Machine Learning Applications for Head and Neck Imaging

Article *in* Neuroimaging Clinics of North America · November 2020

DOI: 10.1016/j.nic.2020.08.003

5 authors, including:

Farhad Maleki
The University of Calgary
37 PUBLICATIONS **298** CITATIONS
SEE PROFILE

William Trung Le
University of Montreal Hospital Research Centre
13 PUBLICATIONS **38** CITATIONS
SEE PROFILE

Thiparom Sananmuang
Mahidol University
9 PUBLICATIONS **29** CITATIONS
SEE PROFILE

Samuel Kadoury
Polytechnique Montréal
204 PUBLICATIONS **4,586** CITATIONS
SEE PROFILE

Machine Learning Applications for Head and Neck Imaging

Farhad Maleki, PhD[a], William Trung Le, BSc[b], Thiparom Sananmuang, MD[c], Samuel Kadoury, PhD[b,d], Reza Forghani, MD, PhD[a,e,f,g,h,*]

KEYWORDS

• Head and neck imaging • Autosegmentation • Classification • Machine learning • Deep learning • Convolutional neural network • Artificial intelligence • Head and neck cancer

KEY POINTS

- Data from head and neck (HN) imaging alongside clinical data can be used to build predictive models improving noninvasive characterization of HN cancers and other anomalies.
- Machine learning (ML) has shown the potential for building predictive models that lead to improved HN tumor characterization, prediction of treatment response, and survival.
- Successful deployment of ML-based models in a clinical setting will require algorithms based on large-scale studies that encompass various sites and variations enabling the future ability to generalize.

INTRODUCTION

Head and neck (HN) imaging is concerned with the evaluation of disorders affecting the complex structures and spaces of the neck, paranasal sinuses, skull base, and the orbits. Like any other anatomic area, the HN can be affected by a large variety of neoplastic and nonneoplastic disorders. However, the HN is notable for its complex anatomy and the concentration of critical organs in close proximity to one another, making advanced imaging and highly specialized expert interpretations particularly important for the diagnosis and treatment of different disorders affecting this region. Cross-sectional imaging plays a fundamental role in the evaluation of major disorders affecting the HN.

For the work-up of HN lesions such as HN cancer, imaging can be used to identify tumors and, when appropriate, suggest a differential diagnosis for distinguishing benign from malignant lesions. Furthermore, in HN cancer, one of the most fundamental roles of cross-sectional imaging is to accurately determine the stage of a tumor, including

Funding: R. Forghani is a clinical research scholar (chercheur-boursier clinicien) supported by the Fonds de recherche en santé du Québec (FRQS) and has an operating grant jointly funded by the FRQS and the Fondation de l'Association des radiologistes du Québec (FARQ).

[a] Augmented Intelligence & Precision Health Laboratory (AIPHL), Department of Radiology & Research Institute of the McGill University Health Centre, 5252 Boulevard de Maisonneuve Ouest, Montreal, Quebec H4A 3S5, Canada; [b] Polytechnique Montreal, PO Box 6079, succ. Centre-ville, Montreal, Quebec H3C 3A7, Canada; [c] Department of Diagnostic and Therapeutic Radiology and Research, Faculty of Medicine Ramathibodi Hospital, Ratchathewi, Bangkok 10400, Thailand; [d] CHUM Research Center, 900 St Denis Street, Montreal, Quebec H2X 0A9, Canada; [e] Department of Radiology, McGill University, 1650 Cedar Avenue, Montreal, Quebec H3G1A4, Canada; [f] Segal Cancer Centre, Lady Davis Institute for Medical Research, Jewish General Hospital, 3755 Cote Ste-Catherine Road, Montreal, Quebec H3T 1E2, Canada; [g] Gerald Bronfman Department of Oncology, McGill University, Suite 720, 5100 Maisonneuve Boulevard West, Montreal, Quebec H4A3T2, Canada; [h] Department of Otolaryngology, Head and Neck Surgery, Royal Victoria Hospital, McGill University Health Centre, 1001 boul. Decarie Boulevard, Montreal, Quebec H3A 3J1, Canada

* Corresponding author. Room C02.5821, 1001 Decarie Blvd, Montreal, Quebec H4A 3J1, Canada.
E-mail address: reza.forghani@mcgill.ca

Neuroimag Clin N Am 30 (2020) 517–529
https://doi.org/10.1016/j.nic.2020.08.003

evaluation of deep spaces in the neck that may not be reliably evaluated by clinical physical examination and endoscopy, upstaging the initial clinical assessment when appropriate. After treatment, imaging is an essential diagnostic tool for the surveillance of tumor recurrence or differentiation of tumor recurrence from treatment-related complications.[1] In most North American institutions, contrast-enhanced computed tomography (CT) scans, magnetic resonance (MR) imaging, and to a lesser extent PET are the mainstay advanced imaging modalities used for initial evaluation and follow-up of nonthyroid malignancies of the HN, whereas ultrasonography is typically the first modality used for the evaluation of thyroid lesions.

In current routine clinical practice, evaluation or interpretation of patient diagnostic medical scans is largely qualitative with only limited use of basic quantitative parameters for lesion evaluation and characterization. Cancer staging is performed using the American Joint Committee on Cancer (AJCC) TNM system, which classifies cancers based on the size and extent of the primary tumor (T), involvement of regional lymph nodes (N), and the presence or absence of distant metastases (M).[2–4] Reflecting the AJCC classification, tumor assessment and staging is largely based on a lesion's anatomic extent, although, in its revised classification that came into effect in 2018, the AJCC created a separate staging system for human papillomavirus (HPV)–associated cancers of the oropharynx, incorporating a molecular marker for tumor staging that reflects the distinct clinicopathologic characteristics of this subtype of head and neck squamous cell carcinoma (HNSCC).[4,5] Although delineation of the anatomic extent of a tumor is clearly important, there is little use of the additional complex quantitative features present on patients' diagnostic images, and incorporation of molecular markers is only rudimentary at this time. The current process is also subjective and depends on the radiologist's or radiation oncologist's level of expertise and experience, resulting in variations that may not be optimal for patient care. With the advances made in computational power and software engineering, there is great potential for leveraging this technology for improving or augmenting diagnostic evaluation in head and neck imaging toward the ultimate goal of more precise, personalized therapy.

This article reviews the various applications of machine learning (ML) in HN imaging. The article begins with a brief overview of the traditional radiomic-ML workflow versus deep learning, followed by a review of several applications of ML in HN imaging. Next, it discusses the challenges and opportunities in using ML for HN imaging, followed by a brief summary.

RADIOMICS AND MACHINE LEARNING

There has been long-standing interest in computerized image analysis for diagnostic decision support, but developments have been accelerated and the potential expanded as a result of the impressive advances in computational power and software engineering in the last decade. One area of much interest and research activity is in texture analysis or, more broadly, radiomics. First introduced in the medical literature in 2012 and defined as high-throughput extraction of quantitative imaging features with the intent of creating mineable databases from radiological images, the definition of radiomics was later expanded to "high-throughput extraction of quantitative features that result in the conversion of images into mineable data and the subsequent analysis of these data for decision support."[6–9] Radiomic approaches perceive medical imaging as a rich source for mining quantitative features as opposed to pictures intended solely for visual and subjective interpretation.[6] These image-based quantitative features, which are referred to as radiomics features, can provide important additional information and serve as biomarkers enabling higher-level characterization of lesions such as tumors, including prediction of different clinical or molecular end points of interest that may not be evident using standard subjective interpretation alone. Radiomic features may even be considered as image extracted omics that can be used along with other features, such as clinical, genomic, or proteomic information, to capture a holistic picture of the lesion phenotype under study.[6,10–12]

Machine learning is commonly used for predictive modeling in radiomics.[13,14] ML refers to quantitative algorithms capable of learning a task given data related to that task.[15] A performance measure, which is commonly referred to as an objective function, error function, or loss function, is used to evaluate and guide the learning process. It is important to emphasize that radiomic feature extraction itself does not necessarily require artificial intelligence (AI) or, more specifically, ML; many of the published texture or radiomic studies are based on more traditional and computer vision approaches derived using clearly defined or explicit mathematical formulas designed by experts, collectively referred to as handcrafted or hand-engineered features. However, ML classifiers are particularly useful for performing prediction modeling or classification tasks based on

radiomic features, with examples including tumor classification, prediction of treatment response, or survival among a multitude of possible tasks. Also, ML approaches such as convolutional neural networks (CNNs) can also be used for direct image analysis and feature extraction, a process that may be referred to as deep radiomics, as discussed in this article.[11,12]

This article categorizes applications of ML in HN cancers as those using traditional ML and those using deep learning methods. Methods such as linear regression, support vector machine (SVM), and random forest are considered as traditional ML. In contrast, methods using neural networks with more than 3 hidden layers are considered as deep learning methods. **Fig. 1** shows a typical workflow for the application of a traditional handcrafted radiomic feature extraction approach followed by the use of an ML method in HN imaging. This process starts with image acquisition. Then radiomic features are mined during feature extraction. Next, available clinical or biological features (if any) can be added to the set of mined radiomics features. Because the size of the feature set often is larger than the sample size (ie, the number of medical images), feature selection or other dimensionality reduction approaches such as principal component analysis is used to extract a reduced set of features. These features are then used for model building. The performance of the model is assessed using previously unseen data; that is, data not used for training and fine-tuning the model.

Traditional ML methods rely on handcrafted features, which are extracted from images as inputs. These features are often the result of years of independent research but may be redundant and nonoptimal for a given task. Deep learning, which is a subset of ML, can be used to perform image analysis and has the potential to alleviate the need for handcrafted features. Given a large and representative dataset, deep learning methods can automatically extract discriminative features for a given task during model training. **Fig. 2** shows a typical deep learning workflow in radiomics. As discussed in later in this article, CNNs have been applied to various image processing tasks, including object detection, semantic segmentation, and instance segmentation.[16–20] CNN-based classification and segmentation models have been developed for HN imaging.[21,22] **Fig. 3** shows a typical CNN-based classification model. Unlike traditional ML approaches that involve a feature extraction step and use the extracted features for model building, feature extraction is part of the deep learning models. These models often are end-to-end models and are trained to learn the optimal features from data.

In a supervised learning context, a dataset of inputs and their desired outputs (ground truths) are used to build and fine-tune a model. Through an iterative process, a random batch of inputs is selected and fed to the model. The model output, which is a probability distribution for the classes, is then recorded. Next, the error (ie, the difference between the model output and the desired output)

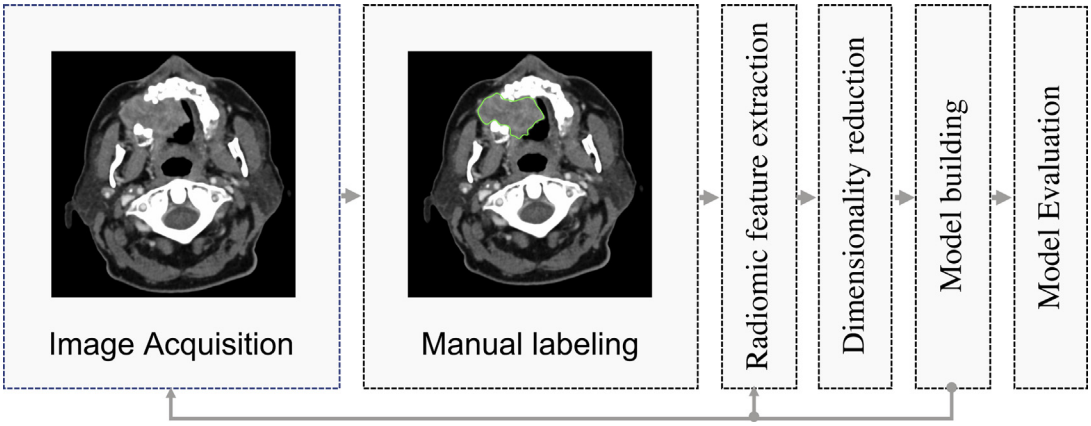

Fig. 1. A typical workflow for the application of traditional handcrafted features and ML methods in radiomics. After acquiring images, they are manually labeled or their region of interest is contoured, depending on the application. Then radiomic features are extracted, most commonly consisting of handcrafted features (variably referred to as texture or radiomic features in the medical literature). Next, dimensionality reduction methods are applied to get a reduced set of features for model building. After training and fine-tuning the model, a collection of sample data that have not been used for building the model is used to evaluate the model performance.

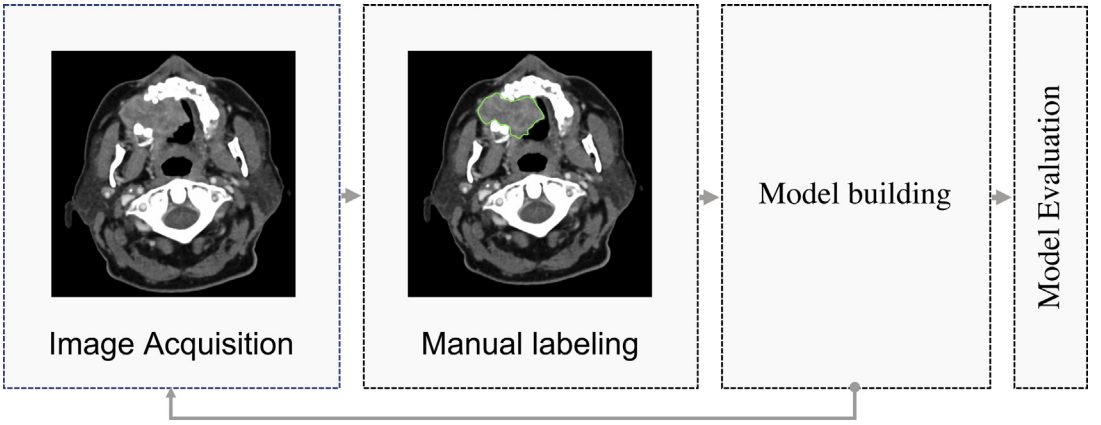

Fig. 2. A typical workflow for deep learning applications in radiomics. After acquiring images, they are manually labeled or contoured, depending on the application. After training and fine-tuning the model, a collection of samples that have not been used for building the model is used to evaluate the model performance. Although in this example the region of interest is manually contoured, with sufficiently large and varied datasets there is potential for deep learning to also perform whole-image analysis, including lesion identification and/or segmentation tasks in an automated or semiautomated fashion.

is measured through a loss function. After that, the model parameters are adjusted using the backpropagation algorithm to minimize the model error.[15] A typical CNN-based model used for image segmentation is shown in Fig. 4. This model is similar to the one for classification (see Fig. 2). The differences are caused by the nature of the desired output. In classification, an input image needs to be assigned to a category among a predetermined set of options. In contrast, in segmentation, the status of each pixel (voxel) of the input images needs to be determined; therefore, the input and output have the same dimension.

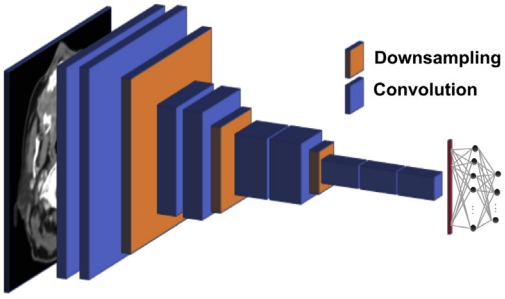

Fig. 3. A typical deep convolutional neural network used for a classification task. A sequence of operators is applied to an input image to produce a dense representation of the input image. This low-dimensional representation is referred to as deep features. These features are then fed to a multilayer perceptron to produce a probability distribution for categories/classes under study.

APPLICATION OF MACHINE LEARNING IN HEAD AND NECK IMAGING

There are several applications of ML that are of clinical interest for the HN. The most common applications include delineation of organ at risk (OAR) or primary tumor for radiation therapy; tumor detection; and tumor phenotyping and precision oncology applications such as prediction of histopathology or molecular phenotype, response to treatment, and survival. These applications often can be categorized as classification or segmentation tasks. Various applications of ML in HN imaging are discussed here. Note that this is not meant to be an exhaustive review of the vast literature or every radiomic or ML study in HN, but key examples from different areas of interest are provided. The studies consist of a combination of more traditional handcrafted radiomics and ML studies as well as studies using deep learning.

Autosegmentation of Organ at Risk

Delineation of OAR is one of the fundamental tasks in a radiation therapy workflow. Accurate delineation of a tumor plays an important role in optimizing and improving treatment outcomes.[23] In the manual delineation of an OAR, a radiation oncologist goes through hundreds of images and tries to delineate the OAR while adjusting various parameters. This task can also be delegated; for example, to a medical physicist, under the supervision of a radiation oncologist. This process is laborious and subject to intraobserver and interobserver variability and is one of the bottlenecks in a

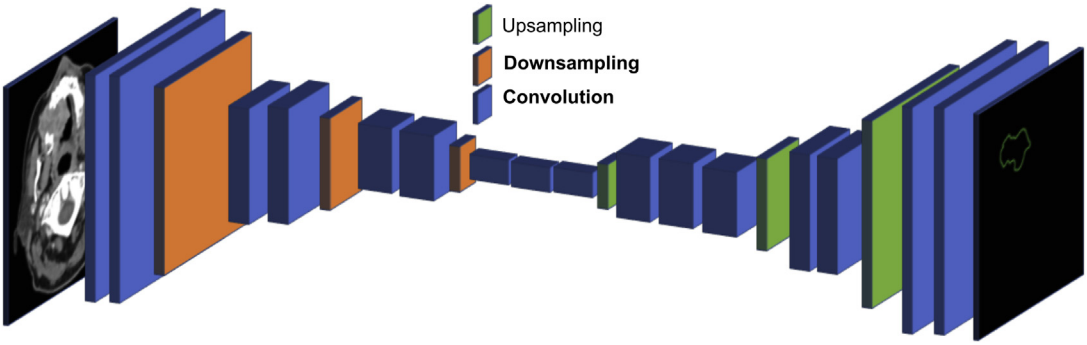

Fig. 4. A typical deep convolutional neural network used for an autosegmentation task. A sequence of operators, such as convolutions, upsampling, and downsampling, is applied to the input image to produce a segmentation map as the output.

radiotherapy workflow. ML has been used for tackling this task.[13,21,22,24–31] Atlas-based and deep learning approaches are 2 main categories of methods used for automatic delineation of OAR. Semiautomatic or automatic segmentation, if sufficiently robust, not only has the potential to improve workflow and productivity but also can reduce variations that could result in more consistent contouring and ultimately optimized treatments and outcomes.

Atlas-based methods for autosegmentation of organ at risk or primary tumor

Atlas-based methods for autosegmentation of OAR rely on an image registration method.[29–31] These methods use a library of contoured images, which are referred to as an atlas, to contour a new image. Through an image registration process, the pixels/voxels of a new image are aligned with pixels/voxels of an image from the atlas. Then, by translating the contour of the image in the library (a process called label fusion), a contour for the new image is generated.

Rigid and deformable registration are 2 main approaches for image registration.[32] Rigid registration uses 3 translational and 3 rotational degrees of freedom to overlay 2 images; therefore, all pixels/voxels from 1 image are uniformly transformed or rotated to overlay the other image. This approach often leads to poor results in the presence of patient movement or anatomic changes.[32] In contrast, deformable registration, generates a vector field transformation that maps pixels/voxels from 1 image to pixels/voxels in the other image.[33]

Yang and colleagues[24] proposed a method for autosegmentation of the parotid gland based on atlas registration and SVM. In a longitudinal study of 15 patients with HN cancer, they acquired MR images before radiation therapy and in 3, 6, and 12 months after starting radiation therapy. They used the pretreatment MR imaging and the binary contour for the parotid gland of each patient as an atlas. For each patient, the pretreatment MR imaging was mapped to the posttreatment MR imaging using a deformable image registration approach. Subsequently, the binary contour was transformed using the same mapping. They trained a radial basis function SVM on the pretreatment MR imaging and its transformed contours. The model was used to differentiate the voxels corresponding with the parotid gland from the surrounding tissues on the posttreatment MR images. The output of the autosegmentation system was evaluated against physicians' manual contours. The investigators reported success in parotid segmentation, with an average volume difference between the autosegmentation and the manual contours of 7.98% and 8.12% for the left and right parotid glands, respectively.

Deep learning methods for autosegmentation of organs at risk

Chen and colleagues[22] proposed a CNN architecture using a multitask learning approach aimed at learning multiple tasks at the same time. Then they transferred the learning weights to single-task learning. Their architecture consisted of 3 components: an encoder, a decoder, and a single-task or multitask layer. The encoder contained 4 downsampling layers, and the decoder contained 4 upsampling layers. The output of each encoder layer was concatenated to the input of the corresponding decoder layer. They used 5-by-5 dilated convolutions[34] and residual connections[16] between consecutive layers. Using a lifelong learning approach, they trained the autosegmentation system in 3 steps. First, they trained a single-task CNN for segmentation of the spinal cord. Then they substituted the last layer

of the network with a multitask layer and preserved the weights for the encoder and decoder layers. After training the multitask segmentation network, they substituted the multitask layer with a single-task layer, preserving the weights for the encoder and decoder layers. The final network was separately trained for the segmentation of 12 HN organs. Using a dataset of 200 HN CT scans, they reported that their proposed approach outperformed two-dimensional (2D) U-Net, three-dimensional (3D) U-Net, and a multitask learning model trained from scratch.

A multiorgan segmentation dataset might be imbalanced because of the variations in the number of annotated organs or differences in the size of organs. Zhu and colleagues[21] developed a 3D CNN model based on the U-Net architecture for the autosegmentation of HN anatomy considering the data imbalances. To address the data imbalance caused by differences in the organ sizes, they adjusted the 3D U-Net architecture. They removed all but 1 of the downsampling layers. They also embedded squeeze-and-excitation layers[35] in the architecture. Also, they proposed and used a hybrid loss function combining Dice coefficient loss and focal loss.[36] To address the data imbalance caused by missing annotations for some anatomic structures, they introduced a class weight into the loss function to assign a higher penalty for the error made in the segmentation of organs with less representation in the dataset. The missing annotations were masked out when calculating the loss function. Using 261 HN CT images from different sites and machines, they trained their model and reported a new and highly effective HN multiorgan segmentation algorithm capable of processing whole-volume CT images.[21]

Tong and colleagues[37] proposed an autosegmentation method for the automatic delineation of 9 HN OARs. The proposed model contains 2 main components: an autoencoder architecture as OARs shape representation model (SRM), and a fully convolutional neural network (FCNN) with a U-shaped architecture similar to that of U-Net.[20] The SRM model, an autoencoder, was trained using segmentation labels and was used as a regularizer when training the FCNN. They used a small dataset of CT images from 32 patients randomly split into a training (22 CT images) and a test set (10 CT images). They reported that their HN autosegmentation method outperformed an atlas-based autosegmentation model[38] that had been the winner of the Medical Image Computing and Computer Assisted Interventions 2015 Head and Neck Auto Segmentation Grand Challenge.[39]

Liang and colleagues[40] developed a deep learning–based method for the detection and segmentation of OARs using CT images from 185 patients with nasopharyngeal carcinoma. Their proposed method had 2 stages: detection and segmentation. The detection stage was based on Fast R-CNN architecture,[41] and the segmentation stage was based on a FCNN.[42] The CT images were used as a stack of 2D axial images. Each 2D image was resized and fed to the network as a 400-by-400 matrix. The investigators compared their proposed method with an FCNN model that directly segmented images without detecting the bounding boxes. They reported that the proposed method achieved significantly higher Dice coefficients compared with the FCNN model. The gain in performance using OAR detection as an intermediary step might be caused by the positions of OARs in CT images being almost consistent; therefore, Fast R-CNN can accurately detect bounding boxes for OARs. Consequently, by eliminating irrelevant information, conducting the segmentation might be an easier task for the model to accomplish.

Deep learning for primary tumor contouring and other advanced radiotherapy applications

In addition to delineation of OARs, studies are also showing the potential value of deep learning for more advanced applications, including automated contouring of primary tumors. In a study by Lin and colleagues,[43] a 3D CNN was trained for automatic primary gross tumor volume (GTV) contouring in patients with nasopharyngeal carcinoma. MR imaging scans from 1021 patients were divided into training, validation, and testing cohorts of 715, 103, and 203 patients, respectively, and GTV contours were defined by consensus of 2 experts. The output from the algorithm was then compared with 8 radiation oncologists in a multicenter evaluation. The AI-generated contours had a high level of accuracy compared with the ground truth (Dice similarity coefficient, 0.79; 2.0-mm difference in average surface distance). Furthermore, during the multicenter evaluation, the use of the AI tool improved contouring accuracy, reduced intraobserver and interobserver variation, and reduced contouring time. In another recent study, by Svecic and colleagues,[44] a deep learning algorithm trained on 337 patients and tested on 50 patients was used as a predictive framework for the evolution of tumor anatomy as well as interfractional dose delivery variations during external beam radiation therapy for HN cancer. Using sequential CT and associated dosimetry data, the probabilistic framework yielded a Dice score of 92% and an overall dose difference of 1.2 Gy in OARs and

tumor volume over the multiday treatment course, achieving a 5% reduction in delivered fraction segments.

Phenotype Classification for Head and Neck

The other common type of radiomic and ML studies are those concerned with classification tasks. A classification model is concerned with predicting the most probable category among 2 or several possible options. There are various tasks in HN disorder diagnosis and treatment workflow that can be modeled as a classification problem. Some samples of these tasks are discussed here.

Human papillomavirus prediction

HPV plays an important role in the pathophysiology of HNSCC, representing a distinct clinicopathologic entity and subtype that mainly affects oropharyngeal subsites. The prognosis for patients with HPV-negative HNSCC is poor compared with HPV-positive patients, and their clinical management can differ from their HPV-negative counterparts. HPV status is usually ascertained by using immunohistochemistry for its surrogate marker P16, a protein that hinders cell division, or via direct testing for HPV by in situ hybridization for viral DNA or polymerase chain reaction for HPV oncogene expression. These processes are invasive because they require collecting biospecimens from the patients. HN imaging and ML algorithms have been used for determining HPV status in a noninvasive manner.

Yu and colleagues[45] developed a model to assess HPV status using HN CT imaging. They extracted radiomic features using IBEX software.[46] After a comprehensive feature selection process, a general linear model was built to predict HPV status using the remaining radiomic features. The selected features indicated that HPV-positive tumors are smaller and more spherical compared with HPV-negative tumors. The investigators evaluated the model performance on a public and a private dataset, achieving area under the curve (AUC) values of 0.86667 and 0.91549 for the public and private datasets, respectively.

Buch and colleagues[47] also investigated HPV status prediction using texture features extracted from contrast-enhanced CT images of 40 patients with oropharyngeal squamous cell carcinoma. They identified histogram median, entropy, and GLCM (gray-level co-occurrence matrix) entropy as discriminative features and concluded that radiomics features could be used to predict the HPV status in squamous cell carcinoma. In another study, Vallieres and colleagues[48] investigated HPV status prediction using 18F-fluorodeoxyglucose PET images of 67 patients with HNSCC. The results of their SVM model supported the feasibility of HPV status prediction for HNSCC. Other studies have also reported variable accuracies for prediction of HNSCC HPV status.[49,50] It is not clear whether ML approaches can reach sufficient accuracy to replace HPV testing, especially given that this testing is done on biopsy specimens and is likely to continue for now. However, a potential role can be foreseen in underserved or under-resourced areas where such testing may not be easy to obtain. Furthermore, the ability to predict HPV status with a relatively high accuracy using noninvasive image analysis and ML suggests the potential value of this approach as a noninvasive biomarker for other tumor phenotypes of interest and precision oncology.

Lymph node classification

Lymph node metastasis and extracapsular lymph node extension are important prognostic factors for HN cancer. Furthermore, more reliable detection of early nodal metastases, especially for oral cavity HNSCCs, has the potential to reduce the need or required extent of neck (nodal) dissection for a subset of those cancers. Kann and colleagues[51] developed a CNN-based model, which they called DualNet, for pretreatment identification of nodal metastasis and extranodal extension using 2875 lymph nodes extracted from 124 patients. The model was then tested on an independent cohort of 131 patients. For the prediction of extranodal extension status, they achieved an AUC value of 0.91 with a negative predictive value of 0.95. For nodal metastasis prediction, they achieved an AUC value of 0.91 with a negative predictive value of 0.82.

Forghani and colleagues[13,52,53] developed an ML approach for the prediction of cervical lymph node metastasis using dual energy CT (DECT) images indirectly by evaluating the radiomic features of the primary HNSCC tumor. They used a retrospective study of 87 patients with histopathology-proven HNSCC. They built a random forest model for nodal status classification that achieved promising results. This study showed that radiomic features of the primary tumor have the potential to be used for nodal status prediction. Their experiments also indicated that multienergy CT scans are superior to single-energy CT scans in predicting cervical lymph node metastasis.

Seidler and colleagues[54] also built random forest and gradient boosting machine models for distinguishing metastatic HNSCC lymph nodes

from lymphoma, inflammatory, and normal nodes. They extracted 412 lymph nodes from DECT images of 50 patients. The built models were capable of distinguishing metastatic HNSCC from normal nodes, lymphoma from normal nodes, inflammatory from normal nodes, and malignant from benign lymph nodes. Their results indicated the utility of radiomic feature analysis of DECTs in the identification and histopathologic classification of cervical lymphadenopathy.

Chen and colleagues[55] also reported a hybrid model combining the handcrafted radiomic features and a 3D CNN for classifying lymph nodes as normal, suspicious, or involved. The investigators reported an accuracy of up to 88%. However, the patients did not undergo surgical neck dissection and therefore gold standard pathology proof was not available as ground truth. Studies have also shown the feasibility of evaluating metastatic lymph nodes from thyroid cancer on ultrasonography or CT.[56–58]

ML applications are promising for the evaluation of cervical lymphadenopathy and, if reliable and generalizable algorithms can be developed, have the potential for significant clinical impact. However, there are significant challenges for developing reliable algorithms for this purpose. From the standpoint of algorithm training and ground truth confirmation, it can be challenging to definitively cross-correlate all lymph nodes on imaging with histopathologically confirmed nodes from surgical specimens, especially the smaller nodes that are frequently of the greatest interest from the standpoint of clinical impact. This problem has the potential to introduce both errors and bias. Segmentation and analysis of numerous lymph nodes can also be prohibitively time and labor intensive and, furthermore, many smaller lymph nodes do not have sufficient volume for reliable radiomic analysis. In this regard, much work remains to be done, and a combined approach based on both direct nodal evaluation and indirect predictive modeling based on the analysis of the primary tumor may have the highest likelihood of successful outcome in terms of developing a reliable algorithm with the highest possible accuracy.

Prognosis, risk assessment, and treatment outcome prediction

In an early retrospective study of 72 patients using 2D CT texture analysis, texture features were found to be associated with overall survival in patients with locally advanced HNSCC treated with induction chemotherapy.[59] More recently, Zhang and colleagues[60] developed a model for pretreatment risk assessment of distant metastasis in patients with nasopharyngeal carcinoma using HN MR imaging images. They extracted 2780 radiomics features, among which 7 were selected for building a logistic regression model to classify patients to low risk or high risk of distant metastasis. They trained the model using a retrospective cohort of 123 untreated patients with nonmetastatic status (AUC = 0.827) and validated the trained model using an independent retrospective cohort of 53 patients (AUC = 0.792). The MR images used in this study were acquired from 2 different MR imaging machines and various protocols.

Early studies also suggest the use of MR imaging, CT, and/or PET radiomics and ML for predicting treatment failure or tumor recurrence following radiation and/or chemotherapy for different HN cancers.[49,61,62] Furthermore, pretreatment CT, PET, or MR imaging texture analysis has been used for the prediction of progression-free survival or overall survival in different mucosal HN cancers or thyroid cancer.[63–71]

Thyroid nodule classification

Thyroid cancer is the most prevalent malignancy in the endocrine system. Ultrasonography is commonly used as the first-line diagnostic modality for the detection and characterization of thyroid nodules. Both the acquisition and interpretation of ultrasonography are operator dependent and subjective. The rapid increase in ultrasonography usage has resulted in a significant increase in workload for radiologists, highlighting the need for automatic processing of the resulting data. Consequently, there has been interest in using ML for the detection and diagnosis of thyroid nodules.[72,73]

As an example, Park and colleagues[73] developed a deep learning–based method for the diagnosis of thyroid nodules. Their approach consisted of 3 components. The first component was a fully convolutional network[42] used for the segmentation of a lesion. Each lesion region was manually selected using a bounding box. After conducting segmentation, the tight bounding box for the lesion was selected and expanded by adding some extra margins. A patch of the image corresponding with the expanded bounding box was used as input for the second and third components. For the second component, AlexNet[74] was used to classify 7 ultrasonography features. For the third component, a modified version of GoogLeNet[75] was used to classify the lesion as benign or malignant. The output of the second component was also used as input to the classifier layer used in the third component. They compared the result of the deep learning–based method with

the results of an SVM classifier and the manual classifications made by a group of 10 radiologists. They observed that the deep learning–based classifier outperformed the SVM approach. Also, there were no significant differences between the results of the deep learning–based method and the classifications made by the expert radiologists. There are numerous publications exploring the potential of deep learning for thyroid nodule detection or classification, including nodule classification based on standardized risk stratification and classification systems such as the Thyroid Imaging Reporting and Data System.[76–82]

BARRIERS AND CHALLENGES FOR DEEP LEARNING APPLICATIONS IN HEAD AND NECK IMAGING

This article reviews various applications of ML in HN imaging. The research in this area shows promise; however, there are still significant challenges that need to be addressed in order to deploy traditional or deep learning methods in a clinical setting.

A challenge that hinders the clinical application of many ML models as a fully automated system is the imbalanced nature of datasets used for training these models. The accuracy of an ML model could be compromised for patients with uncommon tumor characteristics. Most reported research is considered as proof-of-concept models, because they use small sample sizes often from a single site. Conducting large-scale studies, where the phenotype under study is well represented in the training data, is the next step before the clinical deployment of these methods.

Another source of data imbalance in multiorgan studies is the large difference between the organ sizes. Small structures such as the optic chiasm only compose a small fraction of the whole medical image (0.000,01), which introduces data imbalance.[21] For example, it has been reported that the standard U-Net architecture has difficulty in segmenting organs with small volumes, such as the optic chiasm or optic nerves.[21] Generalized Dice coefficients,[83,84] Tversky,[85] focal loss,[36] sparsity label assignment deep multiinstance learning,[86] and exponential logarithm loss[87] are among the remedies used for dealing with data imbalances. Addressing the data imbalance remains an active area of research.

Deep learning and atlas-based methods are 2 main approaches used for the delineation of OAR. In atlas-based methods, the atlas must include a set of contoured template images that represent the population under study. A deviation from the target population may lead to poor results for patients with anatomic characteristics that are not well represented in the atlas. Deep learning methods have led to promising results in OAR delineation, which often is the bottleneck in the applications of radiomics. Alleviating this bottleneck will be an important step in providing adaptive treatment planning and personalized radiation therapy.

The output of autosegmentation tools might degrade because of differences in protocols and machinery used for acquiring medical images.[88] Because the error in the delineation of OARs may affect the radiotherapy treatment, a systematic evaluation of available tools through a pilot study is required before putting them into practice. In such a pilot study, a representative subset of the dataset under study should be contoured by experts and be used to evaluate the performance of each tool. Alternatively, autosegmentation can be used to generate an initial segmentation that is reviewed and corrected by an expert. This semiautomated process helps to speed up the contouring process and avoid the error introduced by the autocontouring step. This process also has the advantage of having an expert human as backup, which may also facilitate adoption.

The comparison of various ML applications in HN imaging remains challenging in the lack of large-scale benchmark datasets. This difficulty is more of an issue when evaluating deep learning models, because there are various hyperparameters that may affect the performance of these models. Also, fine-tuning these models is computationally demanding. For example, Chen and colleagues[22] reported that their architecture outperformed the 3D U-Net in the soft Dice and root mean squared error, although the comparison results might not hold when using larger datasets.

In the absence of large-scale medical imaging datasets, transfer learning can be a viable option for achieving performant and generalizable models. For example, the model proposed by Chen and colleagues[22] that used multitask learning and transfer learning outperformed a model with the same architecture trained from scratch.

Annotating medical images requires expertise that is rare and expensive, which is an obstacle for developing large-scale benchmark datasets. To speed up the annotation process, often, each medical image is annotated by a single practitioner. One limitation in such experiments is that the trained model using these datasets might learn the systematic bias or error in the contours. Ideally, the contours should be generated independently by a group of physicians. For each contour, the image and the contour can be used as a

training example. This approach is suitable for deep learning applications and can help with alleviating systematic errors in contours generated by a single practitioner. This process can be considered as data augmentation applied to contours. Alternatively, when several physicians have contoured an image, the region encompassing the consensus can be used as the contour for that image.

Assembling large-scale HN imaging datasets is more challenging considering the ethical and legal issues pertaining to patient privacy but could yield superior results. A viable approach for addressing this challenge would be multiorganization collaboration through data sharing platforms with a high level of data security, along with robust anonymization tools for removal of all metadata, possibly supplemented by defacing or equivalent software given the small possibility of surface rendering and facial recognition software for patient identification. Other approaches where the primary data or scan do not leave the institution, such as federated learning, can also be used to enhance privacy protection.

In some applications of deep learning, the lack of sensitivity has been reported.[89] Achieving a sufficiently high sensitivity for the application of interest without sacrificing specificity is an essential requirement for any system with the potential for clinical deployment. A more detailed and general discussion of basic challenges for ML algorithm development, as well as handcrafted radiomics, is beyond the scope of this article but can found in various review articles on the topic.[6,11,12]

SUMMARY

ML has shown the potential to use currently unused data available on patients' medical imaging scans for building predictive models that lead to improved diagnosis and ultimately outcome for patients with different HN disorders, especially HN cancers. Most experiments so far have been proof of concepts and conducted using a small number of samples acquired from a single or a few sites following a single protocol. To be able to deploy these models in a clinical setting, conducting large-scale studies that encompass various sites and protocols, is a requirement. To do so requires developing data aggregation and data sharing pipelines that, while addressing patient privacy concerns, make developing such models possible. If these obstacles can be overcome, ML-based assistive diagnostic tools have the potential to improve both the efficiency and quality of health care, paving the way for more precise and personalized therapies.

REFERENCES

1. Forghani R, Johnson JM, Ginsberg LE. Imaging of Head and Neck Cancer. In: Myers J, Hanna E, Myers EN, editors. Cancer of the head and neck. 5th edition. Philadelphia: Wolters Kluwer; 2017. p. 92–148.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA A Cancer J Clin 2019;69:7-34. doi:10.3322/caac.21551.
3. Edge SB, Byrd DR, Compton CC, et al. AJCC cancer staging manual. 7th edition. New York: Springer; 2010.
4. Amin MB, Edge S, Greene F, et al. AJCC cancer staging manual. 8th edition. New York: Springer International Publishing; 2017.
5. Chang Z. Will AI improve tumor delineation accuracy for radiation therapy? Radiology 2019;291(3):687–8.
6. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology 2015;278(2):563–77.
7. Sroussi HY, Epstein JB, Bensadoun R-J, et al. Common oral complications of head and neck cancer radiation therapy: mucositis, infections, saliva change, fibrosis, sensory dysfunctions, dental caries, periodontal disease, and osteoradionecrosis. Cancer Med 2017;6(12):2918–31.
8. Kumar V, Gu Y, Basu S, et al. Radiomics: the process and the challenges. Magn Reson Imaging 2012;30(9):1234–48.
9. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer 2012;48(4):441–6.
10. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014;5:4006.
11. Forghani R, Savadjiev P, Chatterjee A, et al. Radiomics and artificial intelligence for biomarker and prediction model development in oncology. Comput Struct Biotechnol J 2019;17:995–1008.
12. Forghani R. Precision digital oncology: Emerging role of radiomics-based biomarkers and artificial intelligence for advanced imaging and characterization of brain tumors. Radiology: Imaging Cancer 2020;2(4):e190047.
13. Forghani R, Chatterjee A, Reinhold C, et al. Head and neck squamous cell carcinoma: prediction of cervical lymph node metastasis by dual-energy CT texture analysis with machine learning. Eur Radiol 2019;29(11):6172–81.
14. Al Ajmi E, Forghani B, Reinhold C, et al. Spectral multi-energy CT texture analysis with machine learning for tissue classification: an investigation using classification of benign parotid tumours as a testing paradigm. Eur Radiol 2018;28(6):2604–11.

15. Goodfellow I, Bengio Y, Courville A. Deep learning. Cambridge (MA): MIT Press; 2018.
16. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. Paper presented at: Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas (NV), June 26-July 1, 2016.
17. Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks. Paper presented at: Proceedings of the IEEE conference on computer vision and pattern recognition. Honolulu (HI), July 22-25, 2017.
18. Pinheiro PO, Lin T-Y, Collobert R, et al. Learning to refine object segments. Paper presented at: European Conference on Computer Vision. Las Vegas (NV), June 26-July 1, 2016.
19. Pinheiro POO, Collobert R, Dollar P. Learning to segment object candidates. Paper presented at: Advances in Neural Information Processing Systems. Montreal (QC), December 7-10, 2015.
20. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. Paper presented at: International Conference on Medical image computing and computer-assisted intervention. Munich (Germany), October 5-9, 2015.
21. Zhu W, Huang Y, Zeng L, et al. AnatomyNet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. Med Phys 2019;46(2):576–89.
22. Chan JW, Kearney V, Haaf S, et al. A convolutional neural network algorithm for automatic segmentation of head and neck organs at risk using deep lifelong learning. Med Phys 2019;46(5):2204–13.
23. Ng WT, Lee MCH, Chang ATY, et al. The impact of dosimetric inadequacy on treatment outcome of nasopharyngeal carcinoma with IMRT. Oral Oncol 2014;50(5):506–12.
24. Yang X, Wu N, Cheng G, et al. Automated segmentation of the parotid gland based on atlas registration and machine learning: a longitudinal MRI study in head-and-neck radiation therapy. Int J Radiat Oncol Biol Phys 2014;90(5):1225–33.
25. Wu VWC, Ying MTC, Kwong DLW. Evaluation of radiation-induced changes to parotid glands following conventional radiotherapy in patients with nasopharygneal carcinoma. Br J Radiol 2011;84(1005):843–9.
26. Guidi G, Maffei N, Vecchi C, et al. A support vector machine tool for adaptive tomotherapy treatments: prediction of head and neck patients criticalities. Physica Med 2015;31(5):442–51.
27. Scalco E, Fiorino C, Cattaneo GM, et al. Texture analysis for the assessment of structural changes in parotid glands induced by radiotherapy. Radiother Oncol 2013;109(3):384–7.
28. Eisbruch A, Kim HM, Terrell JE, et al. Xerostomia and its predictors following parotid-sparing irradiation of head-and-neck cancer. Int J Radiat Oncol Biol Phys 2001;50(3):695–704.
29. Kearney V, Chen S, Gu X, et al. Automated landmark-guided deformable image registration. Phys Med Biol 2014;60(1):101.
30. Kearney V, Huang Y, Mao W, et al. Canny edge-based deformable image registration. Phys Med Biol 2017;62(3):966–85.
31. Obeidat M, Narayanasamy G, Cline K, et al. Comparison of different QA methods for deformable image registration to the known errors for prostate and head-and-neck virtual phantoms. Biomed Phys Eng Express 2016;2(6):067002.
32. Fortin D, Basran PS, Berrang T, et al. Deformable versus rigid registration of PET/CT images for radiation treatment planning of head and neck and lung cancer patients: a retrospective dosimetric comparison. Radiat Oncol 2014;9(1):50.
33. Oh S, Kim S. Deformable image registration in radiation therapy. Radiat Oncol J 2017;35(2):101.
34. Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv: 151107122. 2015.
35. Hu J, Shen L, Sun G. Squeeze-and-Excitation Networks. Paper presented at: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City (UT), June 18-22, 2018.
36. Lin T-Y, Goyal P, Girshick R, et al. Focal loss for dense object detection. Paper presented at: The IEEE International Conference on Computer Vision. Venice (Italy), October 22-29, 2017.
37. Tong N, Gou S, Yang S, et al. Fully automatic multi-organ segmentation for head and neck cancer radiotherapy using shape representation model constrained fully convolutional neural networks. Med Phys 2018;45(10):4558–67.
38. Han X, Hoogeman MS, Levendag PC, et al. Atlas-Based Auto-segmentation of Head and Neck CT Images. Paper presented at: Medical Image Computing and Computer-Assisted Intervention–MICCAI. New York, September 6-10, 2008.
39. Mannion-Haworth R, Bowes M, Ashman A, et al. Fully automatic segmentation of head and neck organs using active appearance models. MIDAS J 2016.
40. Liang S, Tang F, Huang X, et al. Deep-learning-based detection and segmentation of organs at risk in nasopharyngeal carcinoma computed tomographic images for radiotherapy planning. Eur Radiol 2019;29(4):1961–7.
41. Ren S, He K, Girshick R, et al. Towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell 2016;39(6):1137–49.
42. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Paper

presented at: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston (MA), June 7-12, 2015.
43. Lin L, Dou Q, Jin YM, et al. Deep learning for automated contouring of primary tumor volumes by MRI for nasopharyngeal carcinoma. Radiology 2019;291(3):677–86.
44. Svecic A, Roberge D, Kadoury S. Prediction of interfractional radiotherapy dose plans with domain translation in spatiotemporal embeddings. Med Image Anal 2020;64:101728.
45. Yu K, Zhang Y, Yu Y, et al. Radiomic analysis in prediction of Human Papilloma Virus status. Clin Transl Radiat Oncol 2017;7:49–54.
46. Zhang L, Fried DV, Fave XJ, et al. IBEX: an open infrastructure software platform to facilitate collaborative work in radiomics. Med Phys 2015;42(3):1341–53.
47. Buch K, Fujita A, Li B, et al. Using texture analysis to determine human papillomavirus status of oropharyngeal squamous cell carcinomas on CT. AJNR Am J Neuroradiol 2015;36(7):1343–8.
48. Vallieres M, Kumar A, Sultanem K, et al. FDG-PET image-derived features can determine HPV status in head-and-neck cancer. Int J Radiat Oncol Biol Phys 2013;87(2):S467.
49. Bogowicz M, Riesterer O, Ikenberg K, et al. Computed Tomography Radiomics Predicts HPV status and local tumor control after definitive radiochemotherapy in head and neck squamous cell carcinoma. Int J Radiat Oncol Biol Phys 2017;99(4):921–8.
50. Haider SP, Mahajan A, Zeevi T, et al. PET/CT radiomics signature of human papilloma virus association in oropharyngeal squamous cell carcinoma. Eur J Nucl Med Mol Imaging 2020. https://doi.org/10.1007/s00259-020-04839-2.
51. Kann BH, Aneja S, Loganadane GV, et al. Pretreatment identification of head and neck cancer nodal metastasis and extranodal extension using deep learning neural networks. Sci Rep 2018;8(1):14036.
52. Forghani R, De Man B, Gupta R. Dual-energy computed tomography: physical principles, approaches to scanning, usage, and implementation: part 1. Neuroimaging Clin 2017;27(3):371–84.
53. Forghani R, De Man B, Gupta R. Dual-energy computed tomography: physical principles, approaches to scanning, usage, and implementation: part 2. Neuroimaging Clin 2017;27(3):385–400.
54. Seidler M, Forghani B, Reinhold C, et al. Dual-energy CT texture analysis with machine learning for the evaluation and characterization of cervical lymphadenopathy. Comput Struct Biotechnol J 2019;17:1009–15.
55. Chen L, Zhou Z, Sher D, et al. Combining many-objective radiomics and 3D convolutional neural network through evidential reasoning to predict lymph node metastasis in head and neck cancer. Phys Med Biol 2019;64(7):075011.
56. Lee JH, Baek JH, Kim JH, et al. Deep learning-based computer-aided diagnosis system for localization and diagnosis of metastatic lymph nodes on ultrasound: a pilot study. Thyroid 2018;28(10):1332–8.
57. Lee JH, Ha EJ, Kim JH. Application of deep learning to the diagnosis of cervical lymph node metastasis from thyroid cancer with CT. Eur Radiol 2019;29(10):5452–7.
58. Lee JH, Ha EJ, Kim D, et al. Application of deep learning to the diagnosis of cervical lymph node metastasis from thyroid cancer with CT: external validation and clinical utility for resident training. Eur Radiol 2020;30(6):3066–72.
59. Zhang H, Graham CM, Elci O, et al. Locally advanced squamous cell carcinoma of the head and neck: CT texture and histogram analysis allow independent prediction of overall survival in patients treated with induction chemotherapy. Radiology 2013;269(3):801–9.
60. Zhang L, Dong D, Li H, et al. Development and validation of a magnetic resonance imaging-based model for the prediction of distant metastasis before initial treatment of nasopharyngeal carcinoma: A retrospective cohort study. EBioMedicine 2019;40:327–35.
61. Li S, Wang K, Hou Z, et al. Use of radiomics combined with machine learning method in the recurrence patterns after intensity-modulated radiotherapy for nasopharyngeal carcinoma: a preliminary study. Front Oncol 2018;8:648.
62. Kuno H, Qureshi MM, Chapman MN, et al. CT texture analysis potentially predicts local failure in head and neck squamous cell carcinoma treated with chemoradiotherapy. AJNR Am J Neuroradiol 2017;38(12):2334–40.
63. Mao J, Fang J, Duan X, et al. Predictive value of pretreatment MRI texture analysis in patients with primary nasopharyngeal carcinoma. Eur Radiol 2019;29(8):4105–13.
64. Cheng NM, Fang YH, Chang JT, et al. Textural features of pretreatment 18F-FDG PET/CT images: prognostic significance in patients with advanced T-stage oropharyngeal squamous cell carcinoma. J Nucl Med 2013;54(10):1703–9.
65. Leijenaar RT, Carvalho S, Hoebers FJ, et al. External validation of a prognostic CT-based radiomic signature in oropharyngeal squamous cell carcinoma. Acta Oncol 2015;54(9):1423–9.
66. Parmar C, Grossmann P, Rietveld D, et al. Radiomic machine-learning classifiers for prognostic biomarkers of head and neck cancer. Front Oncol 2015;5:272.
67. Vallieres M, Kay-Rivest E, Perrin LJ, et al. Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. Sci Rep 2017;7(1):10117.

68. Park VY, Han K, Lee E, et al. Association between radiomics signature and disease-free survival in conventional papillary thyroid carcinoma. Sci Rep 2019;9(1):4501.
69. Zdilar L, Vock DM, Marai GE, et al. Evaluating the effect of right-censored end point transformation for radiomic feature selection of data from patients with oropharyngeal cancer. JCO Clin Cancer Inform 2018;2:1–19.
70. Zhuo EH, Zhang WJ, Li HJ, et al. Radiomics on multimodalities MR sequences can subtype patients with non-metastatic nasopharyngeal carcinoma (NPC) into distinct survival subgroups. Eur Radiol 2019; 29(10):5590–9.
71. Haider SP, Zeevi T, Baumeister P, et al. Potential Added Value of PET/CT Radiomics for Survival Prognostication beyond AJCC 8th edition staging in oropharyngeal squamous cell carcinoma. Cancers (Basel) 2020;12(7):E1778.
72. Acharya UR, Swapna G, Sree SV, et al. A review on ultrasound-based thyroid cancer tissue characterization and automated classification. Technol Cancer Res Treat 2014;13(4):289–301.
73. Park VY, Han K, Seong YK, et al. Diagnosis of thyroid nodules: performance of a Deep Learning convolutional neural network Model vs. Radiologists. Sci Rep 2019;9(1):1–9.
74. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Paper presented at: Advances in Neural Information Processing Systems. Stateline (NV), December 3-8, 2012.
75. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. Paper presented at: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition2015.
76. Chi J, Walia E, Babyn P, et al. Thyroid nodule classification in ultrasound images by fine-tuning deep convolutional neural network. J Digit Imaging 2017; 30(4):477–86.
77. Li H, Weng J, Shi Y, et al. An improved deep learning approach for detection of thyroid papillary cancer in ultrasound images. Sci Rep 2018;8(1):6600.
78. Song W, Li S, Liu J, et al. Multitask cascade convolution neural networks for automatic thyroid nodule detection and recognition. IEEE J Biomed Health Inform 2019;23(3):1215–24.
79. Ko SY, Lee JH, Yoon JH, et al. Deep convolutional neural network for the diagnosis of thyroid nodules on ultrasound. Head Neck 2019;41(4):885–91.
80. Buda M, Wildman-Tobriner B, Hoang JK, et al. Management of thyroid nodules seen on US images: deep learning may match performance of radiologists. Radiology 2019;292(3):695–701.
81. Akkus Z, Cai J, Boonrod A, et al. A survey of deep-learning applications in ultrasound: artificial intelligence-powered ultrasound for improving clinical workflow. J Am Coll Radiol 2019;16(9 Pt B): 1318–28.
82. Sun C, Zhang Y, Chang Q, et al. Evaluation of a deep learning-based computer-aided diagnosis system for distinguishing benign from malignant thyroid nodules in ultrasound images. Med Phys 2020. https://doi.org/10.1002/mp.14301.
83. Sudre CH, Li W, Vercauteren T, et al. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: Cardoso J, Arbel T, Carneiro G, et al, editors. Deep learning in medical image analysis and multimodal learning for clinical decision support. Cham (Switzerland): Springer; 2017. p. 240–8.
84. Crum WR, Camara O, Hill DLG. Generalized overlap measures for evaluation and validation in medical image analysis. IEEE Trans Med Imaging 2006; 25(11):1451–61.
85. Salehi SSM, Erdogmus D, Gholipour A. Tversky loss function for image segmentation using 3D fully convolutional deep networks. Paper presented at: International Workshop on Machine Learning in Medical Imaging2017.
86. Zhu W, Lou Q, Vang YS, et al. Deep multi-instance networks with sparse label assignment for whole mammogram classification. Paper presented at: International Conference on Medical Image Computing and Computer-Assisted Intervention2017.
87. Wong KCL, Moradi M, Tang H, et al. 3D segmentation with exponential logarithmic loss for highly unbalanced object sizes. Paper presented at: International Conference on Medical Image Computing and Computer-Assisted Intervention2018.
88. Lee H, Lee E, Kim N, et al. Clinical evaluation of commercial atlas-based auto-segmentation in the head and neck region. Front Oncol 2019;9:239.
89. Wang H, Zhou Z, Li Y, et al. Comparison of machine learning methods for classifying mediastinal lymph node metastasis of non-small cell lung cancer from 18F-FDG PET/CT images. EJNMMI Res 2017;7(1):11.