European Radiology (2019) 29:5452–5457
https://doi.org/10.1007/s00330-019-06098-8

IMAGING INFORMATICS AND ARTIFICIAL INTELLIGENCE

Application of deep learning to the diagnosis of cervical lymph node metastasis from thyroid cancer with CT

Jeong Hoon Lee[1] · Eun Ju Ha[2] · Ju Han Kim[1]

Received: 3 January 2019 / Revised: 1 February 2019 / Accepted: 11 February 2019 / Published online: 15 March 2019
© European Society of Radiology 2019

Abstract

Purpose To develop a deep learning–based computer-aided diagnosis (CAD) system for use in the CT diagnosis of cervical lymph node metastasis (LNM) in patients with thyroid cancer.

Methods A total of 995 axial CT images that included benign ($n = 647$) and malignant ($n = 348$) lymph nodes were collected from 202 patients with thyroid cancer who underwent CT for surgical planning between July 2017 and January 2018. The datasets were randomly split into training (79.0%), validation (10.5%), and test (10.5%) datasets. Eight deep convolutional neural network (CNN) models were used to classify the images into metastatic or benign lymph nodes. Pretrained networks were used on the ImageNet and the best-performing algorithm was selected. Class-specific discriminative regions were visualized with attention heatmap using a global average pooling method.

Results The area under the ROC curve (AUROC) for the tested algorithms ranged from 0.909 to 0.953. The sensitivity, specificity, and accuracy of the best-performing algorithm were all 90.4%, respectively. Attention heatmap highlighted important subregions for further clinical review.

Conclusion A deep learning–based CAD system could accurately classify cervical LNM in patients with thyroid cancer on preoperative CT with an AUROC of 0.953. Whether this approach has clinical utility will require evaluation in a clinical setting.

Key Points

• *A deep learning–based CAD system could accurately classify cervical lymph node metastasis. The AUROC for the eight tested algorithms ranged from 0.909 to 0.953.*

• *Of the eight models, the ResNet50 algorithm was the best-performing model for the validation dataset with 0.953 AUROC. The sensitivity, specificity, and accuracy of the ResNet50 model were all 90.4%, respectively, in the test dataset.*

• *Based on its high accuracy of 90.4%, we consider that this model may be useful in a clinical setting to detect LNM on preoperative CT in patients with thyroid cancer.*

Keywords Artificial intelligence · Lymphatic metastasis · Thyroid cancer · Multidetector computed tomography

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00330-019-06098-8) contains supplementary material, which is available to authorized users.

✉ Eun Ju Ha
radhej@naver.com

[1] Division of Biomedical Informatics, Seoul National University Biomedical Informatics (SNUBI), Seoul National University College of Medicine, Seoul 110799, Republic of Korea

[2] Department of Radiology, Ajou University School of Medicine, Wonchon-Dong, Yeongtong-Gu, Suwon 443-380, South Korea

Abbreviations

AUROC	Area under the receiver operating characteristic curve
CAD	Computer-aided diagnosis
CAM	Class activation mapping
CNN	Convolutional neural network
FNA	Fine needle aspiration
LNM	Lymph node metastasis
PTC	Papillary thyroid carcinoma
US	Ultrasonography

Introduction

The incidence of thyroid cancer has increased substantially over recent decades [1]. Because of this, accurate preoperative staging has become a major issue for surgical planning [2, 3]. Although differentiated thyroid cancers have a good prognosis and low mortality rate, cervical lymph node metastasis (LNM) has been reported in up to 60–70% of patients and is considered an important risk factor for locoregional recurrence [4–8]. Therefore, current guidelines recommend preoperative computed tomography (CT) [9] in addition to ultrasonography (US) for patients with a clinical suspicion of multiple or bulky lymph node involvement [3, 8, 10].

However, previous studies have reported inconsistent results of CT, with a variable sensitivity of 35.0–77.0% and specificity of 70.0–96.0% for the diagnosis of cervical LNM in patients with thyroid cancer [11, 12]. Therefore, confusion still exists with respect to determining the surgical extent when prominent cervical lymph nodes are encountered on CT. Although prophylactic lateral neck dissection may increase the chance of removing such tumors, it increases the risk of surgical complications and decreases the quality of life after surgery [3, 10]. Therefore, the extent of lymph node dissection required should be more accurately determined before the initial surgery.

Deep learning techniques have recently been introduced in medical imaging for accurate and consistent interpretation of imaging features [13–15]. In thyroid imaging, a few studies have reported promising results of computer-aided diagnosis (CAD) systems for characterizing thyroid nodules on US [16–18]. A recent study by Lee et al. reported the possibility of using a deep learning–based CAD system for the classification of metastatic lymph nodes on US [19]. However, to our knowledge, no deep learning model for the classification of cervical LNM on CT images has been developed for patients with thyroid cancers. Therefore, this study aimed to develop a deep learning–based model to differentiate benign from metastatic lymph nodes based on preoperative CT images, and to validate the diagnostic performance of this system.

Materials and methods

Patients and datasets

The protocol of this retrospective study was approved by the Ethics Committee of the Institutional Review Board at the Ajou University Medical Center. A total of 995 axial CT images (including 647 benign and 348 metastatic lymph nodes) were retrieved from 202 patients with thyroid cancer who underwent CT for surgical planning between July 2017 and January 2018. All lymph node characterizations were confirmed by fine-needle aspiration and/or surgery. The 995 axial contrast-enhanced CT images were randomly split into training (543 benign and 244 metastatic lymph nodes), validation (52 benign and 52 metastatic lymph nodes), and test (52 benign and 52 metastatic lymph nodes) datasets.

All CT images were obtained using 64- to 128-channel multidetector CT scanners (SOMATOM Definition Flash, Siemens Medical Solutions; Brilliance, Philips Medical Systems). Contrast-enhanced CT scanning was performed 40 s after initiation of an intravenous injection of a 90 mL bolus of iodinated nonionic contrast material (300–350 mgI/mL) into the right arm, with a subsequent injection of 20–30 mL normal saline for flushing at 3 mL/s using an automated injector. Finally, CT images were obtained with 0.5–0.6 mm collimation, and were reconstructed into axial images every 2.0 mm on a 512 × 512 matrix using iterative reconstruction algorithms available with each vendor's CT scanner.

Image preprocessing, training details, and performance evaluation

A radiologist used the CT images to manually identify lesions on the cervical lymph node, and then drew a rectangular mask around them. To avoid erroneous learning of noises such as bones and blood vessels, only the masked regions were used as input data for deep convolutional neural network (CNN) models. Since the training data had a larger number of benign lymph nodes than metastatic lymph nodes, metastatic lymph nodes in the training data were reoriented in order to create pseudo-new data [9], and images were then input into the deep CNN model.

Eight deep learning algorithms were used to select the most suitable model to differentiate benign from metastatic lymph nodes based on preoperative CT images. The CNN models, such as VGG16, VGG19, Xception, InceptionV3, InceptionResNetV2, DenseNet121, DenseNet169, and ResNet, were used as a backbone model pretrained on the ImageNet database [20–26]. In this study, generated activation maps by CAM on test dataset were applied to evaluate the region of interest for further clinical review. A simple workflow scheme for this process is shown in Fig. 1. A more technical and detailed description of our method is shown in the online supplementary.

The eight deep learning algorithms were trained in the same training sample. All training sessions were carried out using deep learning package Keras (http://keras.io/) with the tensorflow machine learning back end [27]. The batch size was set to 30 and the learning rate to 0.005, decreasing by a factor of 10 when no further decrease was observed in the validation dataset. We have selected the models with the best accuracy in the validation set for each algorithm. The performances of the eight algorithms were measured by the area under the receiver operating characteristic curve (AUROC) of the validation data set, and the sensitivity, specificity, positive predictive value (PPV), and

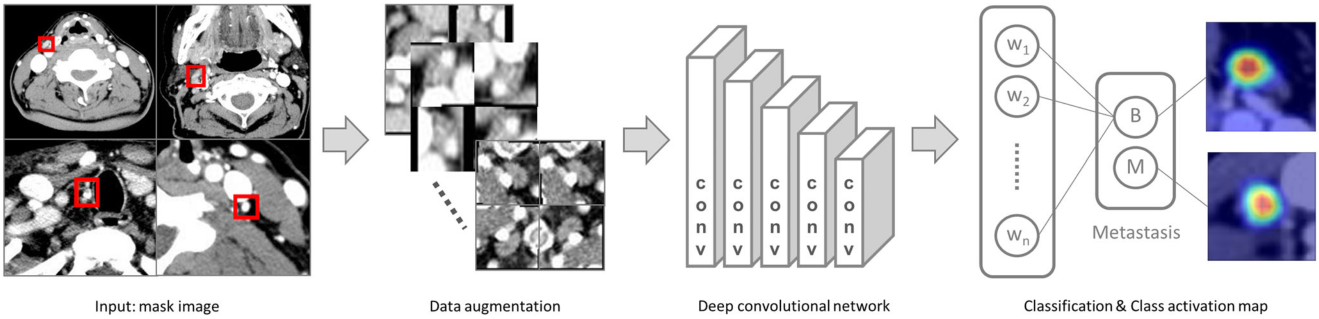

Fig. 1 Workflow scheme of the data-learning process for computed tomography [9] lymph node images. Flow diagrams show the process from data acquisition to lymph node metastasis prediction and generation of an attention heatmap

negative predictive value (NPV), and accuracy of the test dataset. The receiver operating characteristic curve figure and outcome analyses were performed using the pROC package of R statistical software (https://cran.r-project.org/web/packages/pROC).

Results

Diagnostic performance

Table 1 and Fig. 2 show the results of the eight tested algorithms. The AUROC for the tested algorithms ranged from 0.909 to 0.953. Of the eight models, the ResNet50 and InceptionV3 algorithms were the best-performing models for the validation dataset with 0.953 and 0.952 AUROC, respectively (Table 1). A threshold value of 0.869 and 0.899 for the highest sensitivity and specificity pair was determined by obtaining Youden's J statistic. Based on this threshold, the sensitivity, specificity, and accuracy of the model were measured for the diagnosis of metastatic lymph nodes on CT images in the test data set. The sensitivity, specificity, and accuracy of the ResNet50 model were all 90.4%, respectively, in the test dataset. The sensitivity, specificity, and accuracy of the InceptionV3 model were 88.5% 90.4%, and 89.4%, respectively. The PPV and NPV were 90.4% and 90.4%, respectively, in ResNet50 model and 90.4% and 89.4%, respectively, in InceptionV3 model. The performance of the test set for all models was not significantly correlated with the parameters and depth ($p > 0.05$).

Attention heatmap and lesion detection

The attention heatmap of the benign and metastatic lymph nodes was generated by CAM and then superimposed on the original CT image so that the location of the actual lymph node and the region highlighted by the model could be compared. As shown in Fig. 3, the attention heatmap highlighted important subregions for further clinical review. This showed that the abnormal characteristics of a malignancy had been learned by the CNN and used as the basis for its classification of LNM.

Discussion

The findings of our study show that a deep learning–based CAD system could accurately classify cervical lymph node metastasis with 0.953 AUROC in the validation set. Based

Table 1 Diagnostic performance of the deep learning algorithms for metastatic lymph nodes on CT images in the validation and test datasets

Algorithm	Validation	Test					Parameters	Depth
	AUROC	Accuracy	Sensitivity	Specificity	PPV	NPV		
ResNet50	0.953	0.904	0.904	0.904	0.904	0.904	25,636,712	168
InceptionV3	0.952	0.894	0.885	0.904	0.902	0.887	23,851,784	159
DenseNet121	0.943	0.837	0.692	0.981	0.973	0.761	8,062,504	121
Xception	0.942	0.894	0.865	0.923	0.918	0.872	22,910,480	126
InceptionResNetV2	0.936	0.817	0.692	0.942	0.923	0.754	55,873,736	572
DenseNet169	0.920	0.894	0.808	0.981	0.977	0.836	14,307,880	169
VGG19	0.918	0.894	0.846	0.942	0.936	0.859	143,667,240	26
VGG16	0.909	0.846	0.865	0.827	0.833	0.860	138,357,544	23

AUROC area under the receiver operating characteristic curve, *PPV* positive predictive value, *NPV* negative predictive value

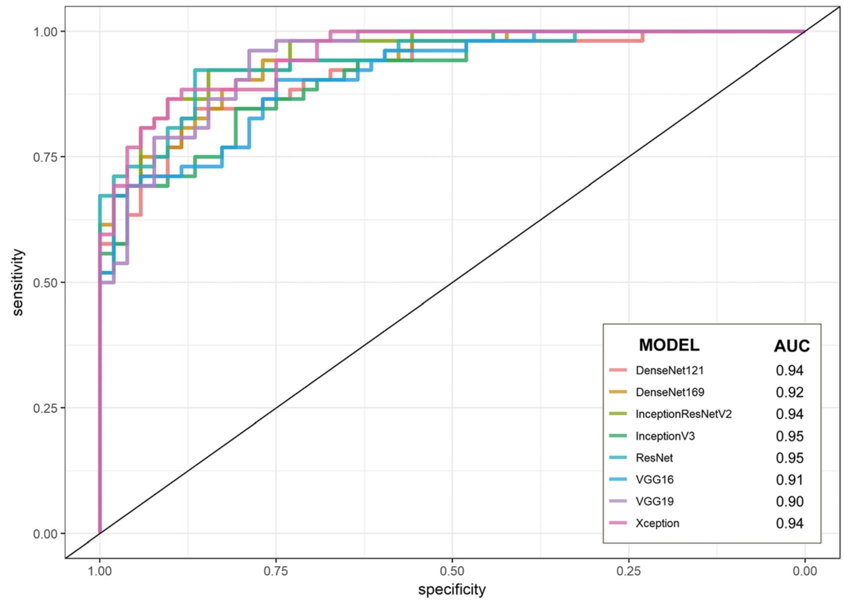

Fig. 2 Receiver operating characteristic (ROC) curves of the eight tested algorithms

on its high accuracy of 90.4%, we consider that this model may be useful in a clinical setting to detect LNM on preoperative CT in patients with thyroid cancer.

Neck US is primarily recommended for the evaluation of primary mass and cervical LNM in patients with thyroid cancer [3, 10]. However, although US is highly accurate in the diagnosis of thyroid cancer, it has not shown sufficient accuracy for the diagnosis of LNM in previous studies [12, 28–30]. Therefore, preoperative CT is recommended in cases where there is clinical suspicion of advanced thyroid cancer, as an adjunct to US [3, 8, 10]. Some surgeons also prefer CT as it can deliver detailed anatomic information regarding nodal location in relation to anatomic landmarks for surgery, as well as providing the advantages of objectivity and unlimited coverage irrespective of location [3, 29]. However, although CT is a standardized and objective imaging technique that is less operator-dependent than other modalities, a recent systematic review and meta-analysis demonstrated that CT achieved a summary sensitivity of 62% (95% confidence interval (CI), 52–70%) and specificity of 87% (95% CI, 80–92%) in patients with thyroid cancers [11]. Several studies also reported variable sensitivity (35.0–77.0%) using CT, depending on the

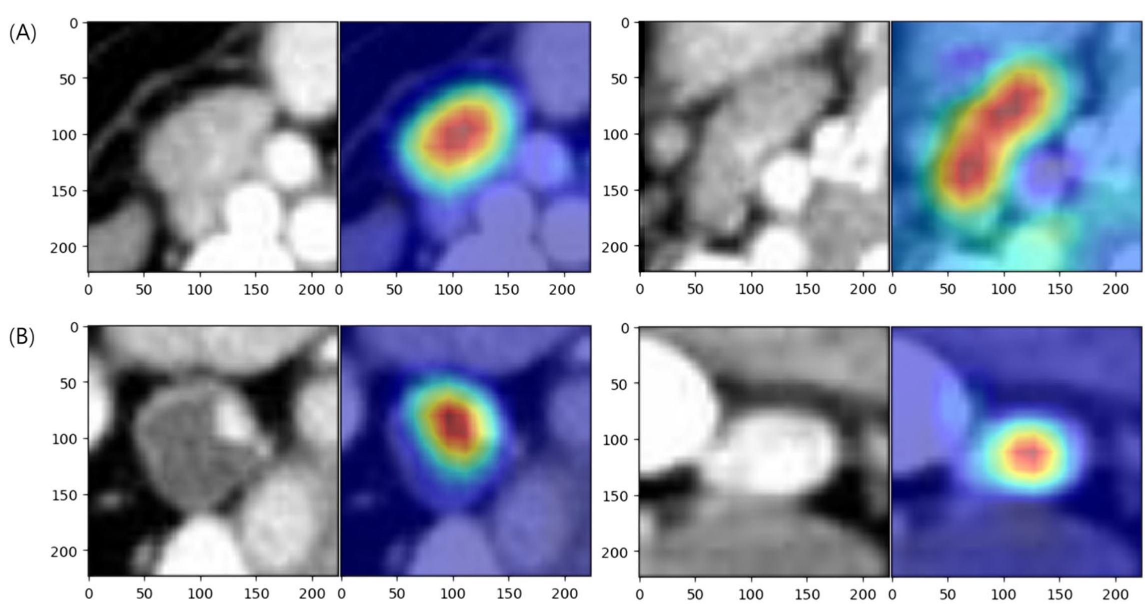

Fig. 3 Attention heatmap drawn by class activation mapping. The red color shows where the model is focused to differentiate benign from metastatic lymph nodes. The red color focuses on the central hilar vessels on benign lymph nodes (**a**) and on a strong enhancing area (focal or diffuse) on metastatic lymph nodes (**b**)

presence or absence of radiologic experts and their level of experience; this may cause the surgical extent to be inappropriately wide or narrow for lateral lymph node dissection [31–33]. While undertreatment of metastatic neck nodes during primary surgery due to under diagnosis will cause local recurrence, overtreatment with prophylactic lateral compartment dissection will increase surgical morbidity [3, 10].

Deep learning is a branch of artificial intelligence in which computers are not explicitly programmed, but instead perform tasks by analyzing relationships between existing data points [15]. Deep learning technology is considered state of the art for the classification of images, and is expected to play a role in assisting clinical decision-making with respect to diagnosis and risk stratification [34, 35]. In this study, a deep learning–based CAD system was implemented to classify cervical LNM in patients with thyroid cancer on CT images. When the diagnostic performance of the CAD system was compared with those in the recent meta-analysis [11], the CAD system in the present study achieved a higher sensitivity (90.4% vs. 62.0%) and specificity (90.4% vs. 87.0%). When compared with a multicenter study performed by experienced radiologists [12], the CAD system also showed a higher sensitivity (90.4% vs. 81.8%) and similar specificity (90.4% vs. 90.4%). Therefore, we consider that the CAD system in the present study may be useful in a clinical setting for detecting LNM before surgery. This deep learning–based CAD system may provide some advantages over the human eye for the evaluation of cervical LNM. First, it could provide accurate and consistent results for the same input image, which in turn could potentially eliminate the obstacle of inter-observer variability [36]. Second, interpretability is increased through the attention heatmap generated by CAM. It is generally difficult to explain the internal relationship between input data and the label predicted by the model. However, using the attention heatmap, we can infer which part of the input image is focused on by the model. Third, the CAD system has scalability. The performance of the CAD system could be upgraded to incorporate reports on incorrect results.

Classification of cervical lymph nodes according to the risk of nodal metastasis requires a certain level of experience and knowledge of CT images. Compared with cervical LNM in head and neck squamous cell carcinoma, metastatic lymph nodes in thyroid cancer have markedly more characteristic CT features, including strong enhancement (focal or diffuse), heterogeneous enhancement, calcification (micro/macro), or cystic change [10]. Benign lymph nodes are also characterized by central hilar fat and central hilar vessel enhancement [10]. Therefore, we consider that the high accuracy of our CAD system for the diagnosis of cervical LNM may be explained by these characteristic CT features in thyroid cancer. The attention heatmap of the benign and metastatic lymph nodes represented evidence of CNN model-based classification, and could assist in clinical decision-making by directly identifying the region of interest.

This study had some limitations. First, the model was developed for classifying lateral LNM. Therefore, diagnostic performance for classifying central LNM could not be evaluated. However, information on lateral LNM is in practice critical for surgical decision-making. Second, this model was not developed to localize the cervical lymph node. Users need to identify lymph nodes on CT and input the images into the CAD system. Third, this study was based on a single center; an external validation study is needed to validate its diagnostic performance and generalizability. Prospective and multiinstitutional datasets are also needed in future studies. Translating technical success to meaningful clinical impact is the next major challenge. Thorough evaluation and further improvement is required to integrate this technique into routine clinical care in the future.

In conclusion, we developed a deep learning–based CAD system for classifying cervical LNM on CT images in patients with thyroid cancer. Whether this approach has clinical utility will require evaluation in a clinical setting.

Funding The authors state that this work was supported by the National Research Foundation of Korea (no. 2017R1C1B5016217).

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Eun Ju Ha.

Conflict of interest The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

Statistics and biometry No complex statistical methods were necessary for this paper.

Informed consent Written informed consent was obtained from all patients before they underwent US.

Ethical approval This study was approved by our institutional review board.

Methodology
• retrospective
• case-control study

References

1. Davies L, Welch HG (2014) Current thyroid cancer trends in the United States. JAMA Otolaryngol Head Neck Surg 140:317–322
2. Li N, Du XL, Reitzel LR, Xu L, Sturgis EM (2013) Impact of enhanced detection on the increase in thyroid cancer incidence in the United States: review of incidence trends by socioeconomic status within the surveillance, epidemiology, and end results registry, 1980-2008. Thyroid 23:103–110
3. Haugen BR, Alexander EK, Bible KC et al (2016) 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid Cancer: the American

Thyroid Association guidelines task force on thyroid nodules and differentiated thyroid cancer. Thyroid 26:1–133
4. Al-Saif O, Farrar WB, Bloomston M, Porter K, Ringel MD, Kloos RT (2010) Long-term efficacy of lymph node reoperation for persistent papillary thyroid cancer. J Clin Endocrinol Metab 95:2187–2194
5. Stack BC Jr, Ferris RL, Goldenberg D et al (2012) American Thyroid Association consensus review and statement regarding the anatomy, terminology, and rationale for lateral neck dissection in differentiated thyroid cancer. Thyroid 22:501–508
6. Durante C, Montesano T, Torlontano M et al (2013) Papillary thyroid cancer: time course of recurrences during postsurgery surveillance. J Clin Endocrinol Metab 98:636–642
7. Tufano RP, Clayman G, Heller KS et al (2015) Management of recurrent/persistent nodal disease in patients with differentiated thyroid cancer: a critical review of the risks and benefits of surgical intervention versus active surveillance. Thyroid 25:15–27
8. Yeh MW, Bauer AJ, Bernet VA et al (2015) American Thyroid Association statement on preoperative imaging for thyroid cancer surgery. Thyroid 25:3–14
9. Wong SC, Gatt A, Stamatescu V, McDonnell MD (2016) Understanding data augmentation for classification: when to warp? https://arXiv.org/abs/1609.08764. Accessed 19 Nov 2018
10. Shin JH, Baek JH, Chung J et al (2016) Ultrasonography diagnosis and imaging-based management of thyroid nodules: revised Korean Society of Thyroid Radiology consensus statement and recommendations. Korean J Radiol 17:370–395
11. Suh CH, Baek JH, Choi YJ, Lee JH (2017) Performance of CT in the preoperative diagnosis of cervical lymph node metastasis in patients with papillary thyroid cancer: a systematic review and meta-analysis. AJNR Am J Neuroradiol 38:154–161
12. Lee Y, Kim JH, Baek JH et al (2018) Value of CT added to ultrasonography for the diagnosis of lymph node metastasis in patients with thyroid cancer. Head Neck. https://doi.org/10.1002/hed.25202
13. Shen D, Wu G, Suk HI (2017) Deep learning in medical image analysis. Annu Rev Biomed Eng 19:221–248
14. Miotto R, Wang F, Wang S, Jiang X, Dudley JT (2017) Deep learning for healthcare: review, opportunities and challenges. Brief Bioinform
15. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436
16. Choi YJ, Baek JH, Park HS et al (2017) A computer-aided diagnosis system using artificial intelligence for the diagnosis and characterization of thyroid nodules on ultrasound: initial clinical assessment. Thyroid 27:546–552
17. Wu MH, Chen CN, Chen KY et al (2016) Quantitative analysis of echogenicity for patients with thyroid nodules. Sci Rep 6:35632
18. Jeong EY, Kim HL, Ha EJ et al (2018) Computer-aided diagnosis system for thyroid nodules on ultrasonography: diagnostic performance and reproducibility based on the experience level of operators. Eur Radiol
19. Lee JH, Baek JH, Kim JH et al (2018) Deep learning-based computer-aided diagnosis system for localization and diagnosis of metastatic lymph nodes on ultrasound: a pilot study. Thyroid 28:1332–1338
20. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. https://arXiv.org/abs/1409.1556. Accessed 19 Nov 2018
21. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognitionProceedings of the IEEE conference on computer vision and. Pattern Recogn:770–778
22. Chollet F (2017) Xception: deep learning with depthwise separable convolutions. https://arXiv.org/abs/1610.02357. Accessed 19 Nov 2018
23. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. https://arXiv.org/abs/1512.00567. Accessed 19 Nov 2018
24. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA (2016) Inception-v4, inception-resnet and the impact of residual connections on learning. https://arXiv.org/abs/1602.07261. Accessed 19 Nov 2018
25. Iandola F, MoskewiczM,Karayev S, Girshick R, Darrell T, Keutzer K (2014) Densenet: Implementing efficient convnet descriptor pyramids. https://arXiv.org/abs/1404.1869. Accessed 19 Nov 2018
26. Russakovsky O, Deng J, Su H et al (2015) Imagenet large scale visual recognition challenge. Int J Comput Vis 115:211–252
27. Abadi M, Barham P, Chen J et al (2016) Tensorflow: a system for large-scale machine learning. https://arXiv.org/abs/1605.08695. Accessed 19 Nov 2018
28. Mulla MG, Knoefel WT, Gilbert J, McGregor A, Schulte KM (2012) Lateral cervical lymph node metastases in papillary thyroid cancer: a systematic review of imaging-guided and prophylactic removal of the lateral compartment. Clin Endocrinol 77:126–131
29. Moon HJ, Kim EK, Yoon JH, Kwak JY (2012) Differences in the diagnostic performances of staging US for thyroid malignancy according to experience. Ultrasound Med Biol 38:568–573
30. Mulla M, Schulte KM (2012) Central cervical lymph node metastases in papillary thyroid cancer: a systematic review of imaging-guided and prophylactic removal of the central compartment. Clin Endocrinol 76:131–136
31. Jeong HS, Baek CH, Son YI et al (2006) Integrated 18F-FDG PET/CT for the initial evaluation of cervical node level of patients with papillary thyroid carcinoma: comparison with ultrasound and contrast-enhanced CT. Clin Endocrinol 65:402–407
32. Ahn JE, Lee JH, Yi JS et al (2008) Diagnostic accuracy of CT and ultrasonography for evaluating metastatic cervical lymph nodes in patients with thyroid cancer. World J Surg 32:1552–1558
33. Lee DW, Ji YB, Sung ES et al (2013) Roles of ultrasonography and computed tomography in the surgical management of cervical lymph node metastases in papillary thyroid carcinoma. Eur J Surg Oncol 39:191–196
34. Ha EJ, Baek JH, Na DG (2017) Risk stratification of thyroid nodules on ultrasonography: current status and perspectives. Thyroid 27:1463–1468
35. Esteva A, Kuprel B, Novoa RA et al (2017) Dermatologist-level classification of skin cancer with deep neural networks. Nature 542:115
36. Lee HJ, Yoon DY, Seo YL et al (2018) Intraobserver and interobserver variability in ultrasound measurements of thyroid nodules. J Ultrasound Med 37:173–178

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.