



Dual-Energy CT Texture Analysis With Machine Learning for the Evaluation and Characterization of Cervical Lymphadenopathy

Matthew Seidler^a, Behzad Forghani^b, Caroline Reinhold^{a,b}, Almudena Pérez-Lara^{a,1}, Griselda Romero-Sanchez^a, Nimesh Muthukrishnan^c, Julian L. Wichmann^d, Gabriel Melki^c, Eugene Yu^e, Reza Forghani^{a,b,c,f,g,*}

^a Department of Radiology, McGill University, Rm C5 118, 1650 Cedar Avenue, Montreal, Quebec H3G 1A4, Canada

^b Department of Radiology and Research Institute of McGill University Health Centre, 1001 boul. Decarie Blvd, Montreal, Quebec H3A 3J1, Canada

^c Segal Cancer Centre and Lady Davis Institute for Medical Research, Jewish General Hospital, Rm C-212.1, 3755 Cote Ste-Catherine Road, Montreal, Quebec H3T 1E2, Canada

^d Department of Diagnostic and Interventional Radiology, University Hospital Frankfurt, Theodor-Stern-Kai 7, 60590 Frankfurt, Germany

^e Joint Department of Medical Imaging, Princess Margaret Hospital, University Health Network, University of Toronto, Rm 3-959, 610 University Ave, Toronto, Ontario M5G 2M9, Canada

^f Gerald Bronfman Department of Oncology, McGill University, Suite 720, 5100 Maisonneuve Blvd West, Montreal, Quebec H4A3T2, Canada

^g Department of Otolaryngology, Head and Neck Surgery, Royal Victoria Hospital, McGill University Health Centre, 1001 boul. Decarie Blvd, Montreal, Quebec H3A 3J1, Canada

ARTICLE INFO

Article history:

Received 29 November 2018

Received in revised form 9 July 2019

Accepted 10 July 2019

Available online 16 July 2019

Keywords:

Lymph nodes

Texture analysis

Radiomics

Machine learning

Artificial intelligence

Dual energy CT

ABSTRACT

Purpose: To determine whether machine learning assisted-texture analysis of multi-energy virtual monochromatic image (VMI) datasets from dual-energy CT (DECT) can be used to differentiate metastatic head and neck squamous cell carcinoma (HNSCC) lymph nodes from lymphoma, inflammatory, or normal lymph nodes.

Materials and methods: A retrospective evaluation of 412 cervical nodes from 5 different patient groups (50 patients in total) having undergone DECT of the neck between 2013 and 2015 was performed: (1) HNSCC with pathology proven metastatic adenopathy, (2) HNSCC with pathology proven benign nodes (controls for (1)), (3) lymphoma, (4) inflammatory, and (5) normal nodes (controls for (3) and (4)). Texture analysis was performed with TexRAD® software using two independent sets of contours to assess the impact of inter-rater variation. Two machine learning algorithms (Random Forests (RF) and Gradient Boosting Machine (GBM)) were used with independent training and testing sets and determination of accuracy, sensitivity, specificity, PPV, NPV, and AUC.

Results: In the independent testing (prediction) sets, the accuracy for distinguishing different groups of pathologic nodes or normal nodes ranged between 80 and 95%. The models generated using texture data extracted from the independent contour sets had substantial to almost perfect agreement. The accuracy, sensitivity, specificity, PPV, and NPV for correctly classifying a lymph node as malignant (i.e. metastatic HNSCC or lymphoma) versus benign were 92%, 91%, 93%, 95%, 87%, respectively.

Conclusion: Machine learning assisted-DECT texture analysis can help distinguish different nodal pathology and normal nodes with a high accuracy.

© 2019 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Identification and accurate characterization of abnormal lymph nodes is an essential and common task in imaging of the neck. In patients with head and neck cancer, nodal status is one of the most

important determinants of outcome and can have significant impact on patient staging and management [1–3]. In routine clinical practice, the most common characteristics used to distinguish pathologic from benign lymph nodes on CT or MRI are size, internal heterogeneity (commonly referred to as internal nodal necrosis), shape, contour, and clustering, among others [1]. Node size is probably the most commonly used criterion, although size criteria have overall error rates of 15 to 20% or more for the determination of lymphadenopathy in the neck [1]. Identification of pathologic lymph nodes becomes especially challenging in small lymph nodes measuring less than 10 mm, even with functional techniques such as PET/CT [1,4–6].

Many of the currently used criteria for determination of pathologic nodes are derived from evaluation of metastatic head and neck

Abbreviations: DECT, Dual-energy CT; VMI, Virtual monochromatic image; HNSCC, Head and neck squamous cell carcinoma; PPV, Positive predictive value; NPV, Negative predictive value; AUC, Area under the receiver operating curve.

* Corresponding author at: Department of Radiology, McGill University Health Centre; Room C02.5821, 1001 Decarie Blvd, Montreal H4A 3J1 QC, Canada.

E-mail address: reza.forghani@mcgill.ca (R. Forghani).

¹ Present address: Department of Radiology, Hospital Regional Universitario de Málaga, Avenida Carlos Haya, S/N, 29010 Málaga, Spain.

<https://doi.org/10.1016/j.csbj.2019.07.004>

2001-0370/© 2019 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

squamous cell carcinoma (HNSCC) lymph nodes [1,6–8] and extrapolated to other pathologic entities. With a few exceptions (e.g. presence of microcalcifications in untreated lymph nodes suggesting metastatic papillary thyroid cancer), pathologic nodes of different primary etiologies are not distinguishable based on their imaging characteristics. Advanced techniques such as dual-energy CT (DECT) [9–11] have been reported to improve diagnostic evaluation of different nodal pathology based on evaluation of different DECT quantitative parameters, including those derived from quantitative information derived from virtual monochromatic images (VMIs) constructed at different energies [12–17]. Therefore, it is possible that advanced image analysis methods such as texture analysis performed on multi-energy datasets may also be advantageous for nodal characterization, similar to a study demonstrating an advantage of multi-energy texture analysis for characterization of benign parotid tumors [18,19].

Texture or radiomic analysis, ideally supported by machine learning approaches for constructing prediction models, has been used in different organ systems to predict tumor characteristics such as molecular features, patient prognosis, and response to treatment [20–27]. Studies have also shown potentially promising results in applying texture analysis for the evaluation of pathologic lymph nodes outside the neck, especially in the mediastinum in the context of lung cancer [28–30]. However, the value of this approach for characterization of nodal pathology in the neck is currently not well established. The goal of our study was to investigate the use of machine learning-assisted rapid-kVp-switching DECT texture analysis employing virtual monochromatic imaging (VMI) datasets for differentiating abnormal from normal lymph nodes and characterization of different nodal pathologies including HNSCC metastasis, lymphoma, and inflammatory lymph nodes.

2. Materials and Methods

2.1. Patient Population

This study was approved by the institutional review board with waiver of informed consent. All of the patients selected were from a tertiary-care hospital and cancer center (Jewish General Hospital, Montreal, Quebec, Canada). A total of 412 lymph nodes from 50 patients who had undergone clinically-indicated DECT scans of the neck between November 2013 and December 2015 were retrospectively analyzed using texture analysis. Ten consecutive patients meeting inclusion criteria for this study (see flowcharts in Figs. 1 and 2) were included from each of the five nodal categories, with multiple lymph nodes (total number of lymph nodes is shown in parenthesis) evaluated per patient as follows: HNSCC with metastatic adenopathy ($n = 31$), HNSCC with benign lymph nodes ($n = 145$; control group 1), lymphoma ($n = 65$), inflammatory ($n = 29$), and normal scans from patients without a known history of malignancy or significant abnormality on the neck CT ($n = 142$; control group 2) (Figs. 1 and 2). All metastatic HNSCC nodes evaluated were pathologically proven, and all benign nodes in HNSCC patients were likewise based on pathology proven neck dissection specimens (Fig. 1). Furthermore, the latter group was selected based on completely absent nodal metastases in the neck on neck dissection, in order to avoid confounding and challenges with distinguishing and correlating normal appearing lymph nodes on imaging with nodes with micrometastases on pathology that may not be possible to perform accurately in a retrospective setting. For the other groups evaluated, it would not be possible to have pathology proof of every node, therefore strict criteria were used (in biopsy proven disease where applicable, e.g. in the case of lymphoma patients), that are

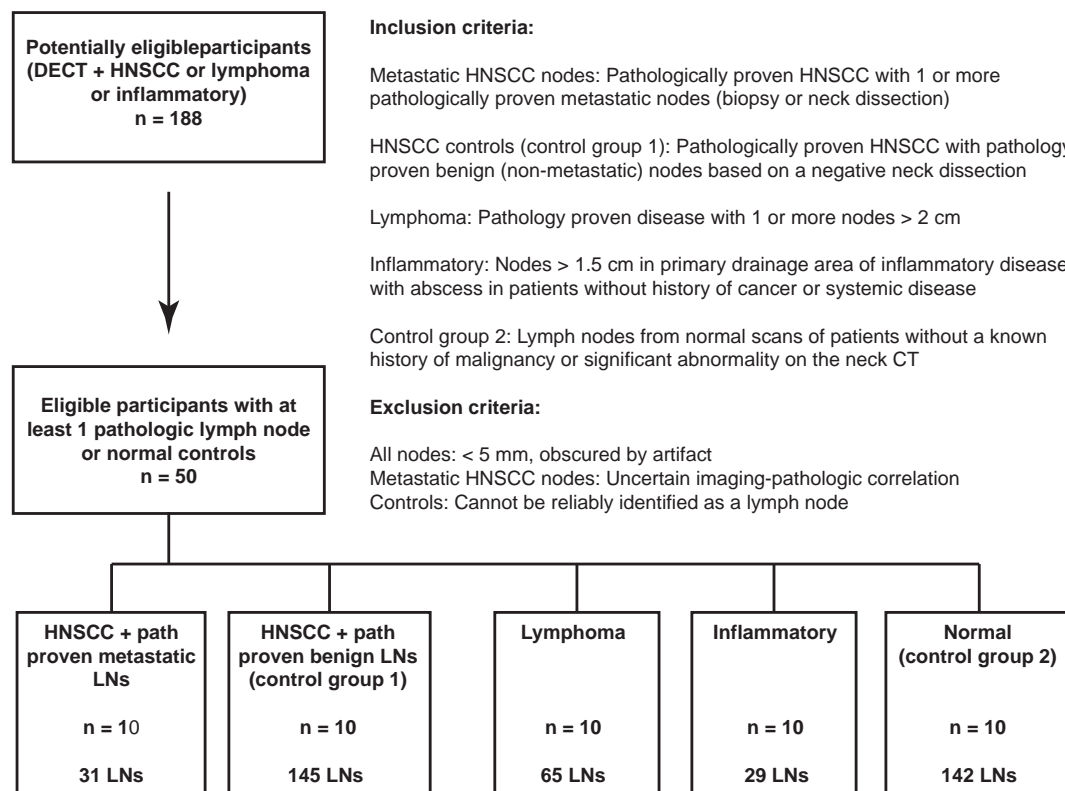


Fig. 1. Flowchart of patient and node selection for texture analysis, with inclusion and exclusion criteria. For HNSCC, only patients with pathological confirmation based on lymph node dissection, either positive or negative, were included. Because lymphoma patients or patients with inflammatory disease do not undergo cervical lymph node dissection as part of their treatment, the above strict criteria were used for selection of nodes in these patient groups (please see text and Supplemental Data for additional details). Two groups of controls were used: (1) pathologically proven benign nodes from HNSCC patients (group 1, controls for metastatic HNSCC nodes) and (2) nodes from normal neck scans of patients without history of cancer or other systemic disease (group 2, controls for lymphoma and inflammatory groups).

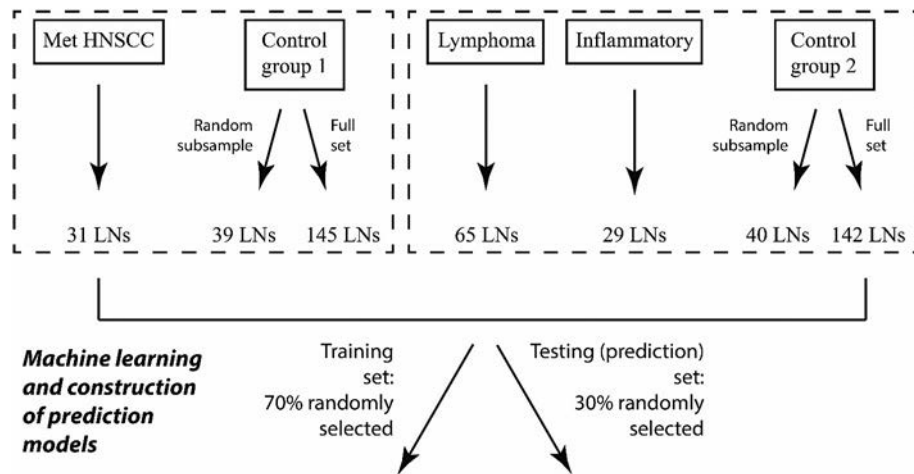


Fig. 2. Flowchart of node selection and groupings for construction of prediction models using machine learning. Because of the large imbalance between the number of nodes in the control groups compared to metastatic HNSCC or inflammatory lymph node groups, for pairwise comparison of these groups and controls, a maximum of 4 lymph nodes per control patient were randomly selected for use by the machine learning algorithm. This resulted in a subset of 39 nodes for control group 1 and 40 nodes for control group 2. For all other groupings and comparisons, the full complement of control nodes was used.

described in detail in the inclusion criteria in Fig. 1 and in the Supplemental Material. The underlying etiology for the inflammatory cases was complicated odontogenic disease or sialolithiasis.

2.2. CT Technique & DECT Image Post-Processing

All patients were scanned using the same 64-section rapid-kVp-switching dual-energy scanner (Discovery CT750 HD; GE Healthcare, Milwaukee, Wisconsin) following injection of 80 mL of iopamidol (Isovue 300; Bracco, Princeton, New Jersey) at a rate of 2 mL/s, with a delay of 65 s. Scans were acquired in dual-energy rapid 80 to 140 kVp switching mode [31,32] and images were reconstructed into 1.25 mm sections in a 25 cm display FOV and 512 × 512 matrix without iterative reconstruction and using a standard kernel (see Supplemental Material for additional details). In order to attempt to take advantage of the energy-dependent changes in tissue attenuation possible with DECT as demonstrated in different studies evaluating head and neck cancer [18,19,31–33], multi-energy VMIs were reconstructed from de-identified scans, ranging from 40 to 140 keV in 5 keV increments at the GE Advantage workstation (4.6; GE Healthcare, Milwaukee, WI). This resulted in 21 different reconstructions per case for multi-energy VMI analysis.

2.3. Selection of Lymph Nodes & Texture Analysis

All lymph nodes meeting inclusion criteria (Figs. 1 and 2) were included in this study. Texture analysis was performed using a filtration-histogram technique with a commercially-available research software (TexRAD Ltd., Cambridge, United Kingdom) by manually delineating a region of interest (ROI) around the largest diameter of the lymph node in the axial plane. During the initial analysis, all contours were first drawn by (A.P.), a fellowship-trained Neuroradiologist with 2 years of neuroradiology experience. These contours were later reviewed and approved (or revised if needed) by a fellowship-trained academic head and neck radiologist with 7 years of experience in oncologic imaging (R.F.). To evaluate the impact of inter-rater variation on texture analysis and the final prediction models, another neuroradiology and head and neck imaging fellowship trained radiologist (G.R.S.) manually re-contoured the lymph nodes for repeat independent texture analysis and prediction model reconstruction. For additional information please refer to the Supplemental Data.

For each lymph node, texture data were extracted from all 21 VMI series using the same ROI. This is possible because of the inherent co-registration of different energy VMI datasets generated from the same DECT scan [18,19]. For each ROI and VMI energy combination, six texture

features based on first-order statistics of the gray level intensity histogram were derived: (1) average intensity values, (2) standard deviation (SD), (3) mean of positive pixels (MPP), (4) entropy, (5) skewness, and (6) kurtosis [19,34]. In addition, for each feature set, either no filter or 5 different spatial scale filter settings were used to highlight features at different anatomic spatial scales ranging between 2 and 6 mm [34–36] (Fig. A1, Supplementary Data). Please see Supplementary Data for additional details.

2.4. Machine Learning & Statistical Analysis

Prediction models were built using either texture data extracted at a single VMI energy of 65 keV, typically considered equivalent and used as a replacement for a conventional 120 kVp single energy neck CT acquisition when a scan is acquired in DECT mode [32,37–40], or multi-energy analysis of the entire 21 VMI datasets [18,31–33,41]. Two independent machine learning approaches, the Random Forests (RF) method [42] and Gradient Boosting Machine (GBM) method [43,44] were used to build the prediction models for the different outcomes, consisting of pairwise evaluations of different pathologic lymph nodes and the normal controls. Finally, RF was used to compare and distinguish different neoplastic from non-neoplastic nodes simultaneously (metastatic HNSCC or lymphoma, from benign nodes proven based on negative neck dissections in HNSCC patients). For an unbiased assessment of the model accuracy [45], 30% of the patients were randomly selected and set aside as the test group. The remaining 70% were used to train each prediction model. These are described in greater detail in the Supplemental Data. Once the final prediction model was found, the accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and area under the receiver operating curve (AUC) of the final models were computed for the training and testing (prediction) sets. The R package (R Development Core Team (2008), Vienna, Austria. ISBN 3–900,051–07–0, <http://www.R-project.org>) was used for the machine learning and statistical analysis. The R package, randomForest, was used for building the Random Forests models [46]. The R package, gbm, was used for building the GBM models [47]. The agreement between prediction models generated using data extracted from contouring by different readers was assessed by calculating Cohen's kappa coefficient (κ) based on binary classification of the nodes into different categories by the prediction models.

3. Results

A total of 412 lymph nodes were evaluated using texture analysis. For the HNSCC group, 31 metastatic lymph nodes and 145

Table 1
Prediction accuracy for distinction of Metastatic HNSCC from normal nodes.

	ML approach	Acc	Sens	Spec	PPV	NPV	AUC
Training set	RF	48/50 (96%)	21/22 (95%)	27/28 (96%)	21/22 (95%)	27/28 (96%)	0.95
	GBM	49/50 (98%)	21/22 (95%)	28/28 (100%)	21/21 (100%)	28/29 (97%)	1.00
Testing set	RF	17/20 (85%; 69, 100)	8/9 (89%; 68, 100)	9/11 (82%; 59, 100)	8/10 (80%; 55, 100)	9/10 (90%; 71, 100)	0.97 (0.89, 1.00)
	GBM	18/20 (90%; 77, 100)	8/9 (89%; 68, 100)	10/11 (91%; 74, 100)	8/9 (89%; 68, 100)	10/11 (91%; 74, 100)	0.96 (0.87; 1.00)

Models constructed are based on analysis of VMIs at 65 keV using texture data extracted from the first set of independent contours.

For the testing sets (prediction models), the lower and upper limits of the 95% confidence interval are provided after the percentage value or the AUC.

ML: Machine Learning; RF: Random Forests; GBM: Gradient Boosting Machine.

histopathology proven benign nodes in HNSCC patients (based on neck dissection) were used for constructing the prediction models. For the other 3 patient groups, 65 lymphoma nodes, 29 inflammatory nodes, and 142 nodes from normal controls were evaluated. For pairwise analysis of HNSCC and inflammatory nodes versus controls, analysis and construction of prediction models using machine learning was also repeated by randomly selecting a maximum of 4 normal nodes per patient in the control groups. This was done to reduce the imbalance between the number of normal and pathologic nodes which could adversely affect prediction model performance and reliability. Patient demographics are shown in Table A1 (Supplementary Data). The following models were evaluated using texture data extracted from 65 keV VMIs only or from 21 sets of multi-energy VMIs ranging between 40 and 140 keV. Although the final features for multi-energy model were selected from various energies other than 65 keV, they were highly correlated with those at 65 keV and no significant improvement in performance was observed and thus, only the 65 keV model performance is discussed in detail.

3.1. Distinguishing Neoplastic or Inflammatory Nodes From Normal Lymph Nodes on 65 keV VMIs

Texture analysis of 65 keV VMIs had a high performance for distinguishing metastatic HNSCC from benign nodes using either machine learning approach (Table 1, Fig. 3). In the independent testing (prediction) set, there was an accuracy, sensitivity, specificity, PPV, NPV, and AUC of 85%, 89%, 82%, 80%, 90%, and 0.97, respectively, using RF and 90%, 89%, 91%, 89%, 91%, and 0.96 respectively, using GBM. The accuracy of the training set was 96% for RF and 98% for GBM and comparison of the performance of the testing compared to the training sets did not suggest any overfitting (overfitting is a modelling error where the algorithm simulates the training data too closely, using noise or random fluctuations in the training data as concepts that may

not be applicable to new datasets and consequently negatively impact algorithm performance in new datasets (i.e. generalization of the model); simply put, overfitting provides a falsely optimistic measure of algorithm performance). Two features (or predictors) were used in the final model: SSF6.entropy.65 and SSF0.skewness.65 (classifier description represents *spatial scale filter.texture features based on first-order statistics of the gray level intensity histogram.VMI energy*).

There was also a high accuracy for distinguishing lymphoma from normal nodes (Table 2), with an accuracy, sensitivity, specificity, PPV, NPV, and AUC of 90%, 100%, 86%, 76%, 100%, and 0.95, respectively, using RF or 93%, 100%, 90%, 83%, 100%, and 0.96, respectively, using GBM. The accuracy of the training set was 88% for RF and 95% for GBM. Two features were used in the final model: SSF2.MPP.65 and SSF2.SD.65. There was a fairly high performance of the prediction models for distinguishing inflammatory from normal nodes (Table 3) with an accuracy, sensitivity, specificity, PPV, NPV, and AUC of 80%, 88%, 75%, 70%, 90%, 0.97, respectively, using RF or 80%, 88%, 75%, 70%, 90%, and 0.96, respectively, using GBM. The accuracy of the training set was 90% for RF and 100% for GBM. Two features were used in the final model: SSF5.entropy.65 and SSF6.mean.65.

When different malignant nodes were combined, using RF, the accuracy, sensitivity, specificity, PPV, and NPV for correctly classifying a lymph node as malignant (i.e. metastatic HNSCC or lymphoma) versus benign in the prediction set was 92%, 91%, 93%, 95%, 87%, respectively (Table 4). Two features were used in the final model: SSF0.skewness.65 and SSF2.SD.65. Additional information, including more detailed information on the training sets, is provided in Tables 1–4.

We concluded this section by re-generating the above testing (prediction) models using texture data extracted from manual contouring by a second reader, in order to assess the impact of inter-rater contour variation on prediction models generated using this approach. As shown in Table 5, there was substantial to almost perfect agreement between the models generated based from the two different readers.

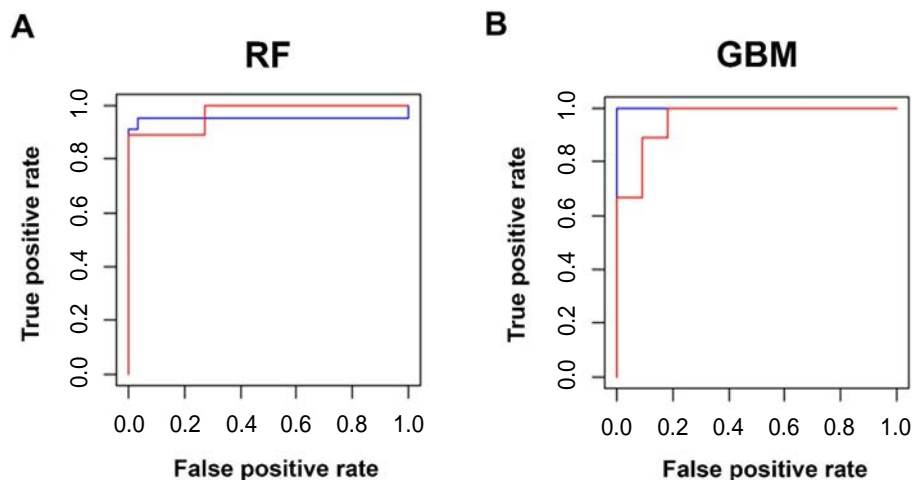


Fig. 3. Example of ROC curve analysis of the diagnostic performance of A, Random Forests (RF) and B, Gradient Boosting Machine (GBM) method for distinguishing metastatic HNSCC from benign lymph nodes. The testing (prediction) sets (red) have similar performance to the training sets (blue), as expected, suggesting that the models are reliable and that there is not overfitting. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2

Prediction accuracy for distinction of Lymphoma from normal nodes.

	ML approach	Acc	Sens	Spec	PPV	NPV	AUC
Training set	RF	129/146 (88%)	38/46 (83%)	91/100 (91%)	38/47 (81%)	91/99 (92%)	0.95
	GBM	138/146 (95%)	42/46 (91%)	96/100 (96%)	42/46 (91%)	96/100 (96%)	0.99
Testing set	RF	55/61 (90%; 83, 98)	19/19 (100%; 100, 100)	36/42 (86%; 75, 96)	19/25 (76%; 59, 93)	36/36 (100%; 100, 100)	0.95 (0.90, 1.00)
	GBM	57/61 (93%; 87, 100)	19/19 (100%; 100, 100)	38/42 (90%; 82, 99)	19/23 (83%; 67, 98)	38/38 (100%; 100, 100)	0.96 (0.91, 1.00)

Models constructed are based on analysis of VMIs at 65 keV using texture data extracted from the first set of independent contours.

For the testing sets (prediction models), the lower and upper limits of the 95% confidence interval are provided after the percentage value or the AUC.

ML: Machine Learning; RF: Random Forests; GBM: Gradient Boosting Machine.

Table 3

Prediction accuracy for distinction of inflammatory from normal nodes.

	ML approach	Acc	Sens	Spec	PPV	NPV	AUC
Training set	RF	44/49 (90%)	18/21 (86%)	26/28 (93%)	18/20 (90%)	26/29 (90%)	0.97
	GBM	49/49 (100%)	21/21 (100%)	28/28 (100%)	21/21 (100%)	28/28 (100%)	1.00
Testing set	RF	16/20 (80%; 62, 98)	7/8 (88%; 65, 100)	9/12 (75%; 50, 100)	7/10 (70%; 42, 98)	9/10 (90%; 71, 100)	0.97 (0.89, 1.00)
	GBM	16/20 (80%; 62, 98)	7/8 (88%; 65, 100)	9/12 (75%; 50, 100)	7/10 (70%; 42, 98)	9/10 (90%; 71, 100)	0.96 (0.87, 1.00)

Models constructed are based on analysis of VMIs at 65 keV using texture data extracted from the first set of independent contours.

For the testing sets (prediction models), the lower and upper limits of the 95% confidence interval are provided after the percentage value or the AUC.

ML: Machine Learning; RF: Random Forests; GBM: Gradient Boosting Machine.

Table 4

Prediction accuracy for Lymph node classification as malignant versus Benign.

	Acc	Sens	Spec	PPV	NPV
Training set	155/170 (91%)	95/102 (93%)	60/68 (88%)	95/103 (92%)	60/67 (90%)
Testing set	65/71 (92%; 85, 98)	39/43 (91%; 82, 99)	26/28 (93%; 83, 100)	39/41 (95%; 89, 100)	26/30 (87%; 75, 99)

Models constructed are based on analysis of VMIs at 65 keV using texture data extracted from the first set of independent contours, using Random Forests, based on texture data from histopathology proven benign lymph nodes (negative node dissection in HNSCC patients), metastatic HNSCC, and lymphoma nodes.

For the testing sets (prediction models), the lower and upper limits of the 95% confidence interval are provided after the percentage value.

3.2. Distinction and Classification of Different Abnormal Nodes on 65 keV VMIs

The accuracy, sensitivity, specificity, PPV, NPV, and AUC of texture analysis of 65 keV VMIs for distinguishing metastatic HNSCC from lymphoma nodes in the prediction set was 93%, 78%, 100%, 100%, 90%, and 0.96, respectively, using RF or 89%, 78%, 95%, 88%, 90%, and 0.95, respectively, using GBM (Table A2, Supplementary Data). Three features were used in the final model: SSF0.mean.65, SSF0.SD.65, and SSF5.skewness.65. The accuracy, sensitivity, specificity, PPV, NPV, and AUC of texture analysis of 65 keV VMIs for distinguishing metastatic HNSCC from inflammatory nodes in the prediction set was 82%, 78%, 88%, 88%, 78%, and 0.92, respectively, using RF or 82%, 78%, 88%, 88%, 78%, and 0.82, respectively, using GBM (Table A3, Supplementary Data). In this case, the features used in the final models were not identical for RF and GBM. For RF, 4 features were used in the final model: SSF6.kurtosis.65, SSF2.entropy.65, SSF0.MPP.65, and SSF3.skewness.65. For GBM, 4 features were used in the final model: SSF6.kurtosis.65, SSF0.mean.65, SSF2.entropy.65, and SSF2.mean.65.

The accuracy, sensitivity, specificity, PPV, NPV, and AUC of texture analysis of 65 keV VMIs for distinguishing lymphoma from inflammatory nodes in the prediction set was 85%, 100%, 50%, 83%, 100%, and 0.88, respectively, using RF or 85%, 100%, 50%, 83%, 100%, and 0.95 respectively, using GBM (Table A4, Supplementary Data). Two features were used in the final model: SSF2.MPP.65 and SSF4.mean.65. Once again, in this case, the features used in the final models were not identical for RF and GBM. For RF, 4 features were used in the final model: SSF3.MPP.65, SSF0.skewness.65, SSF5.skewness.65, and SSF6.kurtosis.65. For GBM, 4 features were used in the final model: SSF3.MPP.65, SSF6.mean.65, SSF6.kurtosis.65, and SSF5.skewness.65.

3.3. Analysis of VMIs at a Single Energy Versus Multi-Energy VMIs

All of the models evaluated using texture data extracted from 65 keV VMIs were also investigated using texture data extracted from 21 sets of multi-energy VMIs ranging between 40 and 140 keV. Although the final features selected were from various energies other than 65 keV, they were highly correlated with those at 65 keV and no significant

Table 5

Comparison of testing (prediction) model performance using texture data extracted from manual contouring by two different readers.

	Testing paradigm	Acc (first reader)	Acc (second reader)	Cohen's kappa coefficient (κ)
RF	Distinction of Metastatic HNSCC from normal nodes	17/20 (85%; 69, 100)	19/20 (95%; 85, 100)	0.8
	Distinction of Lymphoma from normal nodes	55/61 (90%; 83, 98)	55/61 (90%; 83, 98)	0.86
	Distinction of Inflammatory from normal nodes	16/20 (80%; 62, 98)	17/20 (85%; 69, 100)	0.9
	Lymph Node classification as Malignant versus Benign	65/71 (92%; 85, 98)	66/71 (93%; 87, 99)	0.91
GBM	Distinction of Metastatic HNSCC from normal nodes	18/20 (90%; 77, 100)	18/20 (90%; 77, 100)	0.8
	Distinction of Lymphoma from normal nodes	57/61 (93%; 87, 100)	56/61 (92%; 85, 99)	0.96
	Distinction of Inflammatory from normal nodes	16/20 (80%; 62, 98)	18/20 (90%; 77, 100)	0.8

Models constructed are based on analysis of VMIs at 65 keV.

For the testing sets (prediction models), the lower and upper limits of the 95% confidence interval are provided after the percentage value.

RF: Random Forests; GBM: Gradient Boosting Machine.

improvement in performance could be achieved for any of the models described. In some cases, the results were identical (e.g. accuracy for distinction of metastatic HNSCC from normal nodes for reader 2 was 95% for multi-energy analysis and 95% for analysis of VMIs at 65 keV). In some cases, there was a few percentage point variations that were not significant with very similar or in some cases identical confidence intervals.

4. Discussion

In this study, we demonstrate that texture analysis with machine learning can be used to distinguish different pathologic and normal lymph nodes in the neck with a very high accuracy. Not only was this approach useful for distinguishing abnormal from normal lymph nodes, the approach also showed a good performance for distinguishing different types of pathologic nodes, something that is typically not possible using current imaging criteria used for the evaluation of lymphadenopathy clinically. The prediction models had good performance for distinguishing normal from abnormal lymph nodes, which could be particularly useful in the clinical setting. There was substantial to almost perfect agreement between the models generated based on contouring from the two different readers which is promising, and any variation seen can likely be further reduced or eliminated based on future innovations required for implementation into the clinical arena, such as semi-automatic or automatic contouring.

So far, there have been few investigations performing texture or radiomic analysis of lymph nodes, especially in the neck. In a recent study performed on [18F]-FDG-PET scans that included a contrast-enhanced CT, it was reported that there are statistically significant differences between texture features of benign HIV-related lymphadenopathy compared to metastatic HNSCC nodes [48]. Another recent study reported that texture features of the primary HNSCC tumor may be useful for predicting associated cervical lymphadenopathy [49]. Our study confirms the utility of nodal texture analysis, demonstrating high performance for predicting different malignant lymph nodes as well as for distinguishing different types of nodal pathology. Interestingly, considering what has been described for prediction of nodal pathology using texture features of the primary HNSCC tumor [49], there is potential for synergy for a combination approach, using primary tumor texture features combined with those of lymph nodes for further enhancing model performance. This is an interesting topic for future research. A radiomic model combined with machine learning is non-invasive, can be performed on imaging studies already obtained as part of the patient's initial work up, and can be performed pre-operatively. Therefore, the results of this study suggest important potential for future development of artificial intelligence clinical assistant tools for patient diagnostic work up pertaining to the evaluation of cervical lymphadenopathy. Nevertheless, these results require validation in larger patient sets. We observed fairly consistent performance using two different classic machine learning approaches with the use of few features or predictors (2 to 4 depending on the model), and absence of any indication to suggest significant overfitting based on comparison of the training and prediction (test) sets. This increases confidence in the results obtained. These results also provide the basis for future studies focusing on the more challenging task of characterizing small lymph nodes measuring less than 1 cm that frequently cannot be reliably assessed by currently used imaging criteria [1,6].

Although we evaluated various VMI reconstructions based on DECT scans in this study, high performance was achieved based on extraction of data from the 65 keV VMIs, typically considered equivalent to the standard 120-kVp single-energy CT acquisition [32,37–40]. This finding suggests that our approach could potentially also be applied to single-energy CT scans, although this will require independent validation. Interestingly, we did not observe any significant improvement in prediction model performance when we used

the full complement of 21 VMI datasets reconstructed at different energies as might have been expected on theoretical grounds [18] or based on prior published investigations in the literature [19,49–52]. However, this is not entirely surprising given the very high baseline performance using 65 keV VMIs at a single energy, with the upper limits of the confidence intervals in the high 90s or 100%. This leaves little to no room for improvement, even without getting into the additional statistical and mathematical challenges that are posed by the much larger data in the multi-energy VMIs. Especially because of this, we do not believe that this observation or the lack of additional advantages of multi-energy VMIs on the current study can be generalized to other pathology, and the potential advantages of the latter would have to be evaluated on a case by case basis.

Our study has several limitations, the main one being small numbers of patients and lymph nodes evaluated, which is reflected in the large confidence intervals for some of the models. Although for HNSCC we used histopathology proven malignant and benign nodes, this could not be done with lymphoma and inflammatory nodes because, unlike HNSCC, lymphoma patients do not undergo neck dissection and inflammatory lymph nodes are not biopsied except for rare circumstances where unusual inflammatory processes are suspected. However, to overcome these inherent obstacles, we used strict criteria for selection of nodes and avoided any nodes that were borderline or equivocally abnormal. Our study assumes that the characteristics of different pathologic nodes in the same patient are independent, and this is supported by the fact that a given patient can have both normal and malignant lymph nodes or that malignant lymph nodes in the same patient may respond differently to treatment in some cases. In order to reduce potential bias, the nodes in different groups were randomly selected by the machine learning algorithm and a review of the selected nodes in the testing (prediction) sets revealed good representation across different patients. Nonetheless, it is possible that there is some unintended bias from inclusion of multiple pathologic nodes from the same patient and a definitive determination in this regard will ultimately have to be made in future studies with much larger patient numbers. For this initial study, we used a homogenous technique, using data generated from the same institution and scanner. There is furthermore potential for bias related from recruitment from a single center and specialized cancer center. Although the homogenous population and technique used was important in order to demonstrate feasibility and essential for comparison of performance of single energy versus multi-energy VMIs, potential future implementation as a clinical assistant tool will require validation (with additional analysis such as normalization or image pre-processing) across scanners and institutions. We used a vendor-specific commercially available software designed for rapid-kVp-switching DECT, although VMIs at different energies can be generated using DECT systems from different vendors and therefore the multi-energy approach can be implemented in a vendor neutral manner.

5. Conclusions

In conclusion, our investigation demonstrates that DECT texture analysis with machine learning shows a high accuracy for identification and characterization of lymphadenopathy in the neck, laying the foundation for future larger and ideally prospective studies further evaluating this application using both DECT and single energy CT scans.

Funding Information

This work was partly supported by a grant by the Rossy Cancer Network, Montreal, Quebec, Canada. R.F. is a clinical research scholar (chercheur-boursier clinicien) supported by the FRQS (Fonds de recherche en santé du Québec), Montreal, Quebec, Canada.

Declaration of Competing Interest

R.F. has acted as consultant and speaker for GE Healthcare and is a founding partner and stockholder of 4Intel Inc. J.L.W. has received speaker's fees from Siemens Healthcare and GE Healthcare. The other authors have no conflicts of interest to declare.

Acknowledgements

This work was partly supported by a grant by the Rossy Cancer Network. R.F. is a clinical research scholar (chercheur-boursier clinicien) supported by the FRQS (Fonds de recherche en santé du Québec).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.csbj.2019.07.004>.

References

- Som PM, Brandwein-Gensler MS. In: Som PM, Curtin HD, editors. Lymph nodes of the neck, in head and neck imaging. St. Louis, Mo: Mosby; 2011.
- Liao LJ, et al. Analysis of sentinel node biopsy combined with other diagnostic tools in staging cN0 head and neck cancer: a diagnostic meta-analysis. *Head Neck* 2016; 38(4):628–34.
- AJCC Cancer staging manual. 8th ed.. New York; London: Springer International Publishing; 2017.
- Kostakoglu L. PET/CT imaging. In: Som PM, Curtin HD, editors. Head and neck imaging. St. Louis: Mosby Elsevier; 2011.
- Escott EJ. Role of positron emission tomography/computed tomography (PET/CT) in head and neck cancer. *Radiol Clin North Am* 2013;51(5):881–93.
- Forghani R, et al. Imaging evaluation of lymphadenopathy and patterns of lymph node spread in head and neck cancer. *Expert Rev Anticancer Ther* 2014;1–18.
- van den Brekel MW, et al. Cervical lymph node metastasis: assessment of radiologic criteria. *Radiology* 1990;177(2):379–84.
- King AD, et al. Necrosis in metastatic neck nodes: diagnostic accuracy of CT, MR imaging, and US. *Radiology* 2004;230(3):720–6.
- McCollough CH, et al. Dual- and multi-energy CT: principles, technical approaches, and clinical applications. *Radiology* 2015;276(3):637–53.
- Forghani R, De Man B, Gupta R. Dual-energy computed tomography: physical principles, approaches to scanning, usage, and implementation: part 1. *Neuroimaging Clin N Am* 2017;27(3):371–84.
- Forghani R, De Man B, Gupta R. Dual-energy computed tomography: physical principles, approaches to scanning, usage, and implementation: part 2. *Neuroimaging Clin N Am* 2017;27(3):385–400.
- Tawfik AM, et al. Comparison of dual-energy CT-derived iodine content and iodine overlay of normal, inflammatory and metastatic squamous cell carcinoma cervical lymph nodes. *Eur Radiol* 2014;24(3):574–80.
- Liu X, et al. Papillary thyroid Cancer: dual-energy spectral CT quantitative parameters for preoperative diagnosis of metastasis to the cervical lymph nodes. *Radiology* 2015;275(1):167–76.
- Liang H, et al. A retrospective study of dual-energy CT for clinical detecting of metastatic cervical lymph nodes in laryngeal and hypopharyngeal squamous cell carcinoma. *Acta Otolaryngol* 2015;135(7):722–8.
- Yang L, et al. Differentiation of malignant cervical lymphadenopathy by dual-energy CT: a preliminary analysis. *Sci Rep* 2016;6:31020.
- Zhao Y, et al. Preliminary study on the diagnostic value of single-source dual-energy CT in diagnosing cervical lymph node metastasis of thyroid carcinoma. *J Thorac Dis* 2017;9(11):4758–66.
- Tawfik AM, Michael Bucher A, Vogl TJ. Dual-energy computed tomography applications for the evaluation of cervical lymphadenopathy. *Neuroimaging Clin N Am* 2017;27(3):461–8.
- Forghani R, Srinivasan A, Forghani B. Advanced tissue characterization and texture analysis using dual-energy computed tomography. *Horiz Emerg Appl Neuroimaging Clin N Am* 2017;27(3):533–46.
- Al Ajmi E, 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.
- Zhang H, 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.
- Aerts HJ, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014;5:4006.
- Parmar C, et al. Machine learning methods for quantitative radiomic biomarkers. *Sci Rep* 2015;5:13087.
- Dang M, et al. MRI texture analysis predicts p53 status in head and neck squamous cell carcinoma. *AJNR Am J Neuroradiol* 2015;36(1):166–70.
- Buch K, 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.
- Leijenaar RT, et al. External validation of a prognostic CT-based radiomic signature in oropharyngeal squamous cell carcinoma. *Acta Oncol* 2015;54(1429):1423–9.
- Liu J, et al. Use of texture analysis based on contrast-enhanced MRI to predict treatment response to chemoradiotherapy in nasopharyngeal carcinoma. *J Magn Reson Imaging* 2016;44(2):445–55.
- Vallieres M, et al. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Phys Med Biol* 2015;60(14):5471–96.
- Bayanati H, et al. Quantitative CT texture and shape analysis: can it differentiate benign and malignant mediastinal lymph nodes in patients with primary lung cancer? *Eur Radiol* 2015;25(2):480–7.
- Andersen MB, et al. CT texture analysis can help differentiate between malignant and benign lymph nodes in the mediastinum in patients suspected for lung cancer. *Acta Radiol* 2016;57(6):669–76.
- Gao X, et al. The method and efficacy of support vector machine classifiers based on texture features and multi-resolution histogram from (18)F-FDG PET-CT images for the evaluation of mediastinal lymph nodes in patients with lung cancer. *Eur J Radiol* 2015;84(2):312–7.
- Forghani R, et al. Different spectral hounsfield unit curve and high-energy virtual monochromatic image characteristics of squamous cell carcinoma compared with nonossified thyroid cartilage. *AJNR Am J Neuroradiol* 2015;36(6):1194–200.
- Lam S, et al. Optimal virtual monochromatic images for evaluation of Normal tissues and head and neck cancer using dual-energy CT. *AJNR Am J Neuroradiol* 2015;36(8):1518–24.
- Srinivasan A, et al. Differentiation of benign and malignant neck pathologies: preliminary experience using spectral computed tomography. *J Comput Assist Tomogr* 2013;37(5):666–72.
- Ueno Y, et al. Endometrial carcinoma: MR imaging-based texture model for preoperative risk stratification—a preliminary analysis. *Radiology* 2017;161950.
- Parikh J, et al. Changes in primary breast cancer heterogeneity may augment midtreatment MR imaging assessment of response to neoadjuvant chemotherapy. *Radiology* 2014;272(1):100–12.
- De Cecco CN, et al. Texture analysis as imaging biomarker of tumoral response to neoadjuvant chemoradiotherapy in rectal cancer patients studied with 3-T magnetic resonance. *Invest Radiol* 2015;50(4):239–45.
- Matsumoto K, et al. Virtual monochromatic spectral imaging with fast kilovoltage switching: improved image quality as compared with that obtained with conventional 120-kVp CT. *Radiology* 2011;259(1):257–62.
- Forghani R, et al. Low-energy virtual monochromatic dual-energy computed tomography images for the evaluation of head and neck squamous cell carcinoma: a study of tumor visibility compared with single-energy computed tomography and user acceptance. *J Comput Assist Tomogr* 2017;41(4):565–71.
- Forghani R. Advanced dual-energy CT for head and neck cancer imaging. *Expert Rev Anticancer Ther* 2015;1–13.
- Lam S, et al. Multiparametric evaluation of head and neck squamous cell carcinoma using a single-source dual-energy CT with fast kVp switching: state of the art. *Cancers (Basel)* 2015;7(4):2201–16.
- Forghani R, Kelly HR, Curtin HD. Applications of dual-energy computed tomography for the evaluation of head and neck squamous cell carcinoma. *Neuroimaging Clin N Am* 2017;27(3):445–59.
- Breiman L. Random Forests. *Mach Learn* 2001;45(1):5–32.
- Friedman JH. Greedy function approximation: a Gradient Boosting Machine. *Ann Stat* 2001;29(5):1189–232.
- Friedman JH. Stochastic gradient boosting. *Comput Stat Data Anal* 2002;38(4):367–78.
- Hastie T, Tibshirani R, Friedman J. The elements of statistical learning data mining, inference, and prediction. . 2nd ed.Springer series in statistics. Springer Science +Business Media, LLC; 2009.
- Liaw A, Wiener M. Classification and regression by randomForest. *R News* 2002;2(3):18–22.
- Others, G.R.w.c.f. gbm. Generalized Boosted Regression Models R package version 2.1.3 ; 2017.
- Kuno H, et al. CT texture analysis of cervical lymph nodes on contrast-enhanced [(18)F] FDG-PET/CT images to differentiate nodal metastases from reactive lymphadenopathy in HIV-positive patients with head and neck squamous cell carcinoma. *AJNR Am J Neuroradiol* 2019;40(3):543–50.
- Forghani R, 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 Apr 12 2019* [Epub ahead of print].
- Foncubierta-Rodriguez A, et al. Benefits of texture analysis of dual energy CT for computer-aided pulmonary embolism detection. *Conf Proc IEEE Eng Med Biol Soc* 2013;2013:3973–6.
- Oldan J, et al. Pilot study: evaluation of dual-energy computed tomography measurement strategies for positron emission tomography correlation in pancreatic adenocarcinoma. *J Digit Imaging* 2014;27(6):824–32.
- Depeursinge A, et al. Rotation-covariant texture analysis of 4D dual-energy CT as an indicator of local pulmonary perfusion. *Biomedical imaging (ISBI), 2013 IEEE 10th international symposium on; 2013* [San Francisco, CA, USA].