PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

A new breast cancer diagnosis application based on ResNet50

Anjie Le, Zhenghao Li, Haoyun Tang, Haobo Yang

Anjie Le, Zhenghao Li, Haoyun Tang, Haobo Yang, "A new breast cancer diagnosis application based on ResNet50," Proc. SPIE 12079, Second IYSF Academic Symposium on Artificial Intelligence and Computer Engineering, 120792K (1 December 2021); doi: 10.1117/12.2623103

SPIE.

Event: 2nd IYSF Academic Symposium on Artificial Intelligence and Computer Engineering, 2021, Xi'an, China

A New Breast Cancer Diagnosis Application Based on ResNet50

Anjie Le[a], Zhenghao Li[b], Haoyun Tang[c], Haobo Yang[d]
These authors contributed equally
[a]Department of Mathematics, University of Cambridge, Cambridge, UK, al912@cam.ac.uk;
[b]Department of Informatics, King's College London, London, UK, k1898873@kcl.ac.uk;
[c]School of Computing, Harbin Institute of Technology, Harbin, China, 1190201526@stu.hit.edu.cn;
[d]School of Informatics, University of Edinburgh, Edinburgh, UK, s1911593@ed.ac.uk

ABSTRACT

Histopathology is the primary tool employed in breast cancer diagnosis. It involves examining metastatic tissues of lymph nodes under a microscope. Histopathologists are responsible for making tissue diagnoses, and the process is challenging and tedious. To diminish their workload and allocate more time to efficiently maintain patients' care, deploying an intelligent system to support the diagnosis is reasonable. Therefore, we introduced a new breast cancer diagnosis application based on deep learning technology in this paper. The application's foremost objectives were to handle the vast dataset of digital pathology scans and train deep residual networks to classify small patches from the sizable whole slide images with higher accuracy. Experimental outcomes indicated that our model could achieve 97.3%. Another noteworthy feature was a FAQ chatbot that we implemented for patient consulting.

Keywords: cancer; CT scan; machine learning; deep learning; convolutional neural network (CNN); whole slide image (WSI); residual networks (ResNets)

1. INTRODUCTION

Breast cancer is one of the most common cancers among women, which can be diagnosed by histopathologic scans from whole-slide imaging(WSI). WSI refers to scanning conventional glass slides to produce digital cost-effective and high-throughput histopathology slides [1]. With technological advancements, Artificial Intelligence (AI) has made tremendous advances in classifying diseases from WSI [2]. AI can extract features and reasoning, and algorithms can become more precise and accurate as they interact with large training datasets [3]. Undoubtedly, the combination of AI and the medical field can provide significant benefits, including clinical decision support, tasks automation, datasets analysis, diagnosis improvement on speed, accuracy and consistency.

However, healthcare practitioners' adoption of the deep learning diagnostic system is prevented by some challenges. The privacy of patient data is not guaranteed, and it is complex to supervise. Furthermore, gaining complete trust from AI's prediction outputs is not easy because the mechanism and the accuracy of AI diagnosis are not transparent to patients. An excellent deep learning model is the key to an accurate diagnosis. Regarding the model chosen in the early cancer diagnosis field, a study uses NasNetMobile [4] to classify images without preprocessing. It may give incorrect or sub-optimal inputs to the model so that the designed model cannot learn the distribution of the dataset well. Additionally, this study also ignored the deployment of the model and cannot be used directly by practitioners.

To overcome the limitations mentioned above, we aim to build a local run application for interpreting ResNet50 on breast cancer diagnosis. ResNet50 is a residual neural network based on pyramidal cell constructions in the cerebral cortex, and it has 48 Convolution layers, 1 MaxPool layer, and 1 Average Pool layer [5]. Additionally, we used a scanning method [6] and proposed a random cropping method to eliminate human-introduced bias.

The rest of the paper is divided into four segments. Segment 2 describes the detailed methods we used, including dataset analysis, algorithms and model implementations, moreover, an introduction to our application structure. Segment 3 provides the result and discussion based on our implementation. Eventually, a conclusion will be given in segment 4.

1.1 Dataset Description

In our study, we used the dataset that contains two parts: the original WSI of the lymph nodes and a cropped version.

Second IYSF Academic Symposium on Artificial Intelligence and Computer Engineering,
edited by Wei Qin, Proc. of SPIE Vol. 12079, 120792K · © 2021 SPIE
0277-786X · doi: 10.1117/12.2623103

The former was from GrandChallenge.org, namely 'Camelyon 16' [7]. It comprises four hundred WSIs of sentinel lymph nodes acquired and digitised from the Radboud University Medical Center (Nijmegen, the Netherlands) and the University Medical Center Utrecht (Utrecht, the Netherlands) using a 40x objective (resultant pixel resolution of 0.243 microns). Two hundred seventy of WSIs were labelled in the dataset. Note that the WSI is extremely large, with a typical file size of 1.5 GB. The latter was available on Kaggle [8], which initially came from an open-source dataset, namely PatchCamelyon(PCam) [9]. The PCam was extracted from the Camelyon 16 by cropping the WSIs into smaller 96×96px images, with a label indicating pixel of tumour tissue within the centre 32×32px region. The dataset from Kaggle is a selection of PCam datasets, excluding any blurred or duplicated images. There are in total 220k labelled images in the training dataset, of which we used 80% as the training set and 20% for testing.

1.2 Dataset Preprocessing

Our dataset needs to be preprocessed firstly before being fed into the model. We further expand the dataset by randomly selecting four 72×72px images from each of the 96×96px images. This process was specially designed to resolve the problem in real WSI: The tumour tissue is not designated to be in the middle, as in the case of the dataset on Kaggle. A process to normalise the brightness of the images was also done initially, given the fact that the mean brightness was unevenly distributed between images with positive and negative labels. Nevertheless, later experiments confirmed that this process did not increase the accuracy significantly and was abandoned consequently.

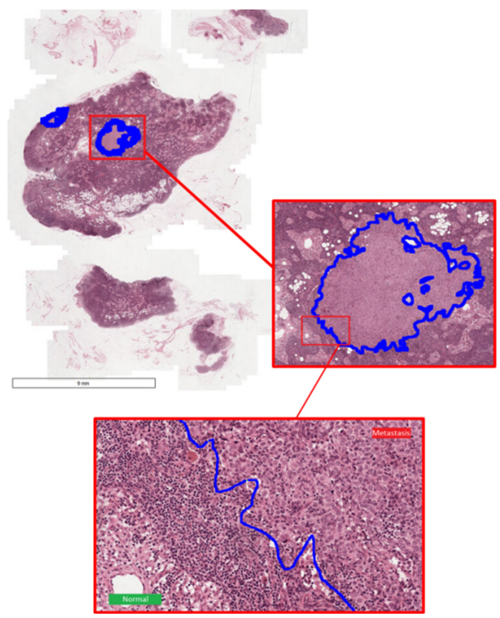

Figure 1. A full Whole Slide Image (top left), with the blue area being the cancer-metastasised area. After several cuts, we get the bottom image, an example of the image we may use to train the model

1.3 Algorithm

1) Image classification based on ResNet-50

In the performance benchmark for models on the ImageNet dataset [10], a dataset used to train models to perform 1000 category image classification [11], Deep Residual Network 50 (ResNet50) [12], can reach an 0.921 accuracy and keep the number of trainable parameters reasonably small. ResNet50 is a Convolutional Neural Network (CNN), which is widely used in image classification tasks. CNN is one of Artificial Neural Networks (ANN) consisting of Convolutional Layers to recognise patterns from image-like data [13]. Hence, we decided to apply ResNet50 to our task.

Table 1. The benchmark on the ImageNet dataset. The top-1 and top-5 accuracy refers to the model's performance on the ImageNet validation dataset [11]

Model	Model Benchmark		
	Top-1 Accuracy	***Top-5 Accuracy***	***Parameters***
NASNetLarge	0.825	0.96	88,949,818
ResNet50	0.749	0.921	25,636,712
NASNetMobile	0.744	0.919	5,326,716
VGG16	0.713	0.901	138,357,544
MobileNet	0.704	0.895	4,253,864

The next step is to select a suitable loss function for training an ANN. It is essential to compare the distance between a predicted result and an actual category label to determine how good our model is. As a result of this, we used the loss function (1).

$$Loss = -(ylog(p) + (1 - y)\log(1 - p)) \tag{1}$$

Where y is the binary indicator, the input image's actual class, p is the predicted probability that the input image belongs to the actual class. Moreover, this loss function is also known as binary cross-entropy.

When dealing with sparse data, Adam is considered to be the best among adaptive optimisers. To speed up the training process, we chose it as our model optimiser. There was also an accuracy metric calculating how often predictions matched the labels [14], and we set the learning rate at 0.001 and batch size of 100.

2) FAQ chatbot based on NLP

A FAQ chatbot using Natural Language Processing(NLP) was implemented for answering some of the most frequently asked questions users may have. The text corpus for training was obtained from "Breast cancer - Mayo Clinic" [15], "Lymph - Wikipedia", and "Dictionary of Cancer Terms - National Cancer Institute" [16].

When the user typed in some sentences, the entire text was first converted into lowercase. Then the words were tokenised and lemmatised. After applying a filter of common stop words, the sentences were turned into and stored as a list of keywords.

Two means of replying were implemented - keyword matching and response generating from a text corpus. Regarding the first way, a predefined response was given whenever the user mentioned any keyword in the predefined list. The latter used Term Frequency-Inverse Document Frequency (TF-IDF) and Cosine Similarity to compare the input keywords with text in the corpus and generate the most relevant response.

1.4 Application Structure

The software architecture is mainly divided into two parts: the diagnosis part and the consultation part. The detailed process is as follows:

1) Detection Part and Consultation Part

In the detection part, the program receives an uploaded image from the user and then predicts results. After the diagnosis, the user can upload more pictures, interact with the chatbot or return to the home menu. In the consultation part, the program receives a user's character string input and returns a response. After each message sending, the user can continue consulting, detect the image again, or return to the home menu.

2) Front-end development and UI design

The application includes three user interfaces. The home page shows a menu with two main functionalities. Clicking each of the labels navigates the user to a new page accordingly. The diagnosis page is divided into two sections; the upper section provides three horizontal windows, each representing the image, result, and emoji. The lower section offers a bar with four buttons to navigate users to upload and diagnose histopathologic scans; they can also jump to other pages. Once users upload their scan image and click the diagnosis button, the first window will show the uploaded image; the second window will present whether the tissue is normal or abnormal. The third window will show an emoji according to the feedback. Another objective of this program is to provide users with a Q&A chatbot that can answer their general medical questions, and it is implemented on the AI chatbot page. Like other software, simply typing and sending will get a response.

2. RESULT AND DISCUSSIONS

2.1 The performance of CNNs

There are currently a large number of CNN models available, as mentioned in section 2.2.1. To select the best model for our purpose, we attempted 10 epochs of training with our dataset on three of the most popular CNN models. We took a record of their performances, as illustrated in Table II.

In particular, the VGG16 model achieved the highest accuracy overall. However, the problem with this model is that the accuracy only improved with small batch size. It took about 2650 seconds to perform one epoch of training with this model. In contrast, one epoch in the ResNet50 or NASNetMobile model could be trained in approximately 2200 seconds, even with a larger batch size.

Considering both training time and accuracy performance, we decided to use the ResNet50 model in our study. The models are trained with 40 epochs. The ResNet50 model achieved a 97% accuracy at last on the testing dataset.

Table 2. The performance of different models on our dataset

Model	**Model Benchmark**		
	Training Loss	***Training Accuracy***	***Testing Accuracy***
ResNet 50 batch size = 200	0.0981	0.9639	0.8772
VGG16 batch size = 64	0.0579	0.9801	0.9764
NASNetMobil batch size = 200	0.3131	0.8664	0.8210

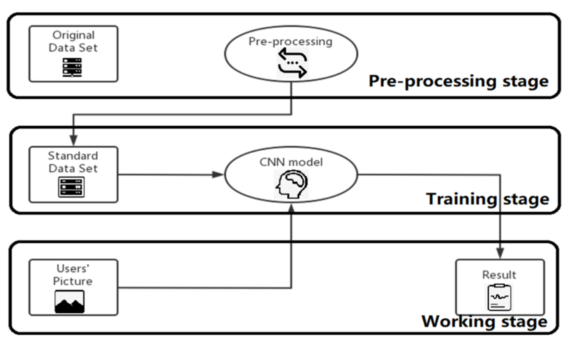

Figure 2. System Structure

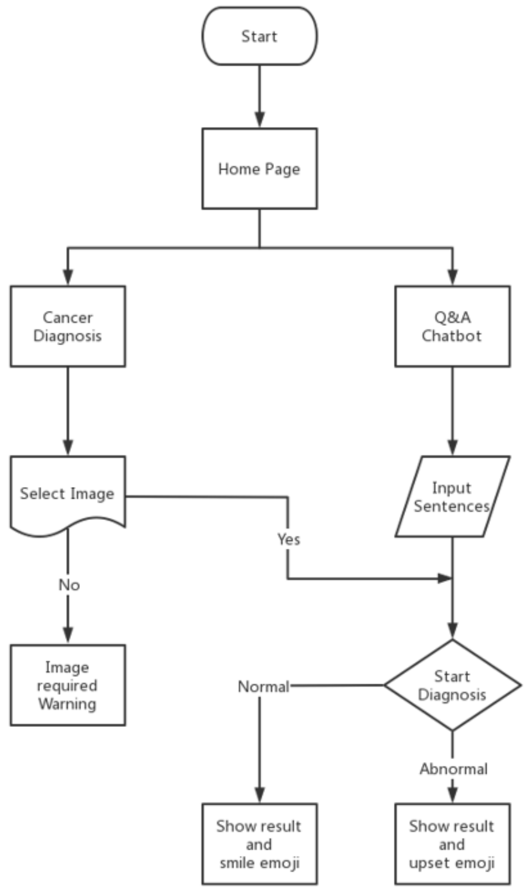

Figure 3. Application Flowchart

2.2 Random Cropping

In this segment, the models are trained from scratch without pre-train parameters on both the uncropped dataset (d1) and cropped dataset(d2). We tested these two models with the testing dataset to determine whether random cropping can push the ANN model to its limit.

Following 40 epochs of training, the model accuracy on d1 reached 99%, 2% higher than the model trained on d2. However, when we deployed these two models on the testing dataset, the accuracy of the model trained on d1 dropped significantly to 95%, smaller than the 96.9% accuracy of the model trained on d2.

We believe it was a consequence of overfitting. The model trained on d1 suffered more from overfitting than the model trained on d2. Hence, random cropping is a reasonable approach to prevent our model from overfitting.

2.3 User Interface

Coding the user-facing aspects of a software program is essential. Qt Designer is a convenient and efficient tool for designing and building graphical user interfaces with widgets from the Qt GUI framework. We used Qt Designer and PyCharm to write and edit Python code to let users interact with our deep learning models. Below are three samples of our interfaces:

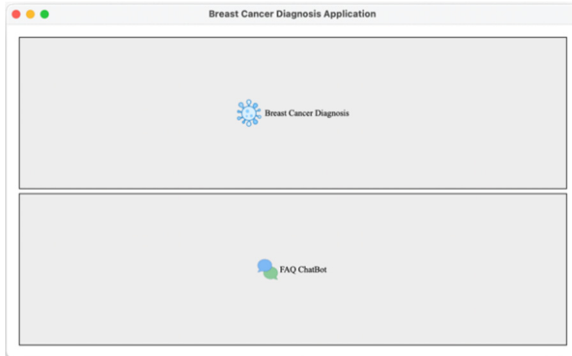

Figure 4. The home menu provides two functionality entries

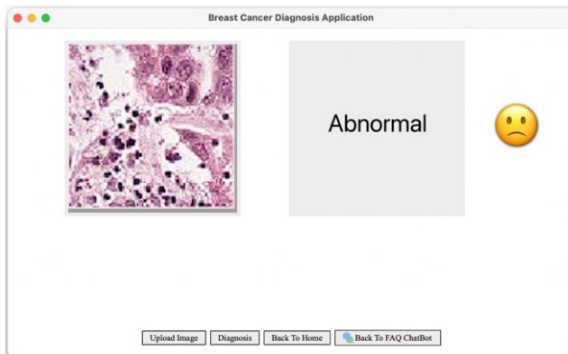

Figure 5. The cancer diagnosis page predicts results from user input

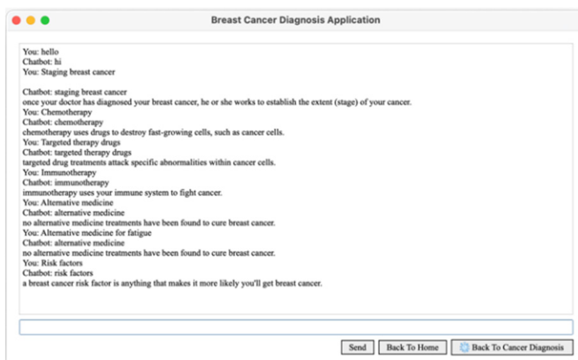

Figure 6. The FAQ chatbot page provides information based on the user's query

3. CONCLUSION

A local run application for automated breast cancer diagnosis using WSI has been presented. The application's fundamental issues were processing the large dataset and training the ResNet50. With a high-accuracy detection model and an easy-to-navigate user interface, our cancer diagnosis application was a success. Compared with previous medical applications, our design is straightforward and intuitive, with considerably higher accuracy and a superior capacity to lessen users' cognitive strain. Meanwhile, the built-in chatbot can make it more pleasant to use.

In the future, we suggest a novel CNN model design, the big-small core architecture, enabling prospective future upgrades to increase the time-performance of our system. As a complement to our big ResNet50 core, we may introduce a smaller CNN core with fewer parameters to improve the final performance of models.

REFERENCES

[1] M. Veta, J. P. W. Pluim, P. J. van Diest, and M. A. Viergever, "Breast Cancer Histopathology Image Analysis: A Review," IEEE Transactions on Biomedical Engineering, vol. 61, no. 5, pp. 1400–1411, May 2014, doi: 10.1109/tbme.2014.2303852.

[2] M. Long, "Artificial Intelligence in Medical Diagnosis | Healthcare Insights," Aidoc, Nov. 17, 2020. https://www.aidoc.com/blog/artificial-intelligence-medical-diagnosis/.

[3] J. Bresnick, "Top 12 Ways Artificial Intelligence Will Impact Healthcare," HealthITAnalytics, Apr. 30, 2018. https://healthitanalytics.com/news/top-12-ways-artificial-intelligence-will-impact-healthcare.

[4] "CNN Starter - NasNet Mobile (0.9709 LB)," kaggle.com. https://www.kaggle.com/CVxTz/cnn-starter-nasnet-mobile-0-9709-lb.

[5] "Understanding ResNet50 architecture," OpenGenus IQ: Learn Computer Science, Mar. 30, 2020. https://iq.opengenus.org/resnet50-architecture/.

[6] "Lymph," Wikipedia, Jul. 24, 2021. https://wikipedia.org/wiki/Lymph (accessed Aug. 04, 2021).

[7] "CAMELYON16 - Grand Challenge," grand-challenge.org. https://camelyon16.grand-challenge.org/Data/.

[8] B. S. Veeling, J. Linmans, J. Winkens, T. Cohen, and M. Welling, "Rotation Equivariant CNNs for Digital Pathology," arxiv.org, Jun. 2018, Accessed: Aug. 09, 2021. [Online]. Available: https://arxiv.org/abs/1806.03962.

[9] "Histopathologic Cancer Detection," kaggle.com. https://www.kaggle.com/c/histopathologic-cancer-detection/data.

[10] K. Team, "Keras documentation: Keras Applications," keras.io. https://keras.io/api/applications/.

[11] "ImageNet," www.image-net.org. https://www.image-net.org/.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," arXiv.org, 2015. https://arxiv.org/abs/1512.03385.

[13] Keiron Teilo O'Shea and R. Nash, "An Introduction to Convolutional Neural Networks," ResearchGate, Nov. 2015. https://www.researchgate.net/publication/285164623_An_Introduction_to_Convolutional_Neural_Networks.

[14] "tf.keras.metrics.SparseCategoricalAccuracy | TensorFlow Core v2.5.0," TensorFlow. https://www.tensorflow.org/api_docs/python/tf/keras/metrics/SparseCategoricalAccuracy (accessed Aug. 09, 2021).

[15] "Breast cancer - Diagnosis and treatment - Mayo Clinic," www.mayoclinic.org. https://www.mayoclinic.org/diseases-conditions/breast-cancer/diagnosis-treatment/drc-20352475. (accessed Aug. 04, 2021).

[16] "NCI Dictionary of Cancer Terms," National Cancer Institute, 2011. https://www.cancer.gov/publications/dictionaries/cancer-terms/def/metastasis