Clinical Evaluation of a Novel Deep Learning Algorithm for Automated Lymph Node Detection in Chest CT

Alex Johnson, MD[1]; Josh Rampton, MD[2]; Andrew Miller, MD[2]; John Miller, MD[2]
University of North Dakota, Grand Forks, ND[1]; Sanford Medical Center, Bismarck, ND[2]

Research Question

The objective of this study was to clinically evaluate an AI-based algorithm designed for the detection and segmentation of hilar and mediastinal pathologic lymph nodes or nodal conglomerates using a lymph node station specific recall method.

Background and Significance

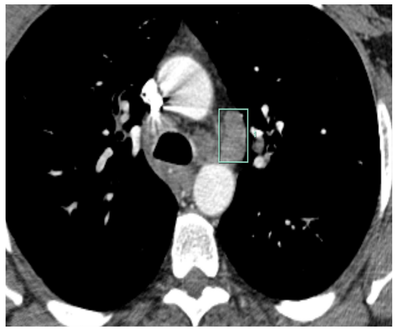
The accurate identification of mediastinal lymph nodes on computed tomography (CT) is critical for the evaluation of chest malignancies. Lymphatic spread plays a key role in staging many cancers; however, in the mediastinum, challenges arise due to low tissue contrast and variations in lymph node shape, size, and location. These factors have hindered the development of reliable software that can assist radiologists in accurately identifying suspicious lymph nodes [1,2]. With the advent of new artificial intelligence (AI) architectures, there is growing potential for the identification of malignant lymph nodes [3].

Design, Framework

The system integrates a RetinaNet detection algorithm [4] using ResNet50 as the backbone, and a UNet segmentation algorithm [5]. The system was trained using 1600 chest CT scans using a classic 70/20/10 split with all datasets annotated by board-certified radiologists.

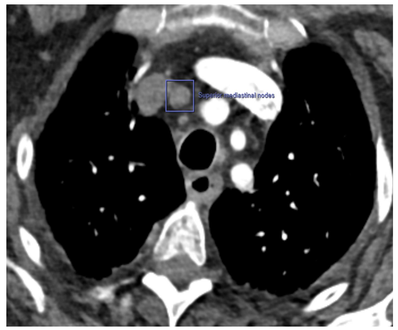
A lymph node station-wise recall was used to evaluate the model's ability to detect adenopathy in specific lymph node stations, rather than at the patient level compared to other authors[6]. This metric counts a false negative whenever the model fails to detect adenopathy in a particular lymph node station, providing a more detailed and clinically relevant analysis. By focusing on station-wise performance, we ensure that missed detections are appropriately captured, which is essential for accurate staging and treatment planning. Ground truth for the study was established by a panel of three board-certified radiologists with more than 39 years of combined clinical chest CT experience.

Figure 1



True Positive – Concordant mark (green box) of an enlarged level V (AP window) node measuring 14 mm in short axis. Human and machine agreement.

Figure 2



False Negative – Axial CT image demonstrating enlarged 2R lymph node (blue box) measuring 9.5 mm in short axis. This lesion was detected by human readers, not the algorithm.

References

1. Walker CM, Chung JH, Abbott GF, Little BP, El-Sherief AH, Shepard JAO, et al. Mediastinal lymph node staging: from noninvasive to surgical. AJR. 2012;199–64.
2. Roth HR, et al. A new 2.5D representation for lymph node detection using random sets of deep convolutional neural network observations. In: Golland P, Hata N, Barillot C, Hornegger J, Howe R, editors. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014. Lecture Notes in Computer Science, vol 8673. Springer, Cham; 2014.
3. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med Image Anal. 2017;42:60–88..
4. Lin TY, Goyal P, Girshick R, He K, Dollar P. Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV); 2017:2980-2988.
5. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention; 2015:234-241.

Methods

The AI-based algorithm was clinically evaluated using a retrospective cohort of 211 pathology rich chest CT scans with all datasets annotated by board-certified radiologists. A positive lymph node was defined as having a short-axis diameter of ≥9 mm. A lymph node station specific recall was used. Therefore, a false negative could be generated in a patient who had nodes detected in other stations not seen by AI. This is proposed as a more effective clinical form of attribution compared to patient wise recall or single node accounting.

Results

Out of 211 cases, 25 were negative for enlarged lymph nodes. The AI system yielded 378 true positive findings and 28 false negatives. There were 46 false positive findings, which included both small non-pathologic nodes and non-node structures incorrectly labeled by the AI. These results yielded a station-specific sensitivity of 93.1% and a false positive rate of 0.218 FP/vol.

Conclusions, Implications, and Dissemination

The results of this study indicate that the AI-based system can detect enlarged lymph nodes from chest CT scans with acceptable sensitivity and relatively lower false positivity. In contrast to previous algorithms reporting sensitivities between 52.9%, 70%, and 95.5% [7,8], this system maintains previously achieved sensitivity. The reduced false positive rate compared to other systems may in part be due to a larger training set. This eliminated the need to use cross-validation likely curbing over-fitting. Further large-scale investigations maybe be warranted.

SANFORD HEALTH

019012-060415