Journal of Clinical Neuroscience 66 (2019) 239–245

Contents lists available at ScienceDirect

Journal of Clinical Neuroscience

journal homepage: www.elsevier.com/locate/jocn

Experimental study

Automated brain histology classification using machine learning

Justin Ker [a], Yeqi Bai [b], Hwei Yee Lee [c], Jai Rao [a], Lipo Wang [b,*]

[a] Department of Neurosurgery, National Neuroscience Institute, 308433, Singapore
[b] School of Electrical and Electronic Engineering, Nanyang Technological University, 639798, Singapore
[c] Department of Pathology, Tan Tock Seng Hospital, 308433, Singapore

ARTICLE INFO

Article history:
Received 2 April 2019
Accepted 22 May 2019

Keywords:
Machine learning
Convolutional neural networks
Automated medical diagnosis
Brain histology
Glioma histology

ABSTRACT

Brain and breast tumors cause significant morbidity and mortality worldwide. Accurate and expedient histological diagnosis of patients' tumor specimens is required for subsequent treatment and prognostication. Currently, histology slides are visually inspected by trained pathologists, but this process is both time and labor-intensive. In this paper, we propose an automated process to classify histology slides of both brain and breast tissues using the Google Inception V3 convolutional neural network (CNN). We report successful automated classification of brain histology specimens into normal, low grade glioma (LGG) or high grade glioma (HGG). We also report for the first time the benefit of transfer learning across different tissue types. Pre-training on a brain tumor classification task improved CNN performance accuracy in a separate breast tumor classification task, with the F1 score improving from 0.547 to 0.913. We constructed a dataset using brain histology images from our own hospital and a public breast histology image dataset. Our proposed method can assist human pathologists in the triage and inspection of histology slides to expedite medical care. It can also improve CNN performance in cases where the training data is limited, for example in rare tumors, by applying the learned model weights from a more common tissue type.

1. Introduction

Breast and brain tumors cause significant morbidity and mortality worldwide. It is estimated that over 23,880 new cases of brain tumors and 268,670 cases of breast cancer will occur in 2018 for the United States alone [1]. Accurate histological diagnosis is crucial in order to determine prognosis and subsequent treatment, which may involve chemotherapy, radiotherapy or further surgery.

After initial medical history-taking, examination and radiological investigations, patients suspected of these tumors are subject to either biopsy or full surgical excision. In the case of breast biopsies, a punch biopsy can be obtained in the clinic outpatient setting, whereas brain tumors require an open or stereotactic procedure under general anesthesia. In both cases, the obtained tissue is sent to the pathology laboratory, where the tissue is smeared onto glass slides and subject to various chemical stains, hormonal, immunological or molecular assays. One of the most widely-used stains is the hematoxylin and eosin (H&E); hematoxylin is basic and stains acidic structures such as DNA/RNA of cell nuclei blue, while eosin is acidic and stains basophilic structures such as cytoplasmic proteins pink. Pathologists review these H&E glass slides under the microscope to distinguish cellular structures, and to look for distortions in normal cellular and tissue architecture that may signify the presence of neoplasm or cancer. Specifically, neoplastic features include strikeout increased mitotic figures in the nucleus, increased nuclear to cytoplasmic ratio, nuclear atypia, variation in nuclear size i.e. nuclear pleomorphism, lack of cellular differentiation, prominent nucleoli, areas of necrosis, abnormally increased formation of new blood vasculature, and invasion into surrounding tissue structures.

The identification of these visual features, allows pathologists to: i) classify the tissue as normal or tumor, ii) classify the grade of the tumor, which conveys the degree of the tumor malignancy, and iii) classify the cell origin of the tumor. In addition, by identifying the cell type on the histology slide, pathologists also determine if the tumor is a primary or metastatic tumor. Subsequent treatment is dependent on tumor-cell origin, tumor grade, and the degree of spread to other organs. In the context of brain tumors, this information guides subsequent clinical treatment. Intracranial germinomas are highly sensitive to radiotherapy and this is the preferred mode of treatment as opposed to full surgical excision; primary central nervous system lymphomas are treated

* Corresponding author at: 100A Eng Neo Ave, Singapore 289562, Singapore.
E-mail addresses: justin.ker@mohh.com.sg (J. Ker), BA0001QI@e.ntu.edu.sg (Y. Bai), hwei_yee_lee@ttsh.com.sg (H.Y. Lee), jai.rao@singhealth.com.sg (J. Rao), ELPWang@ntu.edu.sg (L. Wang).

https://doi.org/10.1016/j.jocn.2019.05.019
0967-5868/

with chemotherapy and radiotherapy; metastatic tumors can be treated with surgery, whole brain radiotherapy, or stereotactic radiosurgery.

Additional assays can test for the presence of hormonal, immunological, and molecular markers, such as estrogen-receptor positivity in the case of breast tumors, or isocitrate dehydrogenase (IDH) mutations in brain gliomas, to influence the choice of chemotherapy, or to provide a survival prognosis. For example, patients with brain glioma tumors and IDH mutations have a longer survival than patients whose tumors are wild-type [2]. Therefore, establishing an accurate tissue diagnosis is crucial for the patient, and dependent on the clinical acumen of an experienced pathologist.

The reading and diagnosing of tumors from histological slides is an intensive, manual task. Obstacles to accurate diagnosis include the experience of the pathologist, the time pressures faced in reporting a diagnosis, the subjectivity in identifying neoplastic features, the subjectivity of classification into tumor grades. One study demonstrated that the concordance between pathologists [3,4] to be extremely varied. With longer lifespans, more patients, and increased prevalence of cancers in much of the developed world, it is likely that these obstacles will only worsen in the future. It would be ideal to use a software algorithm to automate the identification of tumors and tumor cell origin, as well as the classification of tumors into various grades.

Automated computer vision is a branch of artificial intelligence that has seen dramatic successes in natural image classification in recent years. Early efforts in the 1960s were directed at attempting to extract three-dimensional geometric information from two-dimensional images, in order to recognize an object within an image. This evolved to bottom-up low-level edge and segment detectors, with later involvement of more top-down techniques that predicated on a certain degree of feature handcrafting with *a priori* knowledge of an object's structure [5].

In recent years, these techniques have given way to machine learning techniques using convolutional neural networks (CNN), after the successful application of CNNs in natural image recognition competitions, such as the annual ImageNet Large Scale Visual Recognition Challenge (ILSVRC). CNNs are a type of supervised machine learning artificial neural network. Their lineage can be traced back to the artificial neuron first described by McCulloch and Pitts in 1943 [6], and the perceptron reported by Rosenblatt [7]. Fukushima's 1982 publication of the Neocognitron has been recognized as a progenitor of a CNN [8]. Rumelhart, Hinton and Williams published their seminal work on back-propagation [9], a technique that was used to train the weights in artificial neural networks. Back-propagation improved the ability of artificial neural networks to learn useful features from input images, and has been an integral part of CNNs since. CNNs were then formalized by Lecun [10], who applied CNNs in one of the early practical computer vision applications to recognize hand-written characters on checks. Krizhevsky [11] used CNN to great success in the 2012 ILSVRC, and since then CNNs have become a main-stay in subsequent ImageNet competitions. The availability of large, labeled datasets and lowered costs of computational graphical processing units have spurred both the research and application of CNNs across many disciplines.

CNNs learn low level image features such as lines, curves and edges, before combining them into higher level features [12]. This is also thought to be the basis of how the mammalian visual system recognizes objects [13]. Recognition of progressively higher features then allows CNNs to perform accurate classification that has in many cases exceeded human performance in visual recognition and detection. In the medical imaging space, CNNs are uniquely suited to the analysis of medical images, which may be both two (e.g. X-ray images, histology slides, ultrasound scans) or three-dimensional in nature (Computed tomography scans, magnetic resonance imaging scans). The advantage of CNNs is that they perform automatic image feature extraction and classification without laborious manual feature hand-crafting. Moreover, spatial representations and relationships within the input image are preserved, which is useful when analyzing medical images as these invariably contain anatomical relationships. Indeed, CNNs have already exceeded human-level performance in diagnosing chest X-rays [14], diabetic retinopathy [15], skin lesions [16] and histology slides [17].

CNNs have been used to classify histology slides with success. Most reports involve the analysis of breast histology slides [17–19], possibly due to the availability of breast specimen histology datasets, and its comparatively high prevalence compared to tumors from other organs. Bejnordi [17] reported the results of the CAMELYON16 competition, which pitted machine learning networks against human pathologists in identifying metastases, and whole slide classification of axillary lymph node specimens of breast cancer patients. The training set consisted of 399 whole slide histological specimens. The best algorithm had an area under the curve (AUC) score of 0.994, outperforming the best human pathologist (AUC of 0.884). Instead of whole slide classification, Ciresan [20] used CNNs to detect the presence of mitotic figures on histology slides, with a dataset of 50 histology images containing 300 mitoses. The author reported state-of-the-art results, with recall and precision scores of 0.70 and 0.88 respectively. Ertosun [21] used an ensemble of two CNNs to classify histology slides of glioma brain tumors into different tumor grades: low grade glioma grade 2, low grade glioma grade 3, glioblastoma multiforme grade 4. The ensembled networks achieved an accuracy of 96% in classifying grade 4 tumors, and an accuracy of 71% when classifying grade 2 and 3 tumors.

Our objective in this work is to automatically analyze histology slides for tumor presence, tumor grade and the tumor cell origin. This is not a trivial task in medical image analysis. One potential method is to manually hand-craft features that are common to neoplastic cells, and for each tumor cell origin and tumor grade. An algorithm then could look for these specific features to aid in the task of classification. However, this method would be manually laborious and time-intensive. Importantly, the algorithm would only be able to recognize a specific set of features, and may miss features when there is cellular variability, since it is not possible to account for every histological manifestation of a tumor cell. For example, an algorithm may be trained to look for abnormal mitotic figures within a nucleus, that may signify neoplastic transformation, but these figures may present in variable numbers, shapes, sizes and stages of separation.

In *transfer learning*, the features learned from the training of a first artificial neural network are applied to the training of a second artificial neural network. This has several advantages: training time is potentially reduced, classification accuracy may be improved, and the limit of a small number of training samples is circumvented. Interestingly, even when the training samples in the first and second datasets are disparate, transfer learning is still beneficial. Transfer learning has been used in medical image analysis to improve network performance in detecting abnormal thoraco-abdominal lymph nodes on computed tomography (CT) scans, localizing kidneys on ultrasound images, and detecting abnormal polyps on colonoscopy videos [22–24]. These works reported that transfer learning from natural images generally improved network performance in classifying medical images. However, in the area of histopathological analysis, there have only been limited reports that studied transfer learning. Bayramoglu [25] examined how various neural networks (AlexNet, GenderNet, VGG-16, GoogLeNet) performed with transfer learning in classifying cell nuclei into four classes: epithelial, fibroblasts,

inflammatory and miscellaneous. Transfer learning improved network classification accuracy for all four networks, with AlexNet showing the greatest improvement, from 71% to 86%. Our work is distinct in that we apply transfer learning between different tumor types, as opposed between natural and biomedical images. Additionally, we classify portions of whole histological slides, instead of individual cell nuclei.

In order to accomplish our objectives in algorithmically classifying histology slides, we propose using a CNN to classify the histological images, and using transfer learning to improve its accuracy.

We see the techniques developed in this work as a tool to aid pathologists to reduce their work-load, by being both a triage and check of the histology slides read. An initial screen using software algorithms can detect the abnormal, neoplastic tissue on histology slides for further and closer inspection by human eyes. Algorithms can also be used to check that abnormal tissue slides are not missed. Improvements in the efficiency, accuracy and safety of such a health-care workflow will benefit patients, doctors, and the health system as a whole, with potentially reduced health-care costs.

The novel contributions of this work are:

1. We report the successful classification of brain tumors into normal brain tissue and two different tumor grades. Previous work did not include differentiation from normal brain tissue, which would be a crucial step in a practical, automated workflow to aid pathologists.
2. We demonstrate that transfer learning from brain tissue classification can improve network performance in breast tissue classification. This is crucial in histology slide analysis as datasets are often small in size, with certain tissue types being rare and therefore difficult to train using supervised machine learning networks. Our approach provides a technique to circumvent this obstacle by training on a different tissue type and applying the learned features to the target tissue histology slides.

2. Methods

In this work, we modify a well-known convolutional neural network, Google Inception V3, that was pre-trained on a subset of Imagenet images to perform four tasks. *Firstly*, we classify brain histology slides into normal brain or high grade glioma. *Secondly*, we classify the same slides into: normal brain, low grade glioma (LGG) or high grade glioma (HGG). *Thirdly*, we then classify breast histology slides into normal or carcinoma-in-situ, and repeat this task after applying transfer learning to improve the network performance. *Fourthly*, we perform and optimize binary classification of breast Carcinoma-In-Situ (CIS) and Brain HGG. The impetus for the latter task is to identify a primary tumor's cell type. The cell of origin of a primary tumor impacts treatment and prognosis. These are summarized below in Table 1.

Table 1
Summary of Tasks Performed.

Tissue	Classification Task
Brain histology	Task 1: Normal vs High grade glioma (HGG)
Brain histology	Task 2: Normal vs Low grade glioma (LGG) vs HGG
Breast histology	Task 3a: Normal vs Carcinoma-In-Situ (CIS)
Breast histology	Task 3b: Normal vs Carcinoma-In-Situ (CIS), after transfer learning from brain tissue
Brain and breast histology	Task 4: Brain HGG vs Breast CIS

2.1. Dataset

2 datasets are employed for experiments in this paper. The first dataset is made up of 3 classes of H&E stain images from brain tissue, which contains 50 normal images, 45 low grade glioma images, and 59 high grade glioma images. The size of each image in the first dataset is 1600 * 1200 pixels. This dataset was obtained from the Department of Pathology at Tan Tock Seng Hospital, Singapore, after the appropriate institutional research approvals were obtained (see Table 2).

The second dataset is made up of 2 classes of H&E stain images from breast tissue, which contains 55 normal images and 63 Carcinoma-In-Situ. The size of each image in the second dataset is 2048 * 1536 pixels. We used two of the four classes of the breast histology images used by Araújo et al. [18]. The datasets are summarized in Fig. 1.

2.2. Data preprocessing

For the brain histology images, we augmented the original images with random cropping, to generate a training set between 3000 and 4000 images and a test set of 1000 images per class. For the breast histology images, random cropping was also used as the data augmentation method to generate a training set of 200 images and a test set of 50 images. For both brain and breast images, the training data and test data images were then normalized into the interval [0, 1].

2.3. Model, training and implementation

In this work we modified a well-known CNN classifier, Inception V3 [26], that was pre-trained on Imagenet images. The last 4 layers of this model were removed, replaced with a global average pooling layer, and four fully connected (ReLU) layers, followed by either 2 or 3 softmax-activated neurons for two or three class classification respectively. For task 2 (Normal vs LGG vs HGG), we made one further modification by ensembling two models before final classification. Images were sent to a first CNN that differentiated between Normal and Abnormal brain images (i.e. LGG and HGG). Thereafter, the abnormal images were classified into LGG or HGG.

For all 4 tasks, we used cross entropy as the loss function and RMSProp as the optimizer. The learning rate was initialized at 0.001, batch sizes varied from 8 to 16. Dropout was set at a fraction of 0.25 to prevent over-fitting. The above was implemented on a desktop workstation with an Intel Xeon E5 3.7 GHz Quad-core processor, 16 GB DDR4 RAM, and one NVIDIA Geforce GTX 1070 graphics card with 8 GB RAM.

2.4. Evaluation metrics

To assess the accuracy of our network performance, we use metrics common in machine learning evaluation: recall, precision

Table 2
Datasets of H&E images used.

Class	Normal brain	Brain low grade glioma (LGG)	Brain high grade glioma (HGG)	Normal breast	Breast Carcinoma-In-Situ (CIS)
Number of original images	50	45	59	55	63

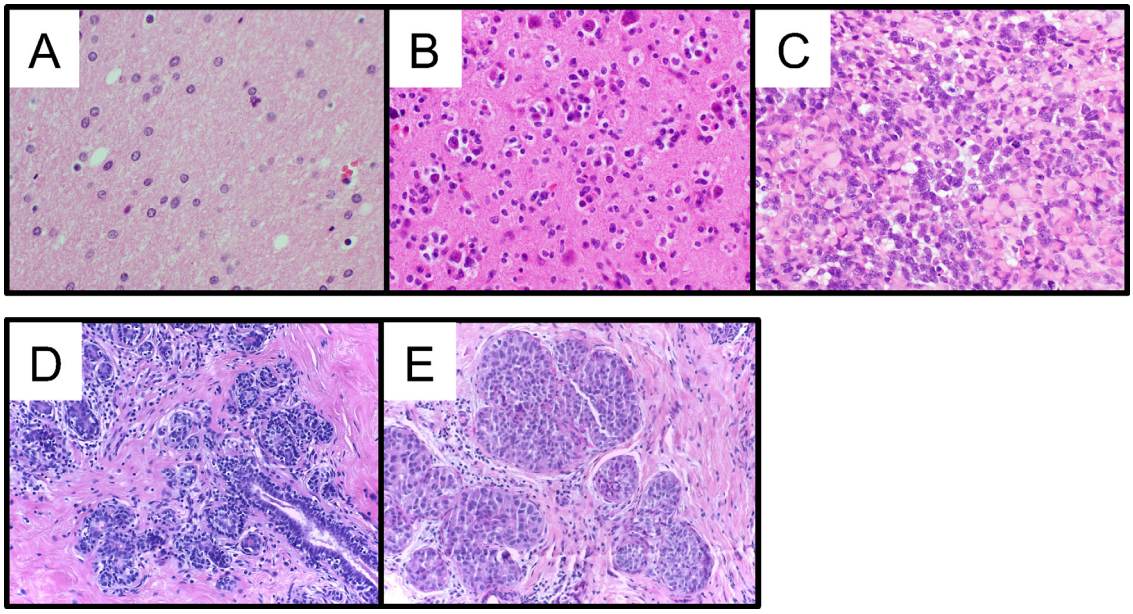

Fig. 1. Histology slides of brain and breast specimen. A, Normal brain. B, Brain Low grade glioma (LGG). C, Brain high grade glioma (HGG). D, Normal breast. E, Breast Carcinoma-In-Situ (CIS).

and the F1 score. R, Recall and P, Precision, and the F1 score are given below. ACC refers to accuracy. TP, TN, FN, FP represent True Positive, False Negative and False Positive respectively.

$$R = \frac{TP}{TP + FN}, \; P = \frac{TP}{TP + FP} \tag{1}$$

$$F_1 = \frac{2}{\frac{1}{Recall} + \frac{1}{Precision}} \tag{2}$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

3. Results

We carried out various binary and multi-class classification experiments as summarized in Table 1 on brain and breast histology images. In Task 1, binary classification was performed on normal and HGG brain histology images. The classifier was able to differentiate all normal from HGG images, and predicted all the test images correctly (Table 3). This raises concerns with overfitting, and drop-out was implemented in the model to address this. Differentiating normal brain from HGG is a straightforward task for a trained human pathologist, and it may be that the CNN is merely reflecting the dramatic differences in cell and tissue architecture between grossly normal and abnormal tissue. Even to an untrained human eye, the disparity between image A (normal) and image C (HGG) in Fig. 1 is obvious. Ertosun et al. [21] published similar work in using a CNN to classify brain tumors into HGG (Glioblastoma multiforme) vs LGG. This is conceivably a more difficult task than classifying normal and HGG, but they still demonstrated recall and precision scores of between 0.94 and 0.98. This suggests that our results in the comparatively easier task of differentiating normal and HGG to be plausible.

Table 3
Task 1: Normal brain vs HGG classifier.

	Predicted Normal brain	Predicted HGG
Actual Normal brain	1000	0
Actual HGG	0	1000

Table 4
Task 2: Normal brain vs LGG vs HGG classifier.

	Predicted Normal brain	Predicted LGG	Predicted HGG
Actual Normal brain	982	18	0
Actual LGG	0	980	20
Actual HGG	0	29	971

In Task 2, we performed three-class classification into normal brain, LGG and HGG. The results are summarized in Table 4, and the ensembled classifier demonstrates satisfactory performance, with the majority of predictions correct. The mistakes in classification match how a human pathologist could conceivably erroneously classify images as well. For the normal brain tissue images, a small number (18 out of 1000) were mistakenly classified as LGG. This can potentially happen in real-life clinical practice, as low grade gliomas are thought to be transformed from normal brain tissue. Therefore, it can be difficult on occasion to distinguish the two, as features of normal and LGG may be present in a LGG brain histology sample. This brings up a technical nuance that may mislead both a CNN classifier and a human pathologist: that on a whole histology slide, there may be areas of normal, low grade or high grade tumor growth. Generally, our CNN classifier predicts LGG and HGG accurately, with a small number of LGG predicted as HGG, and vice versa. Again, the lack of homogeneity in the tumor grade across the entire histology slide may possibly account for these erroneous predictions. We also carried out sub-group analysis to examine how the CNN classifier performed when the classification was binary. These results are summarized in Table 8. In Normal brain versus the combined LGG and HGG group, the F1 score was 0.991. In differentiating LGG and HGG, the F1 score was 0.976. The corresponding precision and recall were 0.980

and 0.971, which match the results obtained by Ertosun et al. [21] in a similar task.

For task 3a, we employed the CNN to classify between normal breast and breast CIS samples. In this task, the CNN performed poorly, with an F1 score of 0.547 (Table 5). The prediction accuracy was low, and we hypothesize that this may be due to the fact that breast tissue (both normal and abnormal) are much more heterogeneous than brain tissue samples. The original normal breast and CIS images contained not just cellular breast tissue areas, but fibrous and fatty areas as well. Breast CIS images (Fig. 1 Image E), and even the normal breast images (Fig. 1 Image D) appear much less uniform overall than normal and abnormal brain tissue (Fig. 1 Images A-C). We hypothesize that non-cellular fibrous and fatty areas, which appear in both normal breast and breast CIS images confound the CNN in classification. Interestingly, Araújo et al. [18] found that enlarging their dataset with more training examples lowered model performance accuracy, possibly due to the introduction of additional non-cellular areas which also introduce confounding features.

We therefore performed a further experiment, task 3b, which repeated the breast tissue classification with transfer-learning from the brain images. Specifically, after task 1 was carried out, the weights learned in the model were used in task 3b to classify normal breast and breast CIS images. Interestingly, this act of transfer learning from a different body organ assisted in improving the classification accuracy of breast tissue histology. With transfer learning, the precision, recall and F1 scores improved to 0.940, 0.887 and 0.913 respectively (from 0.580, 0.581 and 0.547). Accuracy improved from 52% to 91%. One possible explanation is that since the brain tissue images are mostly cellular, the weights learned in the model from task 1 are adapted to look for cellular architectures, which is where the feature differences between normal and tumor tissue (across all organs) lie. We also performed transfer learning in the reverse direction, i.e. using the weights learned from the breast classification model on the brain classification model. However, we found there to be a decrease in the brain CNN accuracy and performance. This suggests that for the benefit of transfer learning to be realized, the model weights to be transferred must be from a CNN that already has good performance. This also suggests that the learned CNN features may be tissue-agnostic, meaning that the cellular features of malignancy (cellular atypia, nuclear pleomorphism etc.) are learned, rather than innate tissue brain or breast features. The advantage is that a well-tuned CNN after transfer learning can be applied to a wide variety of tissue types, even rare tumors with extremely limited datasets (see Table 6).

On task 4, the CNN was applied to differentiate between brain HGG and breast CIS tissue. The reason for performing this classification is that breast and other tumors (such as lung and colon tumors) metastasize to the brain. Metastatic brain lesions are far more common and exact a larger epidemiological burden on patients, than primary brain tumors (such as LGG, HGG). In a metastatic brain tumor, knowing the organ of origin can guide further treatment and prognosis, as tumors from different organs can have varying susceptibility to radiation and chemotherapy. When the organ of tumor origin is not clear or known, being able to discover the type of tumor origin can also guide where to look for the primary tumor, so that it can be treated. Table 7 demonstrates that the CNN is able to distinguish between brain and breast tumor perfectly.

Table 8 summarizes the precision, recall and F1 scores of tasks 1–4. In the brain tissue classification experiments, our results match those of Ertosun et al [21]. For breast tissue classification, Araújo et al. [18] performed both patch and image-wise classification using a CNN, and also a CNN coupled to a support vector machine (SVM). Their best result for patch-wise classification accuracy was 77%, and 77.8% for image-wise classification. In comparison, our CNN had an accuracy of 52%, and 91% when transfer learning from brain tissue classification was applied (F1 scores of 0.547 and 0.913 respectively).

Fig. 2 shows each of the different classes of tissue used in this work, with its corresponding activation map (of the 2nd Concatenate layer of the Inception V3 model) immediately to its right. Image A is normal brain tissue, Image B is LGG and Image C represents HGG. What is striking is that only the cellular areas on the histology slides seem to be activated most strongly (in white).

Table 5
Task 3a: Normal breast vs Breast Carcinoma In Situ (CIS).

	Predicted Normal breast	Predicted Breast CIS
Actual Normal breast	29	21
Actual Breast CIS	27	23

Table 6
Task 3b: Normal breast vs Breast Carcinoma In Situ (CIS), with transfer learning from brain images.

	Predicted Normal breast	Predicted Breast CIS
Actual Normal breast	47	3
Actual Breast CIS	6	44

Table 7
Task 4: Brain HGG vs Breast CIS.

	Predicted Brain HGG	Predicted Breast CIS
Actual Brain HGG	500	0
Actual Breast CIS	0	500

Table 8
Summary of Results of various classification tasks.

	Precision	Recall	F1 Score
Task 1 (Normal Brain vs. HGG)	1.0	1.0	1.0
Task 2 (Normal brain vs. LGG + HGG)	0.982	1.0	0.991
Task 3a (Normal Breast vs Carcinoma-In-Situ)	0.580	0.518	0.547
Task 3b (Normal Breast vs Carcinoma-In-Situ with transfer learning)	0.940	0.887	0.913
Task 4 (Brain HGG vs Breast CIS)	1.0	1.0	1.0

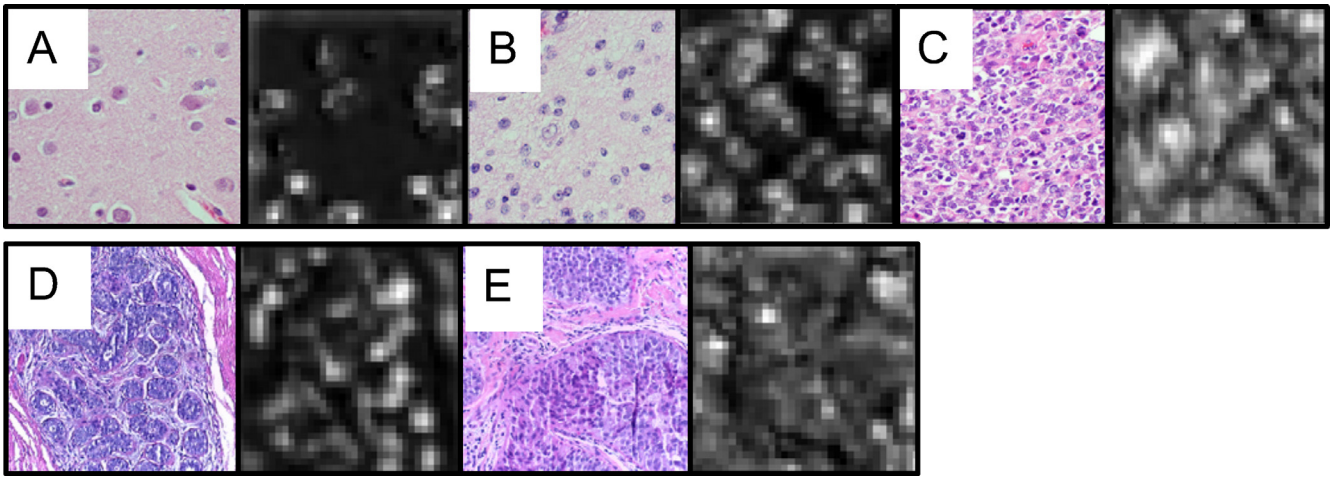

Fig. 2. Brain and breast histology images with the corresponding activation maps of the 2nd Concatenate layer of the Google Inception V3 model. A, normal brain. B, LGG. C, HGG. D, normal breast. E, breast CIS. There is a general trend of cellular areas lighting up as white areas on the activation maps, and of increasing activation with increased cellular density and tumor grade.

Across the progression from normal to LGG and then HGG, there is increased cellularity on the histology slides, with a corresponding density of activation seen on the activation maps. This suggests that the CNN extracts cellular features and uses them for classification, in a similar way to how a human pathologist performs the same visual task. This is startling because unlike a human pathologist, the CNN has no conceptual understanding of what constitutes a cell, and of how a mass of growing cells can represent a tumor mass of increasing grade. Image D represents normal breast tissue and image E is breast CIS. Cellular areas light up in the breast histology activation maps, although these are not as distinct when compared to the brain histology images. Similarly, acellular regions are black on the breast histology activation maps.

Overall, our experiments confirm again the benefit of transfer learning for CNN-based medical image classification. Further, we found that transfer learning, even when applied across different organ systems, to be beneficial in terms of classifier performance. This work was implemented with modest computing resources, and the performance of this system could be easily replicated in a real clinical setting. CNN classifier accuracy was high and training the model took minutes. This study was limited by the relatively small training set, which raises the possibility that over-fitting may contribute some part to the results. We employed the standard technique of dropout to mitigate against this. The next step in this study is to apply these findings in a real hospital scenario to benefit both pathologists and patients.

4. Conclusion

This work presents the implementation of a convolutional neural network to classify human histology images. Digital brain histology images were classified into normal, low grade and high grade glioma. Breast histology images were classified into normal and carcinoma-in-situ. We also demonstrate for the first time, that transfer learning across different body tissues can help improve classification accuracy. Specifically, transfer-learning from brain histology images aided the classification of breast histology images. We present a straight-forward implementation, without requiring expert human pathologist input nor extravagant computing resources. Results of our classification experiments match or exceed those in similar publications. From our work, clinicians can implement an automated histology processing pipeline to assist pathologists, so as to provide fast and automated medical diagnosis.

Funding

National Neuroscience Institute-Nanyang Technological University Neurotechnology Fellowship.

Declaration of Competing Interest

None.

References

[1] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin 2017;67(1):7–30.
[2] Cohen AL, Holmen SL, Colman H. IDH1 and IDH2 mutations in gliomas. Curr Neurol Neurosci Rep 2013;13(5):345.
[3] van den Bent MJ. Interobserver variation of the histopathological diagnosis in clinical trials on glioma: a clinician's perspective. Acta Neuropathol 2010;120(3):297–304.
[4] Elmore JG, Longton GM, Carney PA, Geller BM, Onega T, Tosteson AN, et al. Diagnostic concordance among pathologists interpreting breast biopsy specimens. JAMA 2015;313(11):1122–32.
[5] Huang T. Computer vision: Evolution and promise. 1996.
[6] McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity, 1943. Bull Math Biol 1990;52(1–2):99–115.
[7] Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain. Psychol Rev 1958;65(6):386.
[8] Fukushima K, Miyake S. Neocognitron: a self-organizing neural network model for a mechanism of visual pattern recognition. Competition and cooperation in neural nets. Compet Cooperation Neural Nets 1982.:267–85.
[9] Rumelhart DE, Hinton GE, Williams RJ. Others. Learning representations by back-propagating errors. Cogn Model 1988;5:1–3.
[10] LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, et al. Backpropagation applied to handwritten zip code recognition. Neural Comput 1989;1(4):541–51.
[11] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 2012:1097–105.
[12] Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives. IEEE Trans Pattern Anal Mach Intell 2013;35(8):1798–828.
[13] Hubel DH, Wiesel TN. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol 1962;160(1):106–54.
[14] Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, et al. CheXNet: radiologist-level pneumonia detection on chest X-rays with deep learning. arxiv preprint. 2017.
[15] Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 2016;316(22):2402–10.
[16] Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks Available from:. Nature 2017;542(7639):115–8. https://doi.org/10.1038/nature21056.
[17] Bejnordi BE, Veta M, Diest PJV, Ginneken BV, Karssemeijer N, Litjens G, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA 2017;318(22):2199–210.

[18] Araújo T, Aresta G, Castro E, Rouco J, Aguiar P, Eloy C, et al. Classification of breast cancer histology images using convolutional neural networks. PLoS One 2017;12(6):e0177544.
[19] Cireşan DC, Giusti A, Gambardella LM, Schmidhuber J. Mitosis detection in breast cancer histology images with deep neural networks. In: International Conference on Medical Image Computing and Computer-assisted Intervention. p. 411–8.
[20] Chen H, Dou Q, Wang X, Qin J, Heng PA, et al. Mitosis detection in breast cancer histology images via deep cascaded networks. AAAI 2016:1160–6.
[21] Ertosun MG, Rubin DL. Automated grading of gliomas using deep learning in digital pathology images: a modular approach with ensemble of convolutional neural networks. AMIA Ann Symp Proc. 2015;2015:1899.
[22] Hoo-Chang S, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans Med Imaging 2016;35(5):1285.
[23] Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, et al. Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Trans Med Imaging 2016;35(5):1299–312.
[24] Ravishankar H, Sudhakar P, Venkataramani R, Thiruvenkadam S, Annangi P, Babu N, et al. Understanding the mechanisms of deep transfer learning for medical images; 2016.
[25] Bayramoglu N, Heikkilä J. Transfer learning for cell nuclei classification in histopathology images. Eur Conf Comput Vis 2016:532–9.
[26] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition. p. 2818–26.