CrossMark

Axillary Lymph Node Evaluation Utilizing Convolutional Neural Networks Using MRI Dataset

Richard Ha [1] · Peter Chang [2] · Jenika Karcich [1] · Simukayi Mutasa [1] · Reza Fardanesh [1] · Ralph T. Wynn [1] ·
Michael Z. Liu [3] · Sachin Jambawalikar [3]

Abstract

The aim of this study is to evaluate the role of convolutional neural network (CNN) in predicting axillary lymph node metastasis, using a breast MRI dataset. An institutional review board (IRB)-approved retrospective review of our database from 1/2013 to 6/2016 identified 275 axillary lymph nodes for this study. Biopsy-proven 133 metastatic axillary lymph nodes and 142 negative control lymph nodes were identified based on benign biopsies (100) and from healthy MRI screening patients (42) with at least 3 years of negative follow-up. For each breast MRI, axillary lymph node was identified on first T1 post contrast dynamic images and underwent 3D segmentation using an open source software platform 3D Slicer. A 32 × 32 patch was then extracted from the center slice of the segmented tumor data. A CNN was designed for lymph node prediction based on each of these cropped images. The CNN consisted of seven convolutional layers and max-pooling layers with 50% dropout applied in the linear layer. In addition, data augmentation and L2 regularization were performed to limit overfitting. Training was implemented using the Adam optimizer, an algorithm for first-order gradient-based optimization of stochastic objective functions, based on adaptive estimates of lower-order moments. Code for this study was written in Python using the TensorFlow module (1.0.0). Experiments and CNN training were done on a Linux workstation with NVIDIA GTX 1070 Pascal GPU. Two class axillary lymph node metastasis prediction models were evaluated. For each lymph node, a final softmax score threshold of 0.5 was used for classification. Based on this, CNN achieved a mean five-fold cross-validation accuracy of 84.3%. It is feasible for current deep CNN architectures to be trained to predict likelihood of axillary lymph node metastasis. Larger dataset will likely improve our prediction model and can potentially be a non-invasive alternative to core needle biopsy and even sentinel lymph node evaluation.

Keywords CNN · Axillary metastasis · MRI

✉ Richard Ha
  rh2616@columbia.edu

1   Department of Radiology, 622 West 168th Street, PB-1-301, New York, NY 10032, USA

2   Department of Radiology, T32 Training Grant (NIH T32EB001631), UC San Francisco Medical Center, 505 Parnassus Ave, San Francisco, CA 94143, USA

3   Department of Medical Physics, Columbia University Medical Center, 177 Ft. Washington Ave., Milstein Bldg Room 3-124B, New York, NY 10032-3784, USA

Introduction

Axillary lymph node status is the most important prognostic factor in patients with early-stage breast cancer. Morbidities associated with axillary lymph node dissection have led to the development of sentinel lymph node biopsy (SLNB) to reduce the rate of negative axillary clearances [1, 2]. Reported sensitivity rates of intraoperative SLN evaluation for breast cancer range from 58 to 72% [3–5] and accuracy rate of 75% [6]. These rates are consistent with recently published 33% FN rate for intraoperative SLN [7].

Although SLNB is a minimally invasive procedure, it is still associated with morbidities, which include risk of lymphedema amounting 8.2% at 12 months [8]. Other complications such as seroma, localized swelling, pain and paresthesia, infectious neuropathy, decreased arm strength, and shoulder stiffness have been reported in up to 19.5% of patients with

🖄 Springer

SLNB [9]. There is potential for non-invasive imaging technique for axillary evaluation that may be comparable to SLNB without the associated comorbidities. Prior studies have investigated axillary ultrasound (AUS) and positron emission tomography-computer tomography (PET-CT) for evaluation of the axillary lymph nodes. These modalities have shown only moderate accuracy and sensitivity for detecting metastatic axillary lymph nodes, with 67–77% accuracy and 43.5–72.3% sensitivity for AUS and 81.1% accuracy and 56–62.7% sensitivity for PET-CT [10–12]. In addition, AUS is operator dependent and PET-CT involves potentially harmful ionizing radiation exposure.

Utilizing the breast MRI modality for axillary evaluation reportedly shows low intra- and inter-observer variability and higher diagnostic accuracy (71–85%) and sensitivity 47.8–89% for nodal status [12–15]. Although MRI is the most promising of the imaging modalities, previously published studies are limited by small sample size and subjective identification of the region of interest manually defined within the lymph node by the reader.

In recent years, there has been investigation into quantitative analysis of specific extracted imaging features, termed "radiomics." The field of radiomics has developed largely due to the contribution of machine learning techniques utilizing the extraction of pertinent imaging features and correlating with clinical data. Most recently, a subset of machine learning utilizing a type of artificial neural network called CCN has begun to proliferate due to advances in computer hardware technology for medical imaging analysis. In contrast to traditional algorithms which utilize hand-crafted features based on human-extracted patterns, neural networks allow the computer to automatically construct predictive statistical models, tailored to solve a specific problem subset [16]. The laborious task of human engineers inputting specific patterns to be recognized could be replaced by inputting curated data and allowing the technology to self-optimize and discriminate through increasingly complex layers.

A convolutional neural network (CNN) is a deep artificial neural network that automatically constructs predictive statistical models, tailored to solve a specific problem subset. It allows the technology to self-optimize and discriminate through increasingly complex layers [16]. The purpose of this study is to develop an objective and accurate approach to MRI axillary evaluation applying a novel CNN algorithm.

Methods

Patient Population

An institutional review board-approved retrospective review from 1/2013 to 6/2016 identified biopsy-proven 133 metastatic axillary lymph nodes on core biopsy from 133 patients, which was compliant with Health Insurance Portability and Accountability Act (HIPPA). One hundred forty-two negative control lymph nodes were identified based on benign biopsies and subsequent negative SLN evaluation in 100 patients, and from healthy 42 MRI screening patients with at least 3 years of negative follow-up.

MRI Acquisition and Analysis

MRI was performed on a 1.5-T or 3.0-T commercially available system (Signa Excite, GE Healthcare) using an eight-channel breast array coil. A bilateral sagittal T1-weighted fat-suppressed fast spoiled gradient-echo sequence (17/2.4; flip angle, 35°; bandwidth, 31–25 Hz) was then performed before and after a rapid bolus injection (gadobenate dimeglumine/Multihance; Bracco Imaging; 0.1 mmol/kg) delivered through an IV catheter. Image acquisition started after contrast material injection and was obtained consecutively with each acquisition time of 120 s. Section thickness was 2–3 mm using a matrix of 256×192 and a field of view of 18–22 cm. Frequency was in the antero-posterior direction.

Image Pre-processing

For all patients, lymph nodes were segmented by a breast fellowship trained radiologist with 8 years of experience using 3D Slicer [17] based on the first T1-W post contrast subtraction images. For each segmented lymph node, the slice with the largest cross-sectional area as determined on any orthogonal plane (axial, sagittal, or coronal) was identified. The center of mass for each 2D cross-sectional ROI was used as a landmark to create a uniform 4.0×4.0 cm bounding box around the lymph node of interest. A fixed size bounding box methodology was chosen to preserve relative size of lymph nodes from patient to patient.

All 2D images were rescaled to a 32×32 voxel resolution. The intensity values were normalized by conversion to a z score map. In addition, the ROI mask was dilated by five voxels, and every voxel outside the mask was set to a z score of -5.

Data augmentation employed by this study involves several real-time modifications to the source images at the time of training. Specifically, 50% of all images in a mini-batch were modified randomly by means of (1) addition across all pixels of a scalar between $[-0.1, 0.1]$ in order to simulate the effect of random Gaussian noise from different acquisition parameters; (2) random affine transformation of the original image, which alters each lymph node slightly utilizing a rigid transformation, essentially making the same lymph node appear as a unique input to the network. Given a two-dimensional affine matrix,

$$\begin{bmatrix} s_1 & t_1 & r_1 \\ t_2 & s_2 & r_2 \\ 0 & 0 & 1 \end{bmatrix}$$

the random affine transformation was initialized with random uniform distributions of interval s_1, $s_2 \in [0.8, 1.2]$, t_1, $t_2 \in [-0.3, 0.3]$, and r_1, $r_2 \in [-16, 16]$. These parameters were confirmed on visual inspection as applying enough of a warp to simulate a different lymph node without making the lymph node appear unrealistic. The choice to apply data augmentation to 50% of the example images was made to bias the network towards recognition of real data over augmented data.

Neural Network Architecture

Several neural network architectures were tested with varying network depths and kernel sizes, including a pretrained network architecture based on VGG-16. The final overall network architecture is shown in Figs. 1, 2, and 3. The CNN is implemented completely by series of 3 × 3 convolutional kernels to prevent overfitting [18]. No pooling layers are used; instead downsampling is implemented simply by means of a 3 × 3 convolutional kernel with stride length of 2 to decrease the feature maps by 75% in size. All non-linear functions utilize the rectified linear unit (ReLU) which allows training of deep neural networks by limiting vanishing gradients on backpropagation [19]. Additionally, batch normalization is used between the convolutional and ReLU layers to stabilizing training by limiting vanishing gradients and to prevent covariate shift [20]. Upon downsampling, the number of feature channels is doubled, reflecting increasing representational complexity and to prevent a representation bottleneck. Dropout at 50% was applied to the second to last fully connected layer to limit overfitting and add stochasticity to the training process [21].

Training was implemented using the Adam optimizer, an algorithm for first-order gradient-based optimization of stochastic objective functions, based on adaptive estimates of lower-order moments [22]. Parameters were initialized to equalize input and output variance utilizing the heuristic described by He et al. [23]. L2 regularization is implemented to prevent overfitting of data by limiting the squared magnitude of the kernel weights. To account for training dynamics, the learning rate is annealed and the mini-batch size is increased whenever training loss plateaus. Furthermore, a normalized gradient algorithm is employed to allow for locally adaptive learning rates that adjust according to changes in the input signal [22].

Due to the small sample size, five-fold cross-validation was utilized to evaluate network performance (80% training and 20% testing). This method involves initially splitting the available data into five random groupings. One of the groups is utilized as the initial testing set to fine tune the parameters of the network trained on the other five groups. After parameter tuning is complete, the group utilized as the validation set is changed and the network is retrained on the remaining four groups using the same parameters. The process is repeated until every one of the five groups of data is utilized as a validation set once.

Software code for this study was written in Python using the TensorFlow module (1.0.0). Experiments and CNN training will be done on a Linux workstation with NVIDIA GTX 1070 Pascal GPU with 8 GB on chip memory, i7 CPU and 32-GB RAM.

Results

A total of 142 metastatic lymph nodes and 133 normal lymph nodes were included in this study. For each lymph node, a final softmax score threshold of 0.5 was used for classification. Based on this, mean five-fold cross-validation accuracy was calculated at 84.3%.

Manual inspection of false positive and false negative predictions of the network revealed no discernibly consistent features that consistently lead to false negative or false positive classifications from the network.
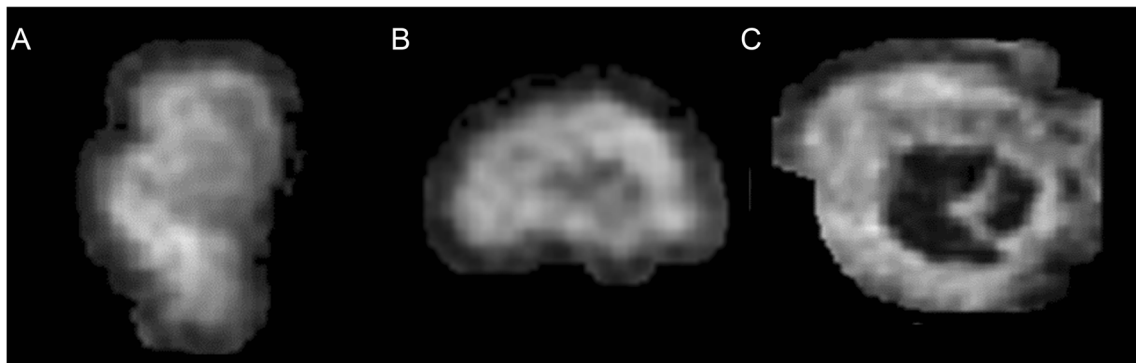

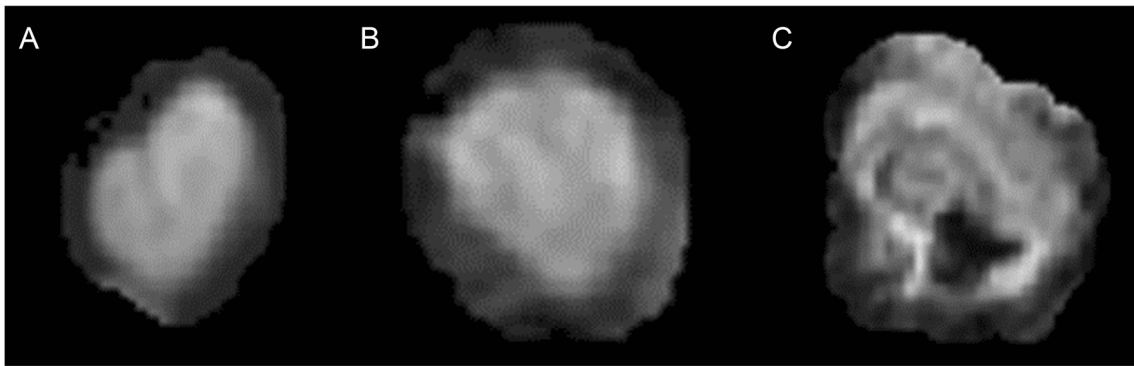Fig. 1 **a**–**c** Representative images after pre-processing of metastatic lymph nodes

Fig. 2 a–c Representative images after pre-processing of negative control lymph nodes

The CNN was trained for a total of 22,000 iterations (approximately 1500 epochs with batch sizes ranging from 12 to 24) before convergence. A single forward pass during test time for classification of new cases can be achieved in 0.0043 s.

Discussion

To our knowledge, this is the first study applying deep machine learning using CNN-based algorithm to predict axillary lymph node metastasis based on imaging data. Our study shows that it is feasible to use a CNN-based algorithm for axillary evaluation using breast MRI dataset yielding a reasonable diagnostic performance (accuracy of 84%) even with a relatively small dataset.

Prior published studies evaluating the axilla with MRI have reported an averaged accuracy rate of 75% (ranging 71–85%) in predicting axillary metastasis [13–15]. In a retrospective study, Hwang et al. analyzed performance of AUS, MRI, and PET-CT in detection of axillary lymph node metastasis (ALNM). AUS, MRI, and PET-CT had accuracies of 77.1, 77.9, and 81.1% respectively. The combination of MRI and PET-CT was most accurate with an accuracy of 83.1%. However, routine use of both MRI and PET-CT for axillary evaluation may not be cost effective.

In a retrospective analysis by Hiecken et al. [14], performance of breast MRI was assessed on both a patient-by-patient and a node-by-node analysis, which included 505 patients. Their patient pool included patients with stages T1–T4. The accuracy of MRI in detection of ALNM was 69.7–71.3%. Abe et al. [15] performed a prospective analysis of 50 patients with stages T1–T3 breast cancer, in a patient-by-patient fashion. The accuracy of MRI in detection of ALNM was 74%. Scaranelo et al. [13] evaluated prospectively the performance of MRI in evaluation of ALNM, in 61 patients. The reported
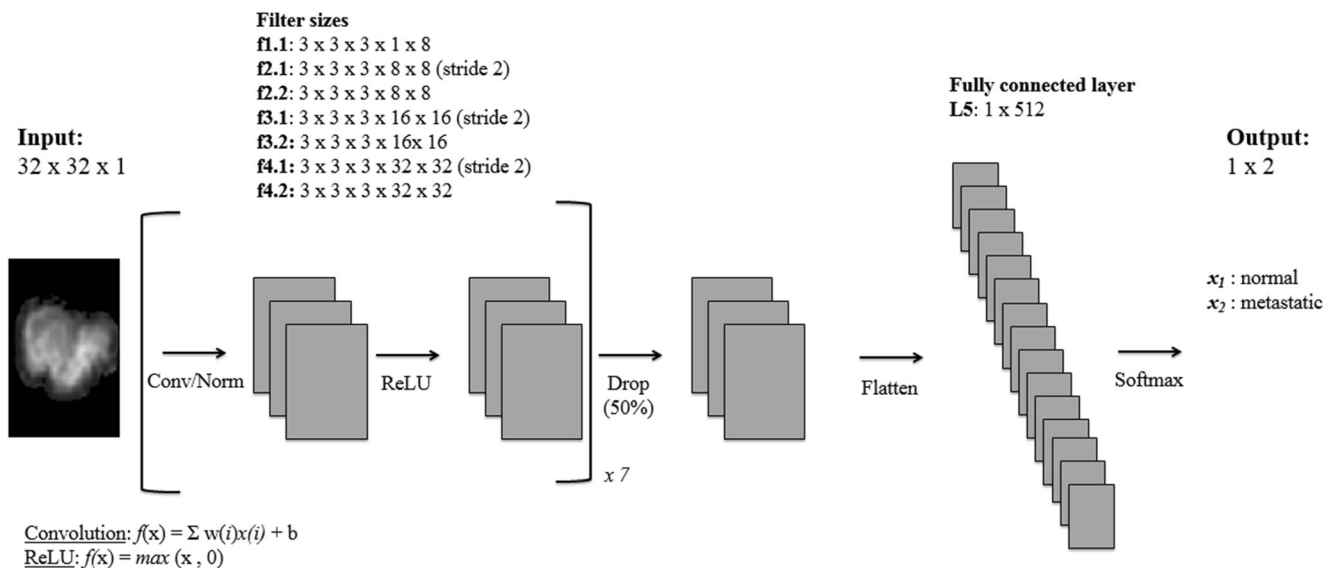

Fig. 3 CNN architecture. An eight hidden layer neural network was constructed based on a 32×32 input image with filter sizes as above. Seven consecutive convolution operations are performed with feature map downsampling done by utilizing convolutional layers with a stride size of two. A fully connected layer with 512 neurons was added as the final hidden layer. Network output consisted of a two class score prediction

accuracy was 85%. The study was limited by a small sample size (61 patients) and subjective evaluation of the lymph nodes. Furthermore, there was poor inter-observer agreement, when interpreting qualitatively the T1-weighted images, ($k = 0.57$ for first reading and $k = 0.78$ for second readings).

In our study, we have shown a validation accuracy rate of 84%, which is comparable to the highest accuracy of previously published data in the literature [13–15]. In comparison to the Scaranelo study, we had a larger sample size, and our study was more objective segmenting the entire lymph node with subsequent systematic analysis, instead of subjective identification of the region of interest manually defined within the lymph node by the reader.

Applying deep machine learning using CNN-based algorithm in our study, we were able to generate reasonable diagnostic performance in predicting axillary lymph node metastasis even with a small dataset. Larger dataset will likely improve our prediction model.

Our study has limitations. It is a small, retrospective study in a single institution. The performance of CNN has been shown to increase logarithmically with larger datasets [15]. Larger MRI datasets are likely to significantly improve the metastatic axillary lymph node prediction model. In addition, patients in this study underwent MRI at different magnetic field strengths (1.5 or 3.0 T), but this was determined randomly based on availability and thus limiting selection bias. Other limitations include inherent limitations of this technology including potentially long training times. Traditional algorithms comparatively take much less time to train; however, this is reversed during testing time, where a CNN can take much less time to execute. Manual inspection of false positive and false negative predictions of the network revealed no discernibly consistent features that consistently lead to false negative or false positive classifications from the network.

In conclusion, it is feasible for current deep CNN architectures to be trained to predict likelihood of axillary lymph node metastasis. Larger dataset will likely improve our prediction model and can potentially be a non-invasive alternative to core needle biopsy and even sentinel lymph node evaluation. Future research with a prospective randomized study is needed to further validate our findings.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Ivens D, Hoe AL, Podd TJ, Hamilton CR, Taylor I, Royle GT: Assessment of morbidity from complete axillary dissection. Br J Cancer 66(1):136–138, 1992

2. Duff M, Hill AD, McGreal G, Walsh S, McDermott EW, O'Higgins NJ: Prospective evaluation of the morbidity of axillary clearance for breast cancer. Br J Surg 88(1):114–117, 2001

3. Weiser MR, Montgomery LL, Susnik B, Tan LK, Borgen PI, Cody HSI: routine intraoperative frozen-section examination of sentinel lymph nodes in breast cancer worthwhile? Ann Surg Oncol 7(9): 651–655, 2000

4. Krishnamurthy S, Meric-Bernstam F, Lucci A, Hwang RF, Kuerer HM, Babiera G, Ames FC, Feig BW, Ross MI, Singletary E, Hunt KK, Bedrosian IA: prospective study comparing touch imprint cytology, frozen section analysis, and rapid cytokeratin immunostain for intraoperative evaluation of axillary sentinel lymph nodes in breast cancer. Cancer 115(7):1555–1562, 2009. https://doi.org/10.1002/cncr.24182.

5. Vanderveen KA, Ramsamooj R, Bold RJA: prospective, blinded trial of touch prep analysis versus frozen section for intraoperative evaluation of sentinel lymph nodes in breast cancer. Ann Surg Oncol 15(7):2006–2011, 2008. https://doi.org/10.1245/s10434-008-9944-8.

6. Pogacnik A, Klopcic U, Grazio-Frković S, Zgajnar J, Hocevar M, Vidergar-Kralj B: The reliability and accuracy of intraoperative imprint cytology of sentinel lymph nodes in breast cancer. Cytopathology 16(2):71–76, 2005

7. Akay CL, Albarracin C, Torstenson T, Bassett R, Mittendorf EA, Yi M, Kuerer HM, Babiera GV, Bedrosian I, Hunt KK, Hwang RF: Factors impacting the accuracy of intra-operative evaluation of sentinel lymph nodes in breast cancer. Breast J 24(1):28–34, 2018. https://doi.org/10.1111/tbj.12829

8. Ballal H, Hunt C, Bharat C, Murray K, Kamyab R, Saunders C: Arm morbidity of axillary dissection with sentinel node biopsy versus delayed axillary dissection. ANZ J Surg, 2018. https://doi.org/10.1111/ans.14382

9. Renaudeau C, Lefebvre-Lacoeuille C, Campion L, Dravet F, Descamps P, Ferron G, Houvenaeghel G, Giard S, Tunon de Lara C, Dupré PF, Fritel X, Ngô C, Verhaeghe JL, Faure C, Mezzadri M, Damey C, Classe JM: Evaluation of sentinel lymph node biopsy after previous breast surgery for breast cancer: GATA study. Breast 28:54–59, 2016. https://doi.org/10.1016/j.breast.2016.04.006.

10. An YS, Lee DH, Yoon JK, Lee SJ, Kim TH, Kang DK, Kim KS, Jung YS, Yim H: Diagnostic performance of 18F-FDG PET/CT, ultrasonography and MRI. Detection of axillary lymph node metastasis in breast cancer patients. Nuklearmedizin 53(3):89–94, 2014. https://doi.org/10.3413/Nukmed-0605-13-06.

11. Cooper KL, Meng Y, Harnan S, Ward SE, Fitzgerald P, Papaioannou D, Wyld L, Ingram C, Wilkinson ID, Lorenz E: Positron emission tomography (PET) and magnetic resonance imaging (MRI) for the assessment of axillary lymph node metastases in early breast cancer: systematic review and economic evaluation. Health Technol Assess 15(4):iii–iiv, 1–134, 2011. https://doi.org/10.3310/hta15040

12. Hwang SO, Lee SW, Kim HJ, Kim WW, Park HY, Jung JH: The comparative study of ultrasonography, contrast-enhanced MRI, and (18)F-FDG PET/CT for detecting axillary lymph node metastasis in T1 breast cancer. J Breast Cancer 16(3):315–321, 2013. https://doi.org/10.4048/jbc.2013.16.3.315

13. Scaranelo AM, Eiada R, Jacks LM, Kulkarni SR, Crystal P: Accuracy of unenhanced MR imaging in the detection of axillary lymph node metastasis: study of reproducibility and reliability. Radiology 262(2):425–434, 2012. https://doi.org/10.1148/radiol.11110639.

14. Hieken TJ, Trull BC, Boughey JC, Jones KN, Reynolds CA, Shah SS, Glazebrook KN: Preoperative axillary imaging with percutaneous lymph node biopsy is valuable in the

contemporary management of patients with breast cancer. Surgery 154(4):831–838, 2013

15. Abe H, Schacht D, Kulkarni K, Shimauchi A, Yamaguchi K, Sennett CA, Jiang Y: Accuracy of axillary lymph node staging in breast cancer patients: an observer-performance study comparison of MRI and ultrasound. Acad Radiol 20(11):1399–1404, 2013. https://doi.org/10.1016/j.acra.2013.08.003

16. LeCun Y, Bengio Y, Hinton G: Deep learning. Nature 521(7553): 436–444, 2015. https://doi.org/10.1038/nature14539.

17. Pieper S, Lorensen B, Schroeder W, et al: The NA-MIC Kit: ITK, VTK, pipelines, grids and 3D slicer as an open platform for the medical image computing community. Proceedings of the 3rd IEEE International Symposium on Biomedical Imaging: From Nano to Macro 1:698–701, 2006.

18. Simonyan K, Zisserman A: Very deep convolutional networks for large-scale image recognition. International Conference on Learning Representations. 2015, p. 1–14

19. Nair V, Hinton GE: Rectified linear units improve restricted Boltzmann machines. https://www.cs.toronto.edu/~hinton/absps/reluICML.pdf

20. Ioffe S, Szegedy C: "Batch normalization: accelerating deep network training by reducing internal covariate shift." International Conference on Machine Learning. 2015

21. Srivastava N, Hinton GE, Krizhevsky A, Sutskever I, Salakhutdinov R: Dropout : a simple way to prevent neural networks from overfitting. J Mach Learn Res 15:1929–1958, 2014

22. Kingma DP, Ba J: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014

23. He K, Zhang X, Ren S, et al: Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. arXiv: 1502.01852 https://arxiv.org/pdf/1502.01852.pdf