dec. 2009.

[19] S. S. Saraf, C. R. Udupi, and S. D. Hajare, "Decision support system based on DCT texture features for diagnosis of esophagitis," *J. Mech. Med. Biol.*, vol. 9, no. 4, pp. 527–538, 2009.

[20] F. van der Sommen, S. Zinger, E. J. Schoon, and P. H. N. de With, "Computer-aided detection of early cancer in the esophagus using HD endoscopy images," *Proc. SPIE*, vol. 8670, Feb. 2013, Art. no. 86700V.

[21] F. van der Sommen, S. Zinger, and E. J. Schoon, "Supportive automatic annotation of early esophageal cancer using local Gabor and color features," *Neurocomputing*, vol. 144, pp. 92–106, Nov. 2014.

[22] F. van der Sommen, S. Zinger, W. L. Curvers, R. Bisschops, O. Pech, B. L. Weusten, J. J. Bergman, and E. J. Schoon, "Computer-aided detection of early neoplastic lesions in Barrett's Esophagus," *Endoscopy*, vol. 48, no. 7, pp. 617–624, Jul. 2016.

[23] F. van der Sommen, S. Zinger, E. J. Schoon, and P. H. de With, "Sweet-spot training for early esophageal cancer detection," *Proc. SPIE*, vol. 9785, Mar. 2016, Art. no. 97851B.

[24] H. Matsunaga, H. Omura, R. Ohura, and T. Minamoto, "Daubechies wavelet-based method for early esophageal cancer detection from flexible spectral imaging color enhancement image," in *Proc. 13th Int. Conf. Inf. Technol., New Gener.*, S. Latifi, Ed. Cham, Switzerland: Springer, 2016, pp. 939–948.

[25] M. H. A. Janse, F. van der Sommen, S. Zinger, E. J. Schoon, and P. H. N. de With, "Early esophageal cancer detection using RF classifiers," *Proc. SPIE*, vol. 9785, Mar. 2016, Art. no. 97851D.

[26] D. Shin, M. H. Lee, A. D. Polydorides, M. C. Pierce, P. M. Vila, N. D. Parikh, D. G. Rosen, S. Anandasabapathy, and R. R. Richards-Kortum, "Quantitative analysis of high-resolution microendoscopic images for diagnosis of neoplasia in patients with Barrett's esophagus," *Gastrointest Endoscopy*, vol. 83, no. 1, pp. 107–114, 2016.

[27] N. Ghatwary, A. Ahmed, X. Ye, and H. Jalab, "Automatic grade classification of Barretts Esophagus through feature enhancement," *Proc. SPIE*, vol. 10134, Mar. 2017, Art. no. 1013433.

[28] A.-F. Swager, F. van der Sommen, S. R. Klomp, S. Zinger, S. L. Meijer, E. J. Schoon, J. J. G. H. M. Bergman, P. H. de With, and W. L. Curvers, "Computer-aided detection of early Barrett's neoplasia using volumetric laser endomicroscopy," *Gastrointestinal endoscopy*, vol. 86, no. 5, pp. 839–846, 2017.

[29] S. Klomp, F. van der Sommen, A.-F. Swager, S. Zinger, E. J. Schoon, W. L. Curvers, J. J. Bergman, and P. H. N. de With, "Evaluation of image features and classification methods for barrett's cancer detection using VLE imaging," *Proc. SPIE*, vol. 10134, Mar. 2017, Art. no. 101340D.

[30] L. A. De Souza, L. C. S. Afonso, C. Palm, and J. P. Papa, "Barrett's esophagus identification using optimum-path forest," in *Proc. 30th Conf. Graph., Patterns Images (SIBGRAPI)*, Oct. 2017, pp. 308–314.

[31] D.-X. Xue, R. Zhang, H. Feng, and Y.-L. Wang, "CNN-SVM for microvascular morphological type recognition with data augmentation," *J. Med. Biol. Eng.*, vol. 36, no. 6, pp. 755–764, 2016.

[32] J. Hong, B.-Y. Park, and H. Park, "Convolutional neural network classifier for distinguishing Barrett's esophagus and neoplasia endomicroscopy images," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2017, pp. 2892–2895.

[33] S. V. Riel, F. Van Der Sommen, S. Zinger, E. J. Schoon, and P. H. de With, "Automatic detection of early esophageal cancer with CNNs using transfer learning," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 1383–1387.

[34] L. C. Garcia-Peraza-Herrera, M. Everson, W. Li, I. Luengo, L. Berger, O. Ahmad, L. Lovat, H.-P. Wang, W.-L. Wang, R. Haidry, D. Stoyanov, T. Vercauteren, and S. Ourselin, "Interpretable fully convolutional classification of intrapapillary capillary loops for real-time detection of early squamous neoplasia," 2018, *arXiv:1805.00632*. [Online]. Available: https://arxiv.org/abs/1805.00632

[35] A. S. Vemuri, "Inter-operative biopsy site relocalization in gastroscopy: Application to oesophagus," Ph.D. dissertation, Ecole Doctorale Sci. Technol. l'Inf. Commun., Nice Sophia Antipolis Univ., Nice, France, 2016.

[36] Z. Zhang, L. Bai, P. Ren, and E. R. Hancock, "High-order graph matching kernel for early carcinoma EUS image classification," *Multimedia Tools Appl.*, vol. 75, no. 7, pp. 3993–4012, 2016.

[37] N. Ghatwary, A. Ahmed, and X. Ye, "Automated detection of Barrett's esophagus using endoscopic images: A survey," in *Proc. Annu. Conf. Med. Image Understand. Anal. (MIUA)*. Cham, Switzerland: Springer, 2017, pp. 897–908.

[38] C. Yip, D. Landau, R. Kozarski, B. Ganeshan, R. Thomas, A. Michaelidou, and V. Goh, "Primary esophageal cancer: Heterogeneity as potential prognostic biomarker in patients treated with definitive chemotherapy and radiation therapy," *Radiology*, vol. 270, no. 1, pp. 141–148, 2014.

[39] A. Cunliffe, S. G. Armato, III, R. Castillo, N. Pham, T. Guerrero, and H. A. Al-Hallaq, "Lung texture in serial thoracic computed tomography scans: Correlation of radiomics-based features with radiation therapy dose and radiation pneumonitis development," *Int. J. Radiat. Oncol., Biol., Phys.*, vol. 91, no. 5, pp. 1048–1056, 2015.

[40] C. Yip, F. Davnall, R. Kozarski, D. B. Landau, G. J. R. Cook, P. Ross, R. Mason, and V. Goh, "Assessment of changes in tumor heterogeneity following neoadjuvant chemotherapy in primary esophageal cancer," *Diseases Esophagus*, vol. 28, no. 2, pp. 172–179, 2015.

[41] R. Trullo, C. Petitjean, D. Nie, D. Shen, and S. Ruan, "Fully automated esophagus segmentation with a hierarchical deep learning approach," in *Proc. IEEE Int. Conf. Signal Image Process. Appl. (ICSIPA)*, Sep. 2017, pp. 503–506.

[42] T. Fechter, S. Adebahr, D. Baltas, I. B. Ayed, C. Desrosiers, and J. Dolz, "Esophagus segmentation in CT via 3D fully convolutional neural network and random walk," *Med. Phys.*, vol. 44, no. 12, pp. 6341–6352, Dec. 2017.

[43] P. Martins, I. Carbone, A. Silva, and A. J. S. Teixeira, "An MRI study of European Portuguese nasals," in *Proc. INTERSPEECH*, 2007, pp. 58–61.

[44] F. Tixier, M. Hatt, C. C. Le Rest, A. Le Pogam, L. Corcos, and D. Visvikis, "Reproducibility of tumor uptake heterogeneity characterization through textural feature analysis in ^{18}F-FDG PET," *J. Nucl. Med.*, vol. 53, no. 5, pp. 693–700, 2012.

[45] B. Ganeshan, K. Skogen, I. Pressney, D. Coutroubis, and K. Miles, "Tumour heterogeneity in oesophageal cancer assessed by CT texture analysis: Preliminary evidence of an association with tumour metabolism, stage, and survival," *Clin. Radiol.*, vol. 67, no. 2, pp. 157–164, 2012.

[46] M. Hatt, F. Tixier, C. C. Le Rest, O. Pradier, and D. Visvikis, "Robustness of intratumour ^{18}F-FDG PET uptake heterogeneity quantification for therapy response prediction in oesophageal carcinoma," *Eur. J. Nucl. Med. Mol. Imag.*, vol. 40, no. 11, pp. 1662–1671, 2013.

[47] X. Dong, L. Xing, P. Wu, Z. Fu, H. Wan, D. Li, Y. Yin, X. Sun, and J. Yu, "Three-dimensional positron emission tomography image texture analysis of esophageal squamous cell carcinoma: Relationship between tumor ^{18}F-fluorodeoxyglucose uptake heterogeneity, maximum standardized uptake value, and tumor stage," *Nucl. Med. Commun.*, vol. 34, no. 1, pp. 40–46, 2013.

[48] S. Tan, H. Zhang, Y. Zhang, W. Chen, W. D. D'Souza, and W. Lu, "Predicting pathologic tumor response to chemoradiotherapy with histogram distances characterizing longitudinal changes in ^{18}F-FDG uptake patterns," *Med. Phys.*, vol. 40, no. 10, 2013, Art. no. 101707.

[49] S. Tan, S. Kligerman, W. Chen, M. Lu, G. Kim, S. Feigenberg, W. D. D'Souza, M. Suntharalingam, and W. Lu, "Spatial-temporal [^{18}F]FDG-PET features for predicting pathologic response of esophageal cancer to neoadjuvant chemoradiation therapy," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 85, no. 5, pp. 1375–1382, 2013.

[50] H. Zhang, S. Tan, W. Chen, S. Kligerman, G. Kim, W. D. D'Souza, M. Suntharalingam, and W. Lu, "Modeling pathologic response of esophageal cancer to chemoradiation therapy using spatial-temporal ^{18}F-FDG PET features, clinical parameters, and demographics," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 88, no. 1, pp. 195–203, 2014.

[51] M. Hatt, M. Majdoub, M. Vallières, F. Tixier, C. C. Le Rest, D. Groheux, E. Hindié, A. Martineau, O. Pradier, R. Hustinx, R. Perdrisot, R. Guillevin, I. El Naqa, and D. Visvikis, "^{18}F-FDG PET uptake characterization through texture analysis: Investigating the complementary nature of heterogeneity and functional tumor volume in a multi–cancer site patient cohort," *J. Nucl. Med.*, vol. 56, no. 1, pp. 38–44, 2015.

[52] C. Lian, S. Ruan, T. Denæux, and P. Vera, "Outcome prediction in tumour therapy based on Dempster-Shafer theory," in *Proc. IEEE 12th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2015, pp. 63–66.

[53] C. Lian, S. Ruan, T. Denæux, F. Jardin, and P. Vera, "Selecting radiomic features from FDG-PET images for cancer treatment outcome prediction," *Med. Image Anal.*, vol. 32, pp. 257–268, Aug. 2016.

[54] S. S. F. Yip, T. P. Coroller, N. N. Sanford, E. Huynh, H. Mamon, H. J. W. L. Aerts, and R. I. Berbeco, "Use of registration-based contour propagation in texture analysis for esophageal cancer pathologic response prediction," *Phys. Med. Biol.*, vol. 61, no. 2, pp. 906–922, 2016.

[55] S. S. F. Yip, T. P. Coroller, N. N. Sanford, H. Mamon, H. J. W. L. Aerts, and R. I. Berbeco, "Relationship between the temporal changes in positron-emission-tomography-imaging-based textural features and pathologic response and survival in esophageal cancer patients," *Frontiers Oncol.*, vol. 6, p. 72, Mar. 2016.

[56] P. S. N. van Rossum, D. V. Fried, L. Zhang, W. L. Hofstetter, M. van Vulpen, G. J. Meijer, L. E. Court, and S. H. Lin, "The incremental value of subjective and quantitative assessment of ^{18}F-FDG PET for the prediction of pathologic complete response to preoperative chemoradiotherapy in esophageal cancer," *J. Nucl. Med.*, vol. 57, no. 5, pp. 691–700, 2016.

[57] P. Desbordes, S. Ruan, R. Modzelewski, P. Pineau, S. Vauclin, P. Gouel, P. Michel, F. Di Fiore, P. Vera, and I. Gardin, "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*, vol. 12, no. 3, 2017, Art. no. e0173208.

[58] M. Nakajo, M. Jinguji, M. Nakajo, T. Shinaji, Y. Nakabeppu, Y. Fukukura, and T. Yoshiura, "Texture analysis of FDG PET/CT for differentiating between FDG-avid benign and metastatic adrenal tumors: Efficacy of combining SUV and texture parameters," *Abdominal Radiol.*, vol. 42, no. 12, pp. 2882–2889, 2017.

[59] K. G. Foley, R. K. Hills, B. Berthon, C. Marshall, C. Parkinson, W. G. Lewis, T. D. L. Crosby, E. Spezi, and S. A. Roberts, "Development and validation of a prognostic model incorporating texture analysis derived from standardised segmentation of PET in patients with oesophageal cancer," *Eur. Radiol.*, vol. 28, no. 1, pp. 428–436, 2018.

[60] M. Nakajo, M. Jinguji, Y. Nakabeppu, M. Nakajo, R. Higashi, Y. Fukukura, K. Sasaki, Y. Uchikado, S. Natsugoe, and T. Yoshiura, "Texture analysis of ^{18}F-FDG PET/CT to predict tumour response and prognosis of patients with esophageal cancer treated by chemoradiotherapy," *Eur. J. Nucl. Med. Mol. Imag.*, vol. 44, no. 2, pp. 206–214, 2017.

[61] R. J. Beukinga, J. B. Hulshoff, L. V. van Dijk, C. T. Muijs, J. G. M. Burgerhof, G. Kats-Ugurlu, R. H. J. A. Slart, C. H. Slump, V. E. M. Mul, and J. T. M. Plukke, "Predicting response to neoadjuvant chemoradiotherapy in esophageal cancer with textural features derived from pretreatment ^{18}F-FDG PET/CT imaging," *J. Nucl. Med.*, vol. 58, no. 5, pp. 723–729, 2016.

[62] G. Doumou, M. Siddique, C. Tsoumpas, V. Goh, and G. J. Cook, "The precision of textural analysis in ^{18}F-FDG-PET scans of oesophageal cancer," *Eur. Radiol.*, vol. 25, no. 9, pp. 2805–2812, 2015.

[63] G. J. Anthony, A. Cunliffe, R. Castillo, N. Pham, T. Guerrero, S. G. Armato, III, and H. A. Al-Hallaq, "Incorporation of pre-therapy ^{18}F-FDG uptake data with CT texture features into a radiomics model for radiation pneumonitis diagnosis," *Med. Phys.*, vol. 44, no. 7, pp. 3686–3694, 2017.

[64] P.-P. Ypsilantis, M. Siddique, H.-M. Sohn, A. Davies, G. Cook, V. Goh, and G. Montana, "Predicting response to neoadjuvant chemotherapy with PET imaging using convolutional neural networks," *PLoS One*, vol. 10, no. 9, 2015, Art. no. e0137036.

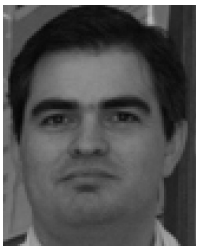
[65] B. Berthon, C. Marshall, M. Evans, and E. Spezi, "ATLAAS: An automatic decision tree-based learning algorithm for advanced image segmentation in positron emission tomography," *Phys. Med. Biol.*, vol. 61, no. 13, pp. 4855–4869, 2016.

[66] A. Zwanenburg, S. Leger, M. Vallières, and S. Löck, "Image biomarker standardisation initiative," 2016, *arXiv:1612.07003*. [Online]. Available: https://arxiv.org/abs/1612.07003

[67] M. A. Nogueira, P. H. Abreu, P. Martins, P. Machado, H. Duarte, and J. Santos, "Image descriptors in radiology images: A systematic review," *Artif. Intell. Rev.*, vol. 47, no. 4, pp. 531–559, 2017.
[68] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, 2015.
[69] I. Domingues, P. H. Abreu, and J. Santos, "BI-RADS classification of breast cancer: A new pre-processing pipeline for deep models training," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 1378–1382.

INÊS DOMINGUES graduated in applied mathematics from the School of Sciences, University of Porto, in 2004, and received the master's degree in electrical and telecommunications engineering from Aveiro University, in 2008, and the Ph.D. degree *(cum laude)* in electrical and computer engineering from the School of Engineering, University of Porto, in 2015. She is currently a Postdoctoral Researcher in the ESTIMA Project, a member of the Medical Physics, Radiobiology and Radiation Protection Group, IPO Porto Research Centre (CI-IPOP), and an Invited Assistant Professor with DEIS-ISEC.

INÊS LUCENA SAMPAIO received the M.D. degree. He has been a Nuclear Medicine Physician Assistant with the Portuguese Institute of Oncology, Porto, since 2013, a member of the Research Group of Medical Physics, Radiobiology and Radiation Protection, Porto Research Centre, Portuguese Institute of Oncology, since 2016, and a member of the Neuroendocrine Tumors Study Group, Portuguese Society of Endocrinology, since 2016. His research interests include neuroendocrine tumors and response therapy evaluation with positron emission tomography, with publications and congress participation as a speaker in these fields.



HUGO DUARTE is specialized in nuclear medicine and he is the Head of the Nuclear Medicine Service of IPO Porto. He collaborates in many research projects as a member of the Research Centre of IPO Porto. He has authored more than 30 papers in the oncology area.

JOÃO A. M. SANTOS received the M.Sc. degree in physics (optoelectronics and lasers) and the Ph.D. degree in physics (condensed matter). He concluded the Residency in Medical Physics at the Portuguese Institute of Oncology of Porto (IPOPFG, EPE), in 2005. He is currently a Medical Physicist Expert and a Coordinator of the Medical Physics, Radiobiology and Radiation Protection Group, IPOPFG Research Centre, and an Affiliated Professor with the University of Porto.

PEDRO H. ABREU received the Informatics Engineering degree and the Ph.D. degree in soccer teams modeling from the University of Porto, in 2006 and 2011, respectively. He is currently an Assistant Professor with the Department of Informatics Engineering, University of Coimbra. He has authored more than 60 publications in international conferences and journals. His research interests include medical informatics and personal healthcare systems applied to oncology.