Neural Networks for Nodal Staging of Non–Small Cell Lung Cancer with FDG PET and CT: Importance of Combining Uptake Values and Sizes of Nodes and Primary Tumor¹

Lauren K. Toney, MD, MS
Hubert J. Vesselle, MD, PhD

Purpose:

To evaluate the effect of adding lymph node size to three previously explored artificial neural network (ANN) input parameters (primary tumor maximum standardized uptake value or tumor uptake, tumor size, and nodal uptake at N1, N2, and N3 stations) in the structure of the ANN. The goal was to allow the resulting ANN structure to relate lymph node uptake for size to primary tumor uptake for size in the determination of the status of nodes as human readers do.

Materials and Methods:

This prospective study was approved by the institutional review board, and informed consent was obtained from all participants. The authors developed a back-propagation ANN with one hidden layer and eight processing units. The data set used to train the network included node and tumor size and uptake from 133 patients with non–small cell lung cancer with surgically proved N status. Statistical analysis was performed with the paired *t* test.

Results:

The ANN correctly predicted the N stage in 99.2% of cases, compared with 72.4% for the expert reader ($P < .001$). In categorization of N0 and N1 versus N2 and N3 disease, the ANN performed with 99.2% accuracy versus 92.2% for the expert reader ($P < .001$).

Conclusion:

The ANN is 99.2% accurate in predicting surgical-pathologic nodal status with use of four fluorine 18 fluorodeoxyglucose (FDG) positron emission tomography (PET)/computed tomography (CT)–derived parameters. Malignant and benign inflammatory lymph nodes have overlapping appearances at FDG PET/CT but can be differentiated by ANNs when the crucial input of node size is used.

©RSNA, 2013

Online supplemental material is available for this article.

¹From the Division of Nuclear Medicine, Department of Radiology, University of Washington, 1959 NE Pacific St, RR-215, UW Mailbox 357115, Seattle, WA 98195-7115. Received November 16, 2012; revision requested January 10, 2013; revision received May 22; accepted June 28; final version accepted July 19. **Address correspondence** to L.K.T. (e-mail: laurenkt@uw.edu).

Lung cancer is responsible for 156 940 deaths in the United States annually (1). Approximately 80% of lung cancers are of non–small cell histology, for which surgical resection offers the best chance of curative treatment.

Accurate staging of non–small cell lung cancer (NSCLC) is essential because it is the most important factor determining prognosis, management, and operability to prevent denial of therapy with curative intent (overstaging) and avoid subsequent morbidity, time, and cost of ineffective therapies—particularly unnecessary thoracotomy (understaging) (2). Although valuable in lung cancer detection (3,4), computed tomography (CT) is limited in the initial staging of NSCLC (5), especially in staging the mediastinum because it informs of nodal spread of tumor by demonstrating lymph node enlargement, which itself can also be caused by benign inflammation.

Fluorine 18 fluorodeoxyglucose (FDG) positron emission tomography (PET) has been shown to be more accurate than CT in both staging of the mediastinum and detecting distant metastases (2,6–20) and is increasingly incorporated into the clinical staging of NSCLC. Although its accuracy is much superior to that of CT in nodal staging, there still remains some overlap in the appearance of malignant and benign lymph nodes at FDG PET because (a) some inflammatory lymph nodes may have mild uptake of FDG and (b) small cancerous nodes may not appear very metabolically active (mildly FDG avid) owing to partial volume effects resulting from the limited reconstructed resolution of a PET scanner. Several authors

have attempted to further differentiate inflammatory from cancerous nodes on the basis of an FDG uptake threshold (standardized uptake value [SUV] threshold) (21–23); however, no single threshold can apply to all types of scanners, image reconstruction algorithms, or clinical settings and no prospective multicenter trial has validated any such threshold.

Nonetheless, it is recognized that experienced readers of FDG PET/CT scans have a keen ability to differentiate benign from malignant lymph nodes when staging lung cancer. They do this by considering all of the imaging features of lymph nodes relative to those of the primary cancer, namely the location, FDG uptake level, and size of nodes relative to location, uptake, and size of the primary cancer. These interpretive skills are based on clinical reader experience. Artificial neural networks (ANNs) have been used to emulate the accuracy of an expert FDG PET/CT reader while overcoming some of the subjectivity still present in the interpretation of PET scans (24). An ANN has previously been trained by using uptake and size of the primary tumor together with uptake of lymph nodes against surgical-pathologic proof of lymph node status. The accuracy of the resulting ANN in the determination of the nodal stage of a patient with NSCLC was superior to that of an expert reader (24). We performed the present study to evaluate the effect of

adding lymph node size to three previously explored ANN input parameters (primary tumor maximum SUV or tumor uptake, tumor size, and nodal uptake at N1, N2, and N3 stations) in the structure of the ANN. The goal was to allow the resulting ANN structure to relate lymph node uptake for size to primary tumor uptake for size in the determination of the status of nodes as human readers do.

Materials and Methods

Patient Population

The patient population was prospectively recruited with approval of the institutional review board in a study of the prognostic significance of FDG uptake level in primary NSCLC between February 1998 and September 2002 (24). This study was funded in part by National Institutes of Health grant 1R01 CA80907. No industry support was given for this study. The authors had control of the data and the information submitted for publication.

The data set includes 133 patients with NSCLC with surgically proved nodal status (N status). The mean patient age ($\pm$ standard deviation) was 64.4 years $\pm$ 9.0, and there were 47 women (mean age, 62.0 years $\pm$ 8.5) and 86 men

Advance in Knowledge

- Artificial neural networks (ANNs) can be leveraged in computer-aided diagnosis to provide more accurate and consistent assessment of nodal stage in patients with non–small cell lung cancer (NSCLC), yielding 99.2% accuracy compared with 72.4% accuracy for the expert reader ($P < .001$).

Implications for Patient Care

- An ANN is a software tool that can be implemented online to provide computer-aided diagnosis to radiologists and trainees at different institutions, which could in turn eventually improve the accuracy of noninvasive staging.
- ANNs may provide a method for teaching reliable interpretation of nodal fluorine 18 fluorodeoxyglucose positron emission tomography findings in NSCLC and would enhance the accuracy of interpretation by drawing the attention of interpreters to features potentially important but dismissed.

Published online before print

10.1148/radiol.13122427 Content code: CH

Radiology 2014; 270:91–98

Abbreviations:

ANN = artificial neural network
FDG = fluorine 18 fluorodeoxyglucose
NSCLC = non–small cell lung cancer
SUV = standardized uptake value

Author contributions:

Guarantors of integrity of entire study, L.K.T., H.J.V.; study concepts/study design or data acquisition or data analysis/interpretation, L.K.T., H.J.V.; manuscript drafting or manuscript revision for important intellectual content, L.K.T., H.J.V.; manuscript final version approval, L.K.T., H.J.V.; literature research, L.K.T.; clinical studies, H.J.V.; statistical analysis, L.K.T.; and manuscript editing, L.K.T., H.J.V.

Funding:

This study was funded in part by National Institutes of Health grant 1R01 CA80907.

Conflicts of interest are listed at the end of this article.

(mean age, 65.7 years $\pm$ 9.0). Patients were excluded if they had pleural implants ($n = 5$) or evidence of stage IV disease ($n = 40$) according to the 6th edition of the TNM staging system (25) (distant metastases or separate tumor nodule[s] in different lobe [ipsilateral or contralateral] at FDG PET and confirmed with additional anatomic imaging or percutaneous biopsy). The T and N status of the entire patient population is shown in Table 1. The 133 eligible patients had undergone surgical staging with bronchoscopy and mediastinoscopy as described previously (19). Although FDG PET/CT information was available to guide the invasive staging, the surgical confirmation of nodal status was always performed by using a systematic approach (24) that is independent of nodal size or nodal FDG uptake level at PET.

PET Protocol

All PET studies were performed with a dedicated whole-body PET tomograph (PET Advance; GE Medical Systems, Milwaukee, Wis), with standardized patient preparation (including 12-hour fast, blood glucose level check, and uniform time from FDG injection to imaging) and image acquisition as previously described (19,24).

Data Extraction

Data were collected from the FDG PET scan and accompanying diagnostic CT scan from each patient for the input layer of the neural network. Collected parameters included size and maximal SUV for both the primary tumor and the most metabolically active lymph node within each nodal station (N1, N2, and N3). The nodal stations were defined according to the 6th edition of the TNM staging system (25), and lymph node size was measured as the maximum short-axis diameter. Mean lung background SUV was obtained and used to correct the primary tumor maximum SUV for the partial volume effect as previously described (26). The background lung and mediastinal uptake were used to normalize N1 and N2–N3 nodal uptake, respectively, by division as previously described (24). Normalization by subtraction has also

Table 1

T and N Status of the Entire Patient Population ($n = 133$)

T Status	N Status				Total
	N0	N1	N2	N3	
T1	24	7	5	0	36 (27.1)
T2	30	15	19	5	69 (51.9)
T3	5	3	4	2	14 (10.5)
T4	7	1	5	1	14 (10.5)
Total	66 (49.6)	26 (19.5)	33 (24.8)	8 (6.0)	133

Note.—T and N status were confirmed by means of invasive staging according to the sixth edition of the TNM system (25). Data are numbers of patients, with percentages in parentheses.

been tested in an ANN model (24), but accuracies were lower; therefore, this method was not considered in the current model.

In the absence of uptake at the N1, N2, or N3 stations, the corresponding background uptake value was used for that N station (mean lung background for the N1 station and mean mediastinal background for N2 and N3 stations). FDG PET/CT scans were read on a dedicated workstation by the same experienced reader (H.J.V., with 10 years of experience) who extracted SUV measurements and had access to the patient's thoracic CT scans. The expert reader was blinded to the N stage because the FDG PET/CT scans were read prospectively before the patients went to the operating room for staging.

Neural Networks

For a more detailed discussion of the function, construction, and theory of ANNs, please see Appendix E1 (online). In brief, an ANN is a prediction tool that can be used in decision processes that require the assimilation of a large number of inputs to arrive at a single output. This tool can be useful as a check for experienced readers or as a guide for the less experienced, toward the goal of reducing subjectivity in scan interpretation. On a deeper level, ANNs may serve to distill the concrete input factors and the less well-defined “intangible” imaging features that are synthesized in an “expert” scan interpretation. Using R (R Core Team; R Foundation for Statistical Computing, Vienna,

Austria), an open source statistical programming language, we developed a feed-forward multilayer ANN using supervised learning by means of the back-propagation algorithm for training. The feed-forward type of ANN refers to a type of network whereby logical units of one layer may only communicate with the subsequent layer. Our network contained three layers: an input layer, a hidden layer containing eight perceptrons, and a final output layer. A “skip-layer” connection between the input and output layers (bypassing the hidden layer) was also used. All analyses were performed with a Macintosh computer (Apple Computer, Cupertino, Calif) equipped with a 2.26-GHz processor and 2 GB of RAM.

The input data set included the primary tumor size and maximum SUV (the partial volume-corrected maximum SUV and uncorrected maximum SUV were both separately tested) and nodal uptake (maximum SUV) at stations N1, N2, and N3 normalized to background as discussed earlier. Node size is the remaining input parameter and corresponded to the size of the most metabolically active node at each nodal station, even if the size was less than 1 cm.

On the basis of the principle of cross validation (27), a neural network could be both built and tested with one set of commands. Cross validation is a process whereby a network is trained and tested on disjoint subsets of the data so that overtraining, which occurs when a network memorizes rather than

ANN training and testing protocol. In brief, the steps consist of randomly partitioning data set into two equal-sized groups called train and test, training the ANN on the former and testing it on the latter. At each of 100 iterations of this process, weights of the ANN that were obtained at training session are retained for the corresponding testing session. The output of the network is compared with the predictions of the expert PET reader and with true surgical-pathologic status at each iteration, and these accuracies are averaged over all iterations.

learns, can be avoided. These steps are depicted in the Figure and consist of the following: Step 1, decide on a set of input variables (input scenario); step 2, randomly partition the data set of 133 patients into two equal-size groups (train and test); step 3, train the ANN on the data set (train); step 4, while retaining the weights (w_i) of the ANN obtained at the training session, run data set test through the network with these weights; step 5, compare output of ANN on test with the true surgical-pathologic status, quantifying neural network accuracy; step 6, compare expert PET reader output on test with true surgical-pathologic status, quantifying PET reader accuracy; and step 7, repeat steps 2–5 many times (100 times) and report the average accuracy obtained in step 5.

The process of training and testing on disjoint sets of data ensures that the neural network is learning predictive patterns rather than just memorizing

specific examples. The partitioning algorithm ensured that each random partition of data into train and test groups had the same proportion of N0 (49.6%), N1 (19.5%), N2 (24.8%), and N3 (6.0%) cases as in the entire cohort of 133 cases. The network was constructed by using the nnet package in R with the following parameters: perceptrons in hidden layer, eight; maximum number of iterations, 1000; softmax (specifies a log-linear model for the perceptron logic); skip (adds a skip layer connection from input to output layer); decay, 0 (parameter used in error fitting of the weights); and rang, 0 (parameter used in error fitting of the weights).

We also compared the N status determined by the expert reader for the same testing cases to the surgical-pathologic N status to calculate the reader accuracy with FDG PET/CT. The expert reader provided an interpretation of each scan that included a nodal status (N0–N3). Direct

comparison between the ANN and FDG PET/CT reader accuracies was performed to demonstrate the benefits of using an ANN to predict the N status of a patient.

The repetitions used in step 6 were performed to ensure against being biased by a high- or low-accuracy run as well as to provide statistics with regard to the reproducibility of a network.

Statistical Analysis

For each input scenario, we trained and tested the neural network 100 different times, as described earlier. The mean accuracy and standard deviation of each input scenario was reported. Comparison of the accuracies between ANNs defined by two different scenarios was performed by using the paired *t* test.

The performance of the most accurate ANN, scenario 9, averaged over 100 iterations of training and testing on different partitions of the data set, was evaluated and compared with that of the PET reader for the same 100 partitions.

Results

We tested the accuracy of several different scenarios comprising different sets of input variables, including repeat demonstration of several of the accuracies previously reported. The scenarios tested are listed in Table 2.

We observed incremental increases in accuracy as additional input parameters were added to each scenario. We also demonstrate this systematic improvement in accuracy by taking scenario 5, which includes lymph node size, as a starting point and successively adding node uptake ratio, primary tumor size, and primary tumor maximum SUV (scenarios 6, 8, and 9).

Not shown in Table 2 is the accuracy of scenarios with partial volume-corrected primary tumor maximum SUV. In general, these accuracies were lower. For example, in scenario 9 with primary tumor partial volume-corrected maximum SUV, accuracy was only 77.8%.

The performance of the most accurate ANN, scenario 9, compared with

the PET reader as described earlier, is shown in Table 3. The mean accuracy in determining the surgical-pathologic N status was $99.2\% \pm 0.8$, compared with $72.4\% \pm 3.7$ for the expert PET reader ($P < .001$).

For the task of differentiating N0 and N1 disease from N2 and N3 disease (the diagnostic step differentiating resectable from nonresectable nodal status), again with scenario 9, the accuracy of the PET reader increased to $92.2\% \pm 2.3$, compared with the ANN's performance of $99.2\% \pm 0.8$ ($P < .001$). With a previous ANN, the differentiation of N0 and N1 disease from N2 and N3 disease is reported as 91.7% (24).

Discussion

We have demonstrated that an ANN can be used to synthesize multiple parameters from FDG PET and CT to predict NSCLC nodal status with consistently high accuracy (mean, $99.2\% \pm 0.8$) surpassing that of an expert PET reader. The task at hand is assessing the FDG uptake in lymph nodes to determine benign versus malignant status for each lymph node and nodal station, knowing that both malignant and inflammatory nodes can have increased FDG uptake and that both node types can be enlarged as well. The overlap in these appearances is what makes the interpretation of FDG PET/CT scans intrinsically challenging and why experience matters. Our goal was to demonstrate that an ANN can overcome some limitations of a human reader by using a large body of learned experience.

The first feature used in FDG PET/CT interpretation is the level of FDG uptake in a node relative to the local background, which when used alone to train an ANN yields a mean accuracy of 77%. This is slightly higher than the 71% achieved when using only the node size in the ANN as per CT interpretation criteria. This relative performance is in keeping with human observer interpretation accuracies, with FDG PET being superior to CT in the nodal status determination task (12–20). These node features can be used for interpretation

Table 2

Accuracy of the ANN in the Identification of Nodal Status (N0–N3) with Various Sets of Input Parameters (Scenarios)

Scenario No.	Input Parameters for the ANN	Accuracy (%)	
		Current Study	Previous Study*
1	(Nodal uptake/Bkg)	77.4 ± 3.6	77.3 ± 0.8
2	(Nodal uptake/Bkg) + (primary size)	75.9 ± 3.7	81.6 ± 1.7
3	(Nodal uptake/Bkg) + (primary SUV _{max})	81.1 ± 3.4	80.1 ± 1.4
4	(Nodal uptake/Bkg) + (primary SUV _{max}) + (primary size)	86.5 ± 2.9	86.1 ± 2.6
5	(Node sizes)	71.3 ± 3.8	NA
6	(Node sizes) + (nodal uptake/Bkg)	79.7 ± 3.6	NA
7	(Node sizes) + (nodal uptake/Bkg) + (primary SUV _{max})	86.6 ± 3.0	NA
8	(Node sizes) + (nodal uptake/Bkg) + (primary size)	80.5 ± 3.6	NA
9	(Node sizes) + (nodal uptake/Bkg) + (primary SUV _{max}) + (primary size)	99.2 ± 0.8	NA

Note.—Data are means $\pm$ standard deviations. Bkg = background, NA = not applicable, SUV_{max} = maximum SUV.

* Data are from reference 24.

Table 3

Concordance between PET Reader– and ANN–predicted Node Status and Surgical-Pathologic N Status

ANN/PET Reader Status	Surgical-Pathologic Node Status			
	N0	N1	N2	N3
N0	3255/3008	0/910	0/287	0/0
N1	0/256	1292/284	0/85	0/0
N2	53/44	0/98	1600/1178	0/91
N3	0/0	0/0	0/50	400/309

Note.—This table was obtained by summing together the truth tables from each of 100 iterations of the ANN training and testing cycles. The numerators represent the ANN–surgical-pathologic concordance, and the denominators represent the expert–surgical-pathologic concordance. Each iterative testing cycle is performed on 66 cases (the remainder of the 133 cases not used in training); therefore, the sum of the numerators is the total number of case iterations in the determination of the ANN accuracy ($66 \times 100 = 6600$). Likewise, the sum of denominators, 6600, is the total number of case iterations in the determination of the expert reader accuracy.

because, in general, malignant nodes tend to be more FDG avid and larger than benign nodes—although an overlap between populations exists.

Our observation of incremental improvements in accuracy as additional input parameters were added to each scenario is commensurate with previous demonstration of increasing accuracy with respect to the incremental addition of input parameters, moving from scenario 1 to scenario 3 to scenario 4 (24). When combining node size with nodal uptake an even better mean accuracy can be reached, which is consistent with reports in the

literature (28). When the maximum SUV of the primary tumor and its size to the ANN structure (scenarios 7 and 9) are added, significant additional improvements in mean accuracy can be obtained with an ANN. This is due to the fact that the ANN is able to learn to relate the uptake level of nodes to that of the primary tumor, with more FDG-avid tumors giving rise to more avid malignant nodes. Moreover, when the sizes of the primary tumor and nodes are taken into consideration, the best accuracy is reached because the ANN can learn that, for a node to be malignant, its uptake per unit volume must

be commensurate with the uptake per unit volume of the primary tumor. In other words, if a node is FDG avid because it is malignant, it should be populated by the same cancer cells as the primary tumor and the “FDG concentration” should be similar in both the malignant node and primary tumor.

One potential way to exploit the relationship between nodal and primary tumor uptake is to examine their ratio as an input to the ANN. The ratio of the maximum SUV of the mediastinal (N2) lymph node to that of the primary tumor in patients with NSCLC has been more recently evaluated with modest success (29), although Cerfolio and Bryant did not take size into account in their calculations. Given the small size of most lymph nodes relative to the reconstructed resolution of the scanner images, partial volume effects are present in maximum SUV measurements. Partial volume correction to the maximum SUVs of nodes and tumor would be necessary before determining their ratio. We believed that this preprocessing of data might detract from both the simplicity and potency of the ANN, and therefore this approach was not used.

The most common (and only) error made by the ANN was to classify incorrectly patients with N0 disease as having N2 disease (53 times in 6600 case iterations). This error was often simultaneously made by the expert PET reader (44 times in 6600 case iterations). This appears to correspond to a case in the data set with a primary tumor SUV of 6.9 with no uptake at nodal stations beyond background. Our data set contains a separate patient with nearly identical PET findings who at surgery was found to have N2 disease. The error therefore appears to be traceable to application of learning from a spurious example. If the patient had true N2 disease at the time of the PET, then this showcases an example of a false-negative finding at PET; if the patient developed N2 disease between PET and the surgical staging, this error highlights the need to minimize temporal separation of the steps in surgical staging.

The most common error made by the expert PET reader was calling a study N0 when no hypermetabolic N1 nodes were observed at PET ($n = 19$). The ANN correctly classified these cases as N1. It is not immediately clear how this was done based on the only available inputs in these situations (primary tumor size and SUV), which highlights that the ANN may be able to identify complex (possibly nonlinear) relationships between input variables. However, the clinical significance of differentiating N0 from N1 disease is of less importance because both patient groups will be treated with surgical resection.

The second most common example of PET reader error was the interpretation of any uptake within N2 or N3 nodes beyond background as indicative of the presence of N2 or N3 disease in cases in which these nodes were negative at surgical staging. The ANN identified that these nodes were not malignant, possibly by identifying lack of concordance between their uptake and the uptake of lower-station nodes and the primary tumor. The clinical relevance of this is immediately evident because the differentiation of N0 and N1 disease from N2 and N3 disease affects stage, treatment, and prognosis. The accuracy of the ANN in this task was 7% higher than that of the expert PET reader (99.2% vs 92.2%, respectively).

Our study has a number of limitations. Several relate to the dependence of a neural network on a training data set. Portability of the current ANN would therefore depend on the assumption that the current set of patients with resectable NSCLC is generalizable. What is general about the patient population on the data set is that they were enrolled prospectively for a National Institutes of Health trial and out of a thoracic surgery clinic that used FDG PET for every one of its patients being considered for surgical resection. Every patient was also treated on the basis of a strict and well-established surgical algorithm, with pathologic proof of nodal status for every case where nodal disease was the determinant of stage (no distant metastases and no pleural metastases). What is not generalizable

is the fact that this data set is dependent on the method of scanning used. FDG PET scanning protocols vary across institutions, leading to variations in SUVs extracted from scans for the purpose of training an ANN. In particular, the time from FDG injection to imaging, the type of image reconstruction algorithm, the reconstruction filter, the scan length, and the attenuation correction method used will affect SUVs used as inputs to the ANN. However, what would remain generalizable are the architecture design of the network and the general conclusions drawn from it as to the relative importance of the different variables in their effect on the determination of nodal stage.

Any institution wishing to implement a similar ANN could do so by using our ANN architecture and training it with its own data set extracted from a series of FDG PET scans obtained in prospectively enrolled patients with NSCLC before surgery.

It is also possible that the distribution of N status in our population, in particular the small number of cases with N3 disease ($n = 8$), may uniquely affect the training process. However, this case distribution was representative of that of a large-volume academic lung cancer surgery practice, where cases with bulky adenopathy have a higher incidence of distant metastases identified on diagnostic CT scans. Therefore, these cases never get to the point of being considered for surgery and, in our institution, were not scanned with FDG PET (because diagnostic CT had already established that these patients had stage IV disease).

Another potential limitation relates to the update of new TNM staging criteria guidelines in the seventh edition of the TNM system for lung cancer recommended by the staging committee of the International Association for the Study of Lung Cancer in 2009 (30). This update has occurred between the time of study enrollment and data collection and that of the current analysis. The new criteria should not have affected study enrollment because pleural disease and distant metastases would have been excluded under both sets of

staging criteria. Moreover, the updated staging system did not modify the definition of the nodal stages.

The reader may note that our study design did not include an evaluation of interobserver variability. Our FDG PET/CT interpretations were performed by one reader with 10 years of experience, and therefore interobserver variability analysis was not applicable. The purpose of this study was to demonstrate the value of sequentially adding parameters to a neural network by extending the results of a previous study to show the importance of nodal size and uptake in the analysis. This was accomplished by comparing neural network accuracy against surgical-pathologic nodal staging, the standard of reference. A secondary comparison of neural network accuracy with that of one particular expert reader served as an internal reference to highlight the advantage of the neural network. However, it was not our intention to evaluate the accuracy of expert-interpreted FDG PET/CT in staging NSCLC because these data are widely available in the literature (2,6–20). Hence, interobserver variability was not believed to be necessary for this study design.

Finally, there is dependence of the neural network on the particular software platform used and, subsequently, the parameters chosen within the functions used. Our chosen software, R, is open source. For example, we chose to use a skip-layer neural network format, with connections directly between the inputs with the outputs. The skip layer represents traditional linear regression and was included in our model because it made the model more stable across other parameter changes and had overall higher rates of accuracy. Not surprisingly, this could reflect some degree of linear relationship between input and output variables.

Another programming choice was the requirement that both the training and test subsets have the same proportion of N0, N1, N2, and N3 cases. This was easy to implement and would be reproducible if the network was assessed (trained and tested) on an independent data set. We believed that this construct gave the ANN

the best chance to learn from a diversity of cases, in particular allowing it to be exposed to the few N3 cases. This may potentially exaggerate the performance of the network in the sense that the network would likely not perform as well if this step were not taken.

Our current study revalidated the results previously published for ANN architectures described in scenarios 1–4 by achieving similar accuracies with a different software platform. Small differences in mean accuracies between current and previous studies are likely due to differences within the software packages (eg, convergence algorithms and/or criterion, threshold functions, etc). The neural network is superior to an expert reader because it can learn from many cases and keep this entire learned experience to weigh any new case against it. On the contrary, an experienced reader will have turned this experience into a “gestalt” or set of rules but will not be able to actually recall all of the previously read cases for a comparative weighted interpretation of a new case. Our ANN far exceeds the performance of an expert reader in assessing the nodal status of a patient. As such, the process of building a reliably accurate ANN has illuminated the relative importance of different imaging parameters in PET/CT of patients undergoing mediastinal staging for NSCLC.

Disclosures of Conflicts of Interest: L.K.T. No relevant conflicts of interest to disclose. H.J.V. No relevant conflicts of interest to disclose.

References

1. American Cancer Society. Cancer facts & figures 2012. Atlanta, Ga: American Cancer Society, 2012.
2. Maziak DE, Darling GE, Incullet RI, et al. Positron emission tomography in staging early lung cancer: a randomized trial. *Ann Intern Med* 2009;151(4):221–228, W-48.
3. National Lung Screening Trial Research Team, Aberle DR, Adams AM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011;365(5):395–409.
4. Sox HC. Better evidence about screening for lung cancer. *N Engl J Med* 2011;365(5):455–457.
5. Detterbeck FC, Jantz MA, Wallace M, Vansteenkiste J, Silvestri GA. Invasive mediastinal staging of lung cancer: ACCP evidence-based clinical practice guidelines (2nd edition). *Chest* 2007;132(3 Suppl):202S–220S.
6. Chin R Jr, Ward R, Keyes JW, et al. Mediastinal staging of non-small-cell lung cancer with positron emission tomography. *Am J Respir Crit Care Med* 1995;152(6 Pt 1):2090–2096.
7. Kernstine KH, Stanford W, Mullan BF, et al. PET, CT, and MRI with Combidex for mediastinal staging in non-small cell lung carcinoma. *Ann Thorac Surg* 1999;68(3):1022–1028.
8. Pieterman RM, van Putten JW, Meuzelaar JJ, et al. Preoperative staging of non-small-cell lung cancer with positron-emission tomography. *N Engl J Med* 2000;343(4):254–261.
9. De Wever W, Stroobants S, Coolen J, Verschakelen JA. Integrated PET/CT in the staging of nonsmall cell lung cancer: technical aspects and clinical integration. *Eur Respir J* 2009;33(1):201–212.
10. Kernstine KH, Trapp JF, Croft DR. Comparison of positron emission tomography (PET) and computed tomography (CT) to identify N2 and N3 disease in non small cell lung cancer (NSCLC). *J Clin Oncol* 1998;17(suppl):459.
11. Coleman RE. PET in lung cancer staging. *Q J Nucl Med* 2001;45(3):231–234.
12. Gupta NC, Tamim WJ, Graeber GG, Bishop HA, Hobbs GR. Mediastinal lymph node sampling following positron emission tomography with fluorodeoxyglucose imaging in lung cancer staging. *Chest* 2001;120(2):521–527.
13. Hellwig D, Ukena D, Paulsen F, Bamberg M, Kirsch CM; Onko-PET der Deutschen Gesellschaft für Nuklearmedizin. Meta-analysis of the efficacy of positron emission tomography with F-18-fluorodeoxyglucose in lung tumors: basis for discussion of the German Consensus Conference on PET in Oncology 2000 [in German]. *Pneumologie* 2001;55(8):367–377.
14. Lloyd C, Silvestri GA. Mediastinal staging of non-small-cell lung cancer. *Cancer Control* 2001;8(4):311–317.
15. Jeong HJ, Chung JK, Kim YK, et al. Usefulness of whole-body (18)F-FDG PET in patients with suspected metastatic brain tumors. *J Nucl Med* 2002;43(11):1432–1437.
16. Yun M, Kim W, Alnafisi N, Lacorte L, Jang S, Alavi A. 18F-FDG PET in characterizing

- adrenal lesions detected on CT or MRI. *J Nucl Med* 2001;42(12):1795–1799.
17. Kernstine KH, McLaughlin KA, Menda Y, et al. Can FDG-PET reduce the need for mediastinoscopy in potentially resectable nonsmall cell lung cancer? *Ann Thorac Surg* 2002;73(2):394–401; discussion 401–402.
 18. Kiernan PD, Sheridan MJ, Lamberti J, et al. Mediastinal staging of non–small cell lung carcinoma using computed and positron-emission tomography. *South Med J* 2002;95(10):1168–1172.
 19. Vesselle H, Pugsley JM, Vallières E, Wood DE. The impact of fluorodeoxyglucose F 18 positron-emission tomography on the surgical staging of non–small cell lung cancer. *J Thorac Cardiovasc Surg* 2002;124(3):511–519.
 20. von Haag DW, Follette DM, Roberts PF, Shelton D, Segel LD, Taylor TM. Advantages of positron emission tomography over computed tomography in mediastinal staging of non–small cell lung cancer. *J Surg Res* 2002;103(2):160–164.
 21. Vansteenkiste JF, Stroobants SG, De Leyn PR, et al. Lymph node staging in non–small-cell lung cancer with FDG-PET scan: a prospective study on 690 lymph node stations from 68 patients. *J Clin Oncol* 1998;16(6):2142–2149.
 22. Hellwig D, Graeter TP, Ukena D, et al. 18F-FDG PET for mediastinal staging of lung cancer: which SUV threshold makes sense? *J Nucl Med* 2007;48(11):1761–1766.
 23. Bryant AS, Cerfolio RJ. The maximum standardized uptake values on integrated FDG-PET/CT is useful in differentiating benign from malignant pulmonary nodules. *Ann Thorac Surg* 2006;82(3):1016–1020.
 24. Vesselle H, Turcotte E, Wiens L, Haynor D. Application of a neural network to improve nodal staging accuracy with 18F-FDG PET in non–small cell lung cancer. *J Nucl Med* 2003;44(12):1918–1926.
 25. Greene FL, Page DL, Fleming ID, et al, eds. *AJCC cancer staging manual*. 6th ed. New York, NY: Springer-Verlag, 2002; 435.
 26. Vesselle H, Schmidt RA, Pugsley JM, et al. Lung cancer proliferation correlates with [F-18]fluorodeoxyglucose uptake by positron emission tomography. *Clin Cancer Res* 2000;6(10):3837–3844.
 27. McClelland JL, Rumelhard DE. *Parallel distributed processing*. Cambridge, Mass: MIT Press, 1986.
 28. de Langen AJ, Raijmakers P, Riphagen I, Paul MA, Hoekstra OS. The size of mediastinal lymph nodes and its relation with metastatic involvement: a meta-analysis. *Eur J Cardiothorac Surg* 2006;29(1):26–29.
 29. Cerfolio RJ, Bryant AS. Ratio of the maximum standardized uptake value on FDG-PET of the mediastinal (N2) lymph nodes to the primary tumor may be a universal predictor of nodal malignancy in patients with nonsmall-cell lung cancer. *Ann Thorac Surg* 2007;83(5):1826–1829; discussion 1829–1830.
 30. Goldstraw P. The 7th edition of TNM in lung cancer: what now? *J Thorac Oncol* 2009; 4(6):671–673.