Deep Learning Models for Histopatological Images Classification

Maisun Al Zorgani and Hassan Ugail

July 9, 2018

Deep Learning Models for Histopatological Images Classification

Maisun Mohamed Alzorgani
Computer Science, Engineering and Informatics,
University of Bradford
Bradford City, UK
m.m.s.alzorgani@bradford.ac.uk

Professor Hassan Ugail
Director, Centre for Visual Computing,
University of Bradford
Bradford City, UK
h.ugail@bradford.ac.uk

Abstract— **This study aimed to compare five deep learning models for classifying Squamous Cell Carcinoma (SCC) tumor of the head and neck cancer from histopatplogical images. The five deep learning models were pre-trained convolutional neural networks (CNNs) included googleNet, Inception-v3, ResNet-50, ResNet-101 and Inceptionresnet-v2. These pre-trained CNNs were used to build fine-tuning networks with transfer learning. The building transfer learning networks replace the last three layers of pre-trained CNNs which configured for 1000 classes by new layers for binary classes and then fine-tune these layers on the Head and Neck Squamous Cell Carcinoma (HNSCC) tumor images. A total of 1,424 histopatological images of head and neck cancer were used to detect the tumor cells in sections. The transfer learning networks were compared in terms of standard performance. Although the number of images was insufficient, the results were shown good accuracy among different models. A highly successful classification has been achieved by the ResNet-50 model with accuracy rate was %98.95. But ResNet-101, googleNet and Inception-v3 performed classification with accepted accuracy rates of %97.89, %97.19 and 94.04%, respectively.**

Keywords; Deep Convolution Neural Network; head and neck cancer; Image classification; Standard Performance.

I. Introduction

Ferlay (2010) reported that Head and Neck cancer is still one of the most common types of cancer. It is the 6th most common type of cancers worldwide and 3rd in developing countries [4]. Forastiere et al. (2003) divided Head and Neck cancer into carcinoma in situ or in squamous cells, according to the location and the deepness of the cancer in the tissue layer. Squamous Cell Carcinoma indicates deeper location in the tissue layer and is the most common type [5]. They referred that most head and neck tumors are histologically SCC type and therefore are categorized as HNSCC [5].

Motlagh et al described the human errors that occurred in the manual inspection methods have adversely effect on the accuracy of the classification in conventional cancer diagnosis [10]. Automated diagnosis have greater benefits than pathologist's diagnosis to patients; as they get their diagnosis and personalised treatment faster, possibly leading to greater survival rates [9]. Automated classification accuracy of tumors using histopathological images might be made possible by deep learning approaches, which more reliable and economical compared to conventional methods [10].

Automated cancer multi-classification from histopathological images plays a key role in computer-aided breast cancer diagnosis. It helps to analyze and interpretation of the histopathological slides for assisting the doctors to choose more efficient therapeutic approaches [10]. Several studies have developed Computer-aided diagnosis (CAD) methods of histopathology images. The example of using automated CAD system is a study conducted by Wang and colleague on breast cancer identification using four deep learning network architectures including GoogleNet, AlexNet, and VGG16 deep network, which was done to classify benign and malignancy status [15]. Motlagh and colleague suggested a generic CAD framework based on deep networks for learning histopathology images [10]. In their work, compared the performance of Inception and ResNet deep learning models using transfer learning strategy on several large image datasets and found that deep ResNet models were more sensitive and reliable than Inception in all tested cancer data-sets. In addition to these, studies such as ([8], [1], [7]) also showed that deep learning techniques are continuously being applicable to image-based medical diagnosis and improve the performance compared to traditional machine learning techniques.

This work extracts discriminative features from the histopatological images of HNSCC tumour by using five pre-trained CNNs, and then fine tune networks using transfer learning strategy on these image datasets.

II. MATERIALS

A. *Data sets*

In this study, image data-sets were obtained from The Ethical Tissue department at the University of Bradford. A total of 1,424 histopathological images derived from Hematoxylin and eosin stain (H&E) slides. A total of 1,424 histopatological images of head and neck cancer were used to detect the tumor cells in sections. 1184 images are HNSCC tumour and 240 images are normal histology. These image data sets stored in jpg format.

B. Data augmentation

In order to deal the difference in the number of images among two classes, Data augmentation was introduced. Dosovitskiy et al. (2013) examined the importance of data augmentation in deep learning to get enough different samples which needed to train a ConvNet network from the images [3]. The dataset images were rotated (90, 180 and 270), flipped left to right horizontally and then vertically to create a larger sample size and to make the approach recognise tumour cells in different orientations (figure 1). These images with 40X magnification level, and did not magnify to other levels. There is large number of studies employed single magnification level ([6]).

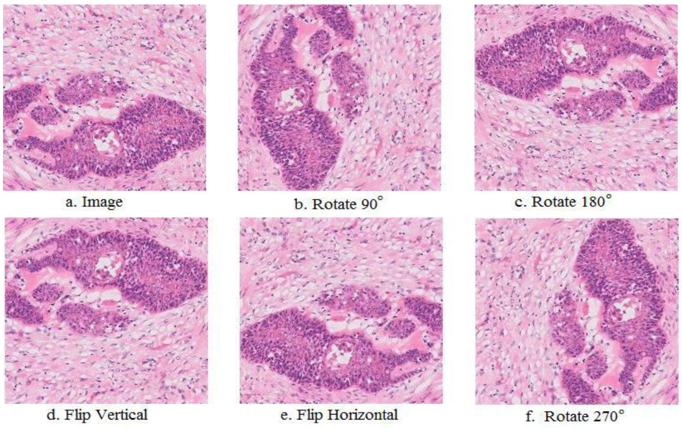

Figure 1 shows Data Augmentation of the original image

III. METHODOLOGY

A. Transfer Learning

Fine-tuning a CNN with transfer learning is commonly used in deep learning applications [16]. It is faster than constructing a new network, because a pre-trained CNN on millions of images could be taken to retrain for new classification using only hundreds of images [16]. Training from scratch is often not the most practical strategy for medical images due to its computational cost, convergence problem [13], and insufficient number of high quality labeled samples [10]. Pre-trained CNNs besides fine-tuning with transfer learning gain the faster convergence and the outperform training from scratch [13]. The transfer learning strategy replaces layers from CNN and then retrain it on the new dataset by fine-tune the weights with the back propagation algorithm. It is possible to fine-tune all the layers of the CNN, or keep some of the earlier layers fixed (due to over-fitting concerns) and only fine-tune some higher-level portion of the network.

The last three layers of the pre-trained CNN are configured for 1000 classes, so these layers must be fine-tuned for the new binary classification problem (HNSCC tumour and normal histology). To achieve that, all the layers were extracted, except the last three, from the pre-trained CNN and then retrained them to classify head and neck tumour. Therefore, the weights of CNN were preserved while the last three layers were updated continuously.

B. Inceptions, ResNet and Inceptionresnet architectures

GoogleNet, Inception-v3, ResNet and Inceptionresnet architectures were considered in this study. These CNNs are trained on a subset of the ImageNet database, which is used in the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) [14]. These models are trained on more than a million images and can classify images into 1000 object categories [14]. These CNNs are Directed Acyclic Graph (DAG) network type. DAG network type is neural networks for deep learning which has a more complex architecture where layers have inputs from, or outputs to, multiple layers [14].

GoogLeNet is a 22 layers deep CNN, trained on the ILSVRC2014 dataset [14]. GoogLeNet uses the so called Inception modules that consist of multiple parallel convolution kernels that process the same input, concatenating feature at different scale [14]. In addition, also features extracted by pooling operation are concatenated by the Inception module [14]. Using the inception module, GoogLeNet is able to achieve high accuracy using limited computational cost. The Inception-v3 is larger, deeper and slower than GoogleNet, but more accurate on the original ILSVRC data set. Inception-v3 is 48 layers deep and consists of multiple convolutional and pooling layers which outputs are concatenated [14]. The connections of ResNet enable training of deeper networks. ResNet-50 is 50 layers deep and ResNet-101 is 101 layers deep [14]. Inception-ResNet-v2 combines a high-efficiency inception module of GoogLeNet with the residual connections of ResNets. Inception-ResNet-v2 is 164 layers deep [14].

C. Pre-processing Step

The 1st step, the training and test images were resized to height and width according to the Image Input size of the pre-trained CNN, before they are input to the pre-training network. GoogLeNet, ResNet-50 and ResNet-101 models require image input of size 224 as height by 224 as width by 3 as channel, so the training and test images were resized to 224×224×3. While, Inception-v3 and Inceptionresnet-v2 models require image input of size 229×229×3.

The 2nd step, the image datasets should be set down to 80% for training and 20% for testing. Bukar & Ugail (2017) demonstrated that training set of data should be considerably larger in order to give more accurate results [2].

D. Standard Performance

The building transfer learning networks were implemented and evaluated according to standard performance such as Accuracy (ACC), Precision (P) and Sensitivity (S). Sokolova and Lapalme (2009) extracted these terms from the confusion matrix [11]. The standard performance measurements were formulated as depicted in the following equations [11];

$$\text{Accuracy} = \frac{(tp + tn)}{(tp + fn + fp + tn)} \tag{1}$$

Accuracy evaluates the overall effectiveness of a classifier.

$$\text{Precision} = \frac{tp}{tp + fp} \tag{2}$$

Precision evaluates the class agreement of data labels with the positive labels given by the classifier.

$$\text{Sensitivity} = \frac{\text{tp}}{\text{tp} + \text{fn}} \tag{3}$$

Sensitivity evaluates the effectiveness of a classifier to identify positive labels

IV. RESULTS

A 2 × 2 confusion matrix was used to represent prediction results of the set of two pathological samples (HNSCC tumor and Normal histology). These 2 × 2 confusion matrixes are shown in figure 2. The results from implemented five models were illustrated in the table 1.

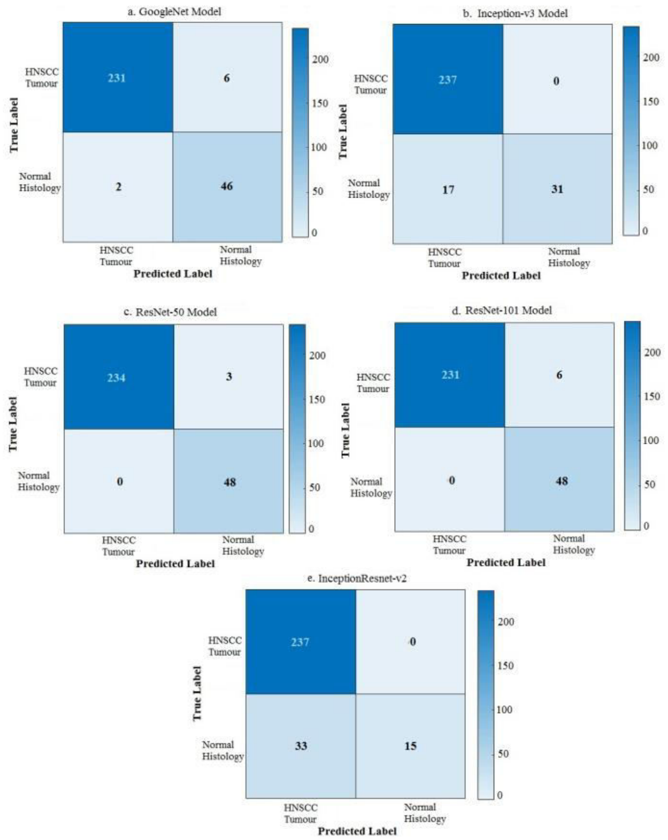

Figure 2 shows the confusion matrixes for Transfer Learning Networks

Deep Learning Model	Standard Performance Measurements		
	Acc (%)	P	S
GoogLeNet	97.19	.9665	.9380
Inception-v2	94.04	.8229	.9444
ResNet-50	98.95	.9937	.9706
ResNet-101	97.89	.9873	.9665
Inceptionresnet-v2	88.42	.6563	.9389

TABLE 1 illustrates the standard performance for Transfer Learning Networks

For comparison, different deep learning methods were trained with the same training sets and tested with the same testing sets. The initial learning rate was selected 0.0001 as a starting point. The software validates the network every 3 iterations. For training cycle, 15 epochs, 8 iterations per epoch and 120 iterations were selected. Comparing the timing performance of the five methods, GoogLeNet is the fastest; it took ~45 minute to train, using a computer with 2.2 GHz dual core i7 CPU. For the same training data, Inception-v2, ResNet-50, ResNet-101 and InceptionResNet-v3 took ~2 h, ~1 h 60 minute, ~2 h 80 minute and 5 h, respectively. Figures 3,4,5,6 and 7 show the training process of the transfer learning models for googleNet, Inception-v3, ResNet-50, ResNet-101 and Inceptionresnet-v2, respectively. As Figures reveal, from the 6th epoch, the accuracy rate were in approximately steady state. The reason for ending training at epoch 15 is that the error fall slowly from the 6th epoch.

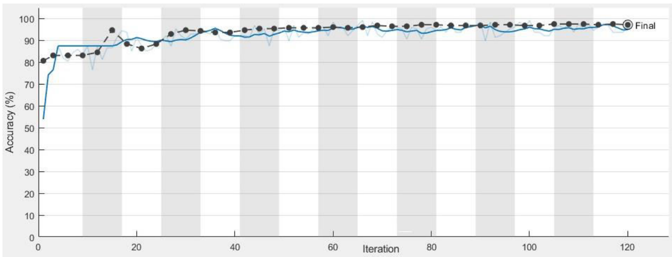

Figure 3 shows the training progress for GoogLeNet Model

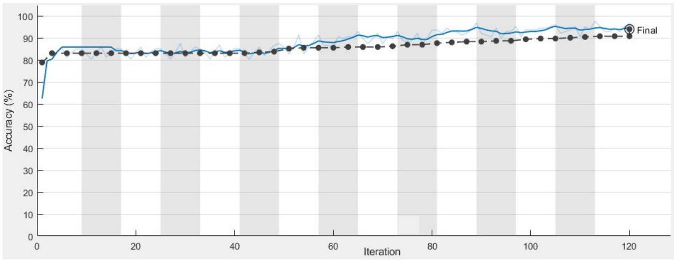

Figure 4 shows the training progress for Inception-v3 Model

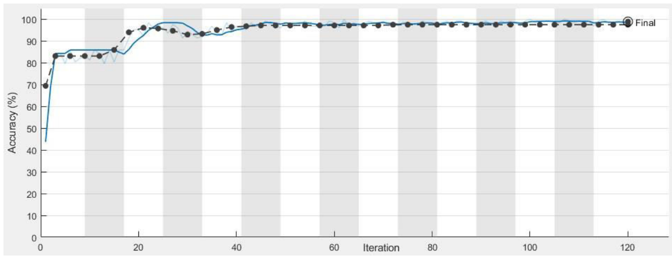

Figure 5 shows the training progress for ResNet-50 Model

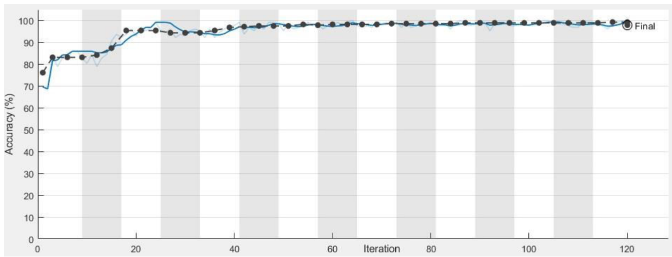

Figure 6 shows the training progress for ResNet-101 Model

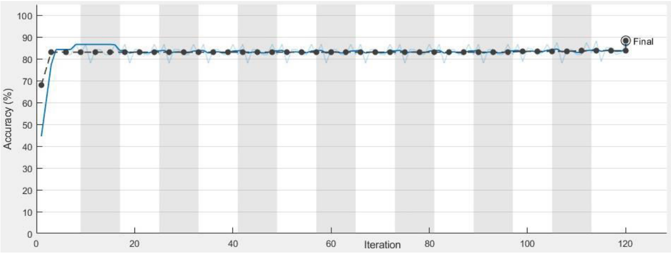

Figure 7 shows the training progress for InceptionResNet-v2 Model

V. DISCUSSION AND CONCLUSION

This work compared the performance of GoogleNet, Inception-v3, ResNet-50, ResNet-101 and Inceptionresnet-v2 deep learning models using transfer learning strategy on histopathology image datasets. Deep ResNet models were found more sensitive and reliable than others models. ResNet-50 and ResNet-101 had significantly high accuracy rates 98.95 % and 97.89 %. Moreover, Motlagh and colleagues (Motlagh et al., 2017), reported the ResNet-50 and ResNet-101 models based on multi classification for four cancer types with the average accuracy rates of 99.3 and 99.5 % , respectively.

In conclusion, Using deep learning ResNet approach with specific settings for cancer detection is an effective and reliable strategy compared to the conventional approaches [9]. This work concerned on the application of the proposed approaches to HNSCC tumour detection. The ResNet-50 approach is more suitable than others, because it is able to examine our histopathological images.

ACKNOWLEDGMENT

We thank Professor Hassan Ugail for his encouragement and appreciation of his guidance and advice.

REFERENCES

[1] Aubreville, M., Knipfer, C., Oetter, N., Jaremenko, C., Rodner, E., Denzler, J., Bohr, C., Neumann, H., Stelzle, F. and Maier, A.," *Automatic classification of cancerous tissue in laserendomicroscopy images of the oral cavity using deep learning*". arXiv preprint arXiv:1703.01622. 2017.

[2] Bukar, A.M. Ugail, H., *"Convnet Features for Age Estimation*", Bradford, UK, ISBN: 9978-989-8533-66-1. 2017

[3] Dosovitskiy, A. ,Springenberg, J. T. and Brox, T.,"*Unsupervised feature learning by augmenting single images*". CoRR, abs/1312.5242. 2013.

[4] Ferlay, J. Shin, H.R. Bray, F. Forman, D. Mathers, C. Parkin, D.M. "*Estimates of worldwide burden of cancer in 2008*": GLOBOCAN 2008, International Journal of Cancer, 127(12):2893-2917. 2010.

[5] Forastiere, A.A. Goepfert, H. Maor, M. Pajak, T.F. Weber, R. Morrison, W. Glisson, B. Trotti, A. Ridge, J.A. Chao, C. *"Concurrent chemotherapy and radiotherapy for organ preservation in advanced laryngeal cancer"*, National English Journal of Medicine, 349:2091-2098. 2003.

[6] Kahya, M. A., Al-Hayani, W. and Algamal, Z. Y., "*Classification of breast cancer histopathology images based on adaptive sparse support vector machine"*. Journal of Applied Mathematics and Bioinformatics, 7(1):49. 2017.

[7] Khosravi, P., Kazemi, E., Imielinski, M., Elemento, O. and Hajirasouliha, I., *"Deep convolutional neural networks enable discrimination of heterogeneous digital pathology images"*. EBioMedicine.

[8] Liu, Y. Gadepalli, K. Norouzi, M. Dahl, G. E. Kohlberger, T. Boyko, A. Venugopalan, S. Timofeev, A. Nelson, P. Q. Corrado, G. S., et al. *"Detecting cancer metastases on gigapixel pathology images*". arXiv preprint arXiv:1703.02442. 2017.

[9] Morroney, R. *"The Future of Pathology: Going Digital, Inspirata: Transforming Cancer Diagnostics*", URL: www.inspirata.com/the-future-of-pathology-going-digital/ Date Seen: 2nd August 2017.

[10] Motlagh, M. H., Jannesari, M., Aboulkheyr, H., Khosravi, P., Elemento, O., Totonchi, M. and Hajirasouliha,I., "*Breast Cancer Histopathological Image Classification: A Deep Learning Approach"*, 2018

[11] Sokolova, M., and Lapalme, G,. "*A systematic analysis of performance measures for classification tasks*". Inf. Process. Manage., 45(4):427–437, July 2009.

[12] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A., "*Going deeper with convolutions.*" In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1-9. 2015.

[13] Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B. and Liang, J., "*Convolutional neural networks for medical image analysis: Full training or fine tuning?"* IEEE transactions on medical imaging, 35(5):1299–1312, 2016.

[14] T. Google-Developers. Pretrained Convolutional Neural Networks. https://uk.mathworks.com/help/nnet/ug/pretrained-convolutional-neural-networks.html, MathWorks,2018

[15] Wang, D., Khosla, A., Gargeya, R., Irshad, H. and Beck, A. H. *"Deep learning for identifying metastatic breast cancer*". arXiv preprint arXiv:1606.05718, 2016.

[16] Zeiler, M. D. and Fergus, R.," *Visualizing and understanding convolutional networks*". In European conference on computer vision, pages 818–833. Springer, 2014.