Cancer metastasis detection using convolutional neural networks and transfer learning

Ali Farahmand
afarah18@earlham.edu
Computer Science Department at Earlham College
Richmond, Indiana

ABSTRACT

Artificial Intelligence (AI) has been used extensively in the field of medicine. More recently, advanced machine learning algorithms have become a big part of oncology as they assist with detection and diagnosis of cancer. Convolutional Neural Networks (CNN) are common in image analysis and they offer great power for detection, diagnosis and staging of cancerous regions in radiology images. Convolutional Neural Networks get more accurate results, and more importantly, need less training data with transfer learning, which is the practice of using pre-trained models and fine-tuning them for specific problems. This paper proposes utilizing transfer learning along with CNNs for staging cancer diagnoses. Randomly initialized CNNs will be compared with CNNs that used transfer learning to determine the extent of improvement that transfer learning can offer with cancer staging and metastasis detection. Additionally, the model utilizing transfer learning will be trained with a smaller subset of the dataset to determine if using transfer learning reduced the need for a large dataset to get improved results.

KEYWORDS

Artificial Intelligence, Cancer Detection, Tumor Detection, Machine Learning, Transfer Learning, Convolutional Neural Networks, Radiology, Oncology

1 INTRODUCTION

Artificial Intelligence (AI) has grown into an advanced field that plays a major role in our healthcare. AI in conjunction with Machine Learning (ML), has been aiding radiologists with detecting cancerous regions [2], determining if the cancerous region is benign or malignant [5], to what degree cancer has spread outside of the initial area [6], how well a patient is responding to treatment [8], and more. Among many ML methods assisting radiologists, Convolutional Neural Networks (CNN) are deep learning algorithms capable of extracting features from images and making classification using those features [4]. CNN's are one of the major deep learning methods on image analysis and have become a popular tool in AI-assisted oncology [2]. Over the years many studies have attempted to improve the accuracy of these implementations by comparing different CNN architectures [14], addressing overfitting of the models, using continuous learning [12], transfer learning [14], etc. This proposal aims to improve cancer staging CNNs by applying transfer learning methods and combining the unique improvements that CNN and transfer learning methods can offer. In this paper, related work for implementation of CNNs and transfer learning for cancer detection is examined and compared to set up an understanding of the algorithms and the tools, the implementation of the CNNs and transfer learning is described, and finally the evaluation method for determining the accuracy of the CNNs is mentioned. Additionally, major risks for the implementation and a proposed timeline of the implementation are included.

2 BACKGROUND

This section focuses on outlining the main components of what is being proposed in this paper. CNN's and transfer learning methods are used frequently in recent related research and it is important to understand the basics of how they work.

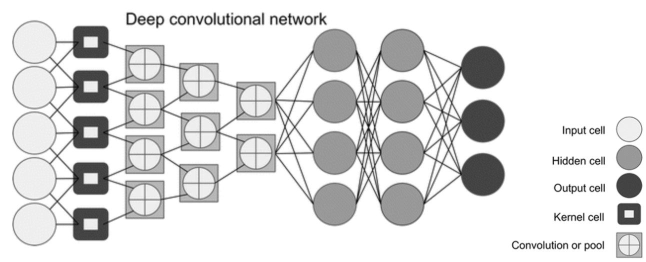

Figure 1: Simple CNN implementation (derived from Choy et al. [4])

2.1 Convolutional Neural Network

Convolutional Neural Networks are a subset of deep learning methods that extract features from images and further use these features for classification. CNNs are optimized for having images as input and since radiology is image focused, CNNs are one of the most common AI methods used in radiology [16]. A CNN consists of convolution and pooling layers. Figure 1 shows the layer layout of a basic CNN [4]. Convolution layers include filters that, through training, learn to create a feature map which outputs detected features from the input [16]. This feature map is then fed to a pooling layer, which downsizes the image by picking either the maximum value from the portion of the image that was covered by the convolution filter or the average value from the portion of the image that was covered by the convolution filter. These two pooling methods are referred to as **Max Pooling** layer and **Average Pooling** layer respectively. The purpose of pooling is to reduce computation and/or avoiding overfitting the model. At the end of the last convolution and pooling layers there is fully connected (FC) layer which is used as the classifier after the feature extracting process. Figure 2 visualises a CNN with two convolution layers and two pooling layers. There are multiple architectures for CNNs which use different layer combinations [16] and these architectures are used in detection, segmentation and diagnosis steps of oncology

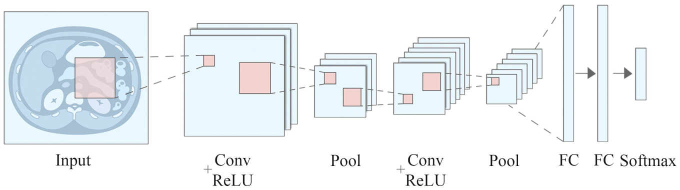

Figure 2: CNN with two conv and pooling layers (derived from Soffer et al. [16])

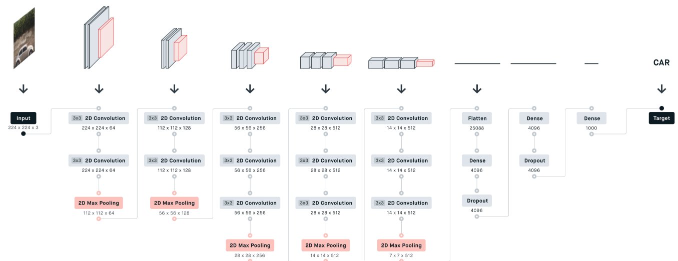

Figure 4: VGG16 architecture (derived from Peltarion website [11] based on Simonyan et al. [15]). VGG16 includes a total sixteen layers of convolution and pooling

[14]. Among the common architectures there are: AlexNet and VGG architectures. AlexNet is the shallow one of the two with five convolutional layers. AlexNet can have different numbers of pooling layers, normally on the convolutional layers that are closer to the FC layer. Figure 3 shows the AlexNet architecture without the pooling layers included. VGG is a deeper CNN with VGG16 having sixteen layers and VGG19 having nineteen layers. Both VGG16 and VGG19 are very clear about how many convolutional and pooling layers are included. Figure 4 shows a VGG16 architecture along with a breakdown of the layers. As shown in figure 4, pooling layers are present after every two or three convolutional layers in the VGG16 architecture. Both AlexNet and VGG16 have been used in cancer detection systems [16]. AlexNet, as the shallower of the two architectures is more commonly used for detection while VGG is used for diagnosis since it is a deeper network and as a result can extract more precise features from the radiology images. I am using a VGG16 architecture for the transfer learning experiment on the my dataset and a model closely similar to the VGG16 for the non-transfer learning part of the software.

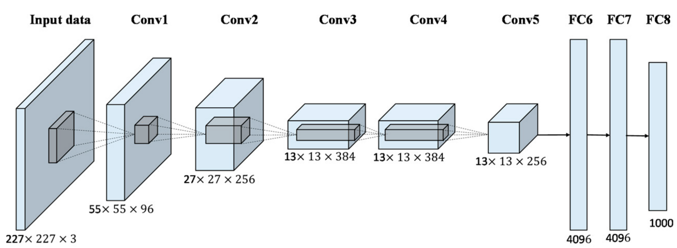

Figure 3: AlexNet architecture (derived from Han et al. [7]). AlexNet includes five convlution layers and a combination of pooling layers after any of the convolution layers

2.2 Transfer Learning

Transfer learning is inspired by the way humans learn new knowledge. The core concept is built around the idea of not isolating different learning environments, as knowledge gained from one learning process can be used in a different learning process with a different but similar goal. CNNs are commonly known to require large amounts of data for reasonable levels of accuracy, and as a result, training CNNs could face problems such as: not having access to enough data, not having access to enough hardware resources for computation, time-consuming training process, etc. Transfer learning can reduce the need for large sets of data while also increasing the accuracy of the CNN [13]. When a CNN without transfer learning is being trained, it is initialized with random weights and biases between the nodes of the network, however, in transfer learning, a pre-trained model is used as the initial state of the network and as a result less data is required to train a capable model for the original problem. This pre-trained model is a network that was trained to solve a different but similar problem. For instance, if we have a functional model that can detect horses in images, the model can be used, with little fine-tuning, for transfer learning into a new model that aims to detect dogs. Transfer learning can be very useful in cancer detecting CNNs as it helps improve and expedite the training process. Transfer learning with [1] and without [13] fine tuning has been used in medical imaging systems and has shown improved results. My software will be aimed to compare results of cancer metastasis detection with and without transfer learning and it will utilize fine-tuning.

3 RELATED WORK

Substantial research has been done on the usability of both CNN and transfer learning methods and how they can improve the results of Computer-Aided Detection (CADe) and Computer-Aided Diagnosis (CADx) systems. Shin et al. use three different CNN architectures along with transfer learning for cancer detection and have published very thorough results in their work [14]. Shi et al. use similar methods to reduce the number of false positives in cancer detection [13]. Bi et al. [2], Hosny et al. [8] and Soffer et al. [16] all have thoroughly explored the current and future applications of CNNs in cancer detection.

4 DESIGN

The process of acquiring images and pre-processing the data is no different than other cancer detection CNNs, I will be using the Breast Histopathologic Cancer Detection dataset from Kaggle [10]. This dataset includes over 200,000 images of breast pathology scans and it includes both images of metastatic cancer which was spread to other tissue and benign tumors that have not spread to other places. Determining if cancer has spread to other regions is a crucial part of staging the cancer in the National Cancer Institution's TNM staging system [3]. Table 1 and Table 2 show how the TNM system is utilized for each patient's staging and how important detecting the spread of cancer and the extent of the spread is.

	Range	Meaning
T	0-4	Size of the tumor, bigger number means bigger tumor
N	0-3	Number of nearby lymph nodes affected by cancer spread
M	0-1	Whether the cancer has spread to a distant organ

Table 1: The meaning of T, N and M numbers in TNM staging system. (derived from: National Cancer Institute's website [3])

Stage	Meaning
Stage 0	Abnormal cells present but no cancer present yet.
Stage I, II and III	Cancer is present
Stage IV	Cancer has spread to a distant organ

Table 2: Staging of the final TNM number. (derived from: National Cancer Institute's website [3])

Figure 5 shows the proposed framework of this project. This framework will be applied to both non-transfer learning and the VGG architectures with the key difference that the pre-trained model will only be used in the VGG architecture for transfer learning. Each architecture will be given exactly the same dataset to compare the results and check if transfer learning improved the results. Afterwards, the VGG architecture with the pre trained model for transfer learning will be trained once more with significantly less data to compare the results and check if transfer learning will indeed reduce the need for a large dataset to achieve better results. As shown in figure 5, the datasets will be pre-processed before being used for feature extraction in the CNN or for the classification, this includes downsampling the photos to a smaller size to reduce the load and improve the speed of training.

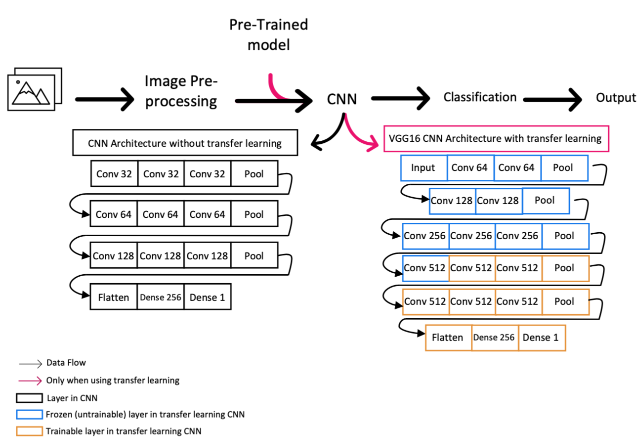

Figure 5: The overall framework of the project

	Layer	Trainable
1	Conv 32	Yes
2	Conv 32	Yes
3	Conv 32	Yes
3	Pooling	Yes
4	Conv 64	Yes
5	Conv 64	Yes
6	Conv 64	Yes
7	Pooling	Yes
8	Conv 128	Yes
9	Conv 128	Yes
10	Conv 128	Yes
11	Flatten	Yes
12	Dense 256	Yes
13	Dense 1	Yes

Table 3: Layers of CNN in non-transfer learning model

	Layer	Trainable
1	Input	No
2	Conv 64	No
3	Conv 64	No
4	Pooling	No
5	Conv 128	No
6	Conv 128	No
7	Pooling	No
8	Conv 256	No
9	Conv 256	No
10	Conv 256	No
11	Pooling	No
12	Conv 512	No
13	Conv 512	Yes
14	Conv 512	Yes
15	Pooling	Yes
16	Conv 512	Yes
17	Conv 512	Yes
18	Conv 512	Yes
19	Pooling	Yes
20	Flatten	Yes
21	Dense 256	Yes
22	Dense 1	Yes

Table 4: Layers of VGG16 CNN ImageNet transfer learning model with frozen and trainable fine-tuning layers

Tables 3 and 4 show the details the CNN layers used for the model without transfer learning and the VGG model using transfer learning. I have decided to use the pre-trained model trained on the ImageNet dataset [9] on the VGG CNN. Table 4 shows that the first eight convolution layers of the VGG CNN were chosen to be frozen with the ImageNet weights and biases for feature extraction and the remaining five convolution layers are chosen to remain trainable for fine-tuning the model for detecting cancer metastasis in this specific dataset.

4.1 Training and Results

The dataset was not balanced with the cases of cancer spread. So I chose about 80,000 of each of the two classes. A subset of 8000 images of each class was utilized as the validation files. Files from the remaining images in the dataset that were not used for training or validation will be used for testing the evaluation and validity of the results. Training on the non-transfer learning model are shown in figure 6. Figure 7 shows the training, validation accuracy and loss for the VGG16 model utilizing transfer learning with the ImageNet dataset. The slight increase in accuracy on the same data can be explained by the bigger model and the frozen layers pre-trained on extracting features from the images. Figure 8 shows the training and validation accuracy along with the loss for the same VGG16 model with the only difference that only half of the dataset was used for training and validation at 40,000 images of each class.

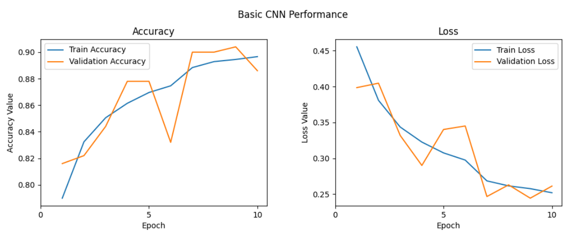

Figure 6: Training and validation accuracy and loss of the non-transfer learning model

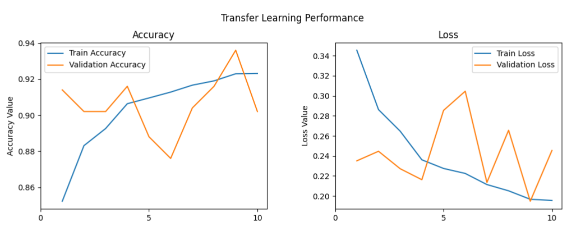

Figure 7: Training and validation accuracy and loss of the transfer learning model on ImageNet VGG16

Figure 8: Training and validation accuracy and loss of the transfer learning model with half of the dataset

The models were saved after training and used on a test dataset. This test dataset is 8000 images of each class that were not used for training and validation steps. The saved models were used for evaluating the models on their performance outside of the training process. Table 5 represents the results of this evaluation. Based on the results, using transfer learning not only improved the final results by roughly three percent it also had improved results when only half of the dataset was used for training.

Model	Evaluation result
Basic CNN	89.68%
VGG CNN using transfer learning with ImageNet Dataset	92.12%
VGG CNN using transfer learning with ImageNet Dataset trained only on half of the dataset	91.84%

Table 5: Final results used on the testing dataset not used in the training and validation steps

4.2 Validation of Results

The testing dataset included 16000 images which were 8000 images of each class. This testing dataset was not used in the training and validation steps, meaning the dataset was never seen before by the models. The benefit of using new images for evaluation is two fold: first, it shows that the training didn't overfit the model to only the training and validation datasets and second, it makes sure that the model is extracting features and classifying the images based on the correct features and not a coincidental set of features in the training and validation datasets.

5 FUTURE WORK

Using the final model trained and created by this software towards more complex problems and datasets would be my first idea of how to move forward after this. The model can be used for transfer learning in datasets that include the TNM staging numbers or datasets that have history of patient's visits for evaluating the response to cancer treatment. Another direction that this work could be helpful to build upon is to use different pre-trained models instead of ImageNet. Using models previously trained on clinical applications as opposed to generalized image datasets such as ImageNet and Ciphar could potentially improve the results as the model would be more specialized in the radiology detection world of datasets rather than images of everyday objects.

6 ACKNOWLEDGMENTS

I would like to thank Charles Peck and David Barbella for helping with this research idea and clarifying the details of the proposed technology.

REFERENCES

[1] Yaniv Bar, Idit Diamant, Lior Wolf, and Hayit Greenspan. 2015. Deep learning with non-medical training used for chest pathology identification. In *Medical Imaging 2015: Computer-Aided Diagnosis*, Vol. 9414. International Society for Optics and Photonics, 94140V.

[2] Wenya Linda Bi, Ahmed Hosny, Matthew B. Schabath, Maryellen L. Giger, Nicolai J. Birkbak, Alireza Mehrtash, Tavis Allison, Omar Arnaout, Christopher Abbosh, Ian F. Dunn, Raymond H. Mak, Rulla M. Tamimi, Clare M. Tempany, Charles Swanton, Udo Hoffmann, Lawrence H. Schwartz, Robert J. Gillies, Raymond Y. Huang, and Hugo J. W. L. Aerts. 2019. Artificial intelligence

in cancer imaging: Clinical challenges and applications. *CA: A Cancer Journal for Clinicians* 69, 2 (2019), 127–157. https://doi.org/10.3322/caac.21552 arXiv:https://acsjournals.onlinelibrary.wiley.com/doi/pdf/10.3322/caac.21552

[3] cancer.gov. [n.d.]. National Cancer Institute website. Retrieved September 22, 2020 from https://www.cancer.gov/about-cancer/diagnosis-staging/staging

[4] Garry Choy, Omid Khalilzadeh, Mark Michalski, Synho Do, Anthony E. Samir, Oleg S. Pianykh, J. Raymond Geis, Pari V. Pandharipande, James A. Brink, and Keith J. Dreyer. 2018. Current Applications and Future Impact of Machine Learning in Radiology. *Radiology* 288, 2 (2018), 318–328. https://doi.org/10.1148/radiol.2018171820

[5] Macedo Firmino, Giovani Angelo, Higor Morais, Marcel R Dantas, and Ricardo Valentim. 2016. Computer-aided detection (CADe) and diagnosis (CADx) system for lung cancer with likelihood of malignancy. *Biomedical engineering online* 15, 1 (2016), 1–17.

[6] Richard Ha, Peter Chang, Jenika Karcich, Simukayi Mutasa, Reza Fardanesh, Ralph T Wynn, Michael Z Liu, and Sachin Jambawalikar. 2018. Axillary lymph node evaluation utilizing convolutional neural networks using MRI dataset. *Journal of Digital Imaging* 31, 6 (2018), 851–856.

[7] Xiaobing Han, Yanfei Zhong, Liqin Cao, and Liangpei Zhang. 2017. Pre-trained alexnet architecture with pyramid pooling and supervision for high spatial resolution remote sensing image scene classification. *Remote Sensing* 9, 8 (2017), 848.

[8] Ahmed Hosny, Chintan Parmar, John Quackenbush, Lawrence H Schwartz, and Hugo JWL Aerts. 2018. Artificial intelligence in radiology. *Nature Reviews Cancer* 18, 8 (2018), 500–510.

[9] ImageNet. [n.d.]. ImageNet Dataset. Retrieved March 17, 2020 from http://www.image-net.org

[10] Kaggle. [n.d.]. Histopathologic Cancer Detection Dataset. Retrieved March 17, 2020 from https://www.kaggle.com/c/histopathologic-cancer-detection/overview

[11] peltarion.com. [n.d.]. Peltarion Website. Retrieved September 22, 2020 from https://peltarion.com/knowledge-center/documentation/modeling-view/build-an-ai-model/snippets---your-gateway-to-deep-neural-network-architectures/vgg-snippet

[12] Oleg S Pianykh, Georg Langs, Marc Dewey, Dieter R Enzmann, Christian J Herold, Stefan O Schoenberg, and James A Brink. 2020. Continuous learning AI in radiology: implementation principles and early applications. *Radiology* (2020), 200038.

[13] Zhenghao Shi, Huan Hao, Minghua Zhao, Yaning Feng, Lifeng He, Yinghui Wang, and Kenji Suzuki. 2019. A deep CNN based transfer learning method for false positive reduction. *Multimedia Tools and Applications* 78, 1 (2019), 1017–1033.

[14] Hoo-Chang Shin, Holger R Roth, Mingchen Gao, Le Lu, Ziyue Xu, Isabella Nogues, Jianhua Yao, Daniel Mollura, and Ronald M Summers. 2016. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Transactions on Medical Imaging* 35, 5 (2016), 1285–1298.

[15] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[16] Shelly Soffer, Avi Ben-Cohen, Orit Shimon, Michal Marianne Amitai, Hayit Greenspan, and Eyal Klang. 2019. Convolutional Neural Networks for Radiologic Images: A Radiologist's Guide. *Radiology* 290, 3 (2019), 590–606. https://doi.org/10.1148/radiol.2018180547