Inceptionism and Residualism in Classification of Breast Fine Needle Aspiration Cytology Cell Samples

Amartya Ranjan Saikia1,*, Lipi B. Mahanta1, Kangkana Bora1 and Anup Kumar Das2
1Centre for Computational and Numerical Studies, Institute of Advanced Study in Science and Technology, Guwahati, India
2Ayursundra Healthcare Pvt. Ltd, Guwahati, India
*Corresponding author email id: ar5saikia@gmail.com

Abstract

Fine needle aspiration cytology (FNAC) entails using a narrow gauge (25-22G) needle to collect a sample of a lesion for microscopic examination. It allows a minimally invasive, rapid diagnosis of tissue but does not preserve its histological architecture. FNAC is commonly used for diagnosis of breast cancer, with traditional practice being based on the subjective visual assessment of the breast cytopathology cell samples under a microscope to evaluate the state of various cytological features. Therefore, there are many challenges in maintaining consistency and reproducibility of findings. However, the advent of digital imaging and computational aid in diagnosis can improve the diagnostic accuracy and reduce the effective workload of pathologists. This paper presents a comparison of various deep convolutional neural network (CNN) based fine-tuned transfer learned classification approach for the diagnosis of the cell samples. The proposed approach has been tested using VGG16, VGG19, ResNet-50 and GoogLeNet-V3 (aka Inception V3) architectures of CNN on an image dataset of 212 images (99 benign and 113 malignant) samples shown in Figure 1, later augmented and cleansed to 2120 images (990 benign and 1130 malignant), where the network was trained using images of 80% cell samples and tested on the rest. This paper presents a comparative assessment of the models giving a new dimension to FNAC study where GoogLeNet-V3 (fine-tuned) achieved an accuracy of 96.25% which is highly satisfactory.

Keywords: Deep Learning, Convolutional Neural Network, Breast Cancer, FNAC.

1. Introduction

Breast malignancy is the second most common type of malignancy after lung cancer in general, second most in women after skin cancer and the fifth most common cause of cancer death worldwide [38]. Breast cancer is also prevalent among men and according to new statistics, about 2,550 new cases of invasive breast cancer are expected to be diagnosed in men in 2018 [39]. According to the World Health Organization projections, breast cancer caused 559,000 deaths worldwide in the year 2008 [22]. The incidence rate of breast cancer is increasing rapidly in developing countries. India being under the radar of high incidence, there was an ardent need to work on this area with cutting edge deep learning approaches. Fine needle aspiration cytology (FNAC) is one of the most commonly used pathological investigations for screening, and diagnosis of breast cancer. The traditional practice of breast FNAC is based on subjective assessment. Here, microscopic appearance of the aspirates is visually evaluated on various cytological criteria. Therefore, there are many challenges in maintaining
consistency and ensuring reproducibility in findings are inevitable. Moreover, with inadequate or non-representative sampling may leading to equivocal diagnosis there exists an overlap in the state of various cytological criteria in benign and malignant lesions [7, 18]. Advancement in AI, digital imaging and computational aid in diagnosis can help to improve the diagnostic accuracy and to reduce the effective workload of a pathologist. In this regard, researchers and practitioners of pathology have been using quantitative analysis for computer-aided diagnosis (CAD) of pathology samples including breast FNAC [6, 16].

Figure 1. FNAC cell samples

This paper implements and analyses several deep convolutional neural network (CNN) based classification models for the diagnosis of the breast FNAC cell and presents a comparison-based study of the same. Finally, the best model is proposed for further use or study.

Figure 2. Overview of the approach

Figure 2 summarizes the contribution of this paper including the creation of image dataset of breast cytopathology samples and CNN based breast FNAC cell sample classification. Limited data samples collected did not led to much of a problem as various data augmentation and cleansing methods were employed. Comparison of various segmentation methods and moving forward with the promising one, after the formation of a proper dataset enabled us to move forward with our research. The paper is organized as - Section 2 describes the prior art for computer vision, and machine learning techniques used in breast cancer diagnosis by FNAC image analysis and deep learning techniques used in cytopathology image analysis, Section 3 describes the image dataset developed by us and used during experimentation for this paper, Section 4 describes our methodology, Section 5 presents details of the experimental setup used, Section 6 presents classification results obtained along with discussion of the findings. Conclusions for the study are presented in Section 7.

2. Prior Art

Advancement in Machine Learning and AI has shown us new promising paths in the field of medical imaging. Researchers and practitioners of pathology have been using quantitative analysis to improve diagnostic accuracy and to reduce the effective workload of a pathologist. With recent developments in cost-effective and high-performance computer technology, the digital pathology has become amenable to the application of quantitative analysis in the form of decision support systems [30,
The CAD systems for digital pathology applications are being developed and deployed for some time now [6, 16]. Review of the prior art shows that computer vision systems commonly used in cytological diagnosis apply the bottom up approach of diagnostic reasoning from evidence to hypothesis [23]. It involves segmentation of primitives such as clusters, cells, and nuclei using image processing and segmentation techniques [10]; quantification of diagnostically significant cytological criteria [13] using techniques that extract morphometric, densitometric, textural and structural features [24]; followed by pattern recognition techniques for prediction of abnormalities and anomalies [20]. Recently deep neural networks like autoencoders [32] are increasingly finding their way into solving whole slide cytopathology image analysis challenges while jointly learning the representative feature space and classification margin. Some contributions related to radiological and interventional image analysis include [1, 21, 26]. The prior art [27] on the breast cancer histopathological image classification dataset (BreakHis database) [28] had earlier used an AlexNet [19] based CNN architecture for classifying whole slide histopathology and achieved 84-90% accuracy. Das et. al [5] in their approach combine predictions by transfer learned GoogLeNet [29] architecture of a CNN for random multiple images of a breast histopathology sample acquired at multiple magnifications to arrive at whole slide diagnosis achieving 94.67% accuracy in multifold validation over the BreakHis database. H Garud et. Al [30] achieved 89.7% mean accuracy in 8-fold validation using GoogLeNet architecture of CNN on an FNAC image dataset.

3. Breast FNAC Image Dataset

For this study FNAC cell sample images were generated at Ayursundra Healthcare Pvt. Ltd, Guwahati, India. Experienced doctors collected the sample from patients and prepared the slides. The processing and staining of cell samples was done at their respective laboratories. The cell samples were collected from patients following all ethical protocols. Images were captured by us using Leica ICC50 HD microscope using 400× resolution and 24 bits color depth, twenty images were captured per slide. Digitized images captured were then reviewed by experienced certified cytopathologists and ten best images were selected per slide. Figure 3 and Figure 4 are Benign and Malignant cell samples respectively.

![Figure 3. FNAC Benign cell samples](image1)

![Figure 4. FNAC Malignant cell samples](image2)

This comprised the FNAC image smear level database containing 212 images (99 Benign and 113 Malignant). Ground truth is then confirmed by a pathologist by marking the cell sample images as ‘Benign’ or
‘Malignant’. Along with the images the corresponding original reports of the patients were also collected which are later used for labelling of the images. For smear level study of the samples with Deep learning, we require more data than the current count to feed into the CNN models. Since, limited by cell samples, data augmentation was the prime focus to be implemented on the current data. Augmentation techniques, such as, Cropping, Shearing, Rotation, Mirroring, Skewing, Inverting and Zooming were implemented and the new count stood to 2120 images (990 Benign and 1130 Malignant). The image data count being better than before, was cleaned now to focus specifically on the Region of Interest (ROI), i.e. the nucleus. An image was split into its RGB and YMCK color channels and we proceeded only with the red channel of the image, as shown in Figure 5, due to its promising clear boundary highlighting only the ROI.

Several other methods including scharr derivative and sobel derivative, were implemented but selection of only the red channel stood out as the best option.

![Figure 5. Red channel Fig (v) Scharr derivative](image)

The actual data is in the ROI, i.e., the nucleus. The Cytoplasm and Red blood cell in the images are noise and hence needs to be removed completely. The images are segmented to achieve the goal and a comparison is made in the Figure 6. The second image of Figure 6 is the result of Otsu Threshold (Yellow) and has better segmentation output than the first image of Figure 6 General Threshold (White), due to its clear visual boundary around the nucleus and better ROI selection. Hence Otsu’s threshold was selected for segmenting the whole dataset.

![Figure 6. Comparing of general threshold vs Otsu’s threshold](image)

4. Training CNN for breast FNAC image classification

In this study, several deep learning models, viz. VGG16 [35], VGG19 [35], ResNet50 [36] and GoogLeNet-V3 [37], were trained to represent breast FNAC features. Fine-tuning of the classification margin were done for these models. Transfer learning from imagenet was deployed to transfer learned features along with the newly trained features from the FNAC cell sample dataset. As for Inception-V3, it is a variant of Inception-V2 [37, 40] which adds BN-auxiliary. BN auxiliary refers to the version in which the fully connected layer of the auxiliary classifier is also-normalized, not just convolutions. GoogLeNet-V3 performs better than other models due to smaller kernels, efficient grid size reduction and presence of several inception grid. The training results are as follows:

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>VGG16</td>
<td>63.20%</td>
</tr>
<tr>
<td>VGG16 fine-tuned</td>
<td>88.67%</td>
</tr>
<tr>
<td>VGG19</td>
<td>73.82%</td>
</tr>
<tr>
<td>VGG19 fine-tuned</td>
<td>88.20%</td>
</tr>
<tr>
<td>ResNet50</td>
<td>85.61%</td>
</tr>
<tr>
<td>ResNet50 fine-tuned</td>
<td>90.56%</td>
</tr>
<tr>
<td>GoogLeNet-V3</td>
<td>71.88%</td>
</tr>
<tr>
<td>GoogLeNet-V3 fine-tuned</td>
<td>96.25%</td>
</tr>
</tbody>
</table>

Table 1. Classification accuracy of different CNN models
Figure 7 shows the training vs validation accuracy graph of fine-tuned VGG16 with ‘relu’ activation and ‘adadelta’ optimizer. For GoogLeNet-V3 fine-tuned model, the hyperparameters in the trained model were:

- batch size – 32,
- number of epochs – 12,
- sgd learning rate - 1e-4,
- momentum – 0.9,
- transformation ratio – 0.05,
- training sample - 900 and validation sample – 90.

5. Experimental Setup

Experimentations were carried out in Intel Optimized Python 3 with Keras (Intel Optimized Tensorflow backend) using Intel® AI DevCloud, a cluster comprising of Intel Xeon Gold 6128 processors. First an image repository was created where all the collected images were stored in 2 different directories which represents 2 different classes. Images were arranged according to the reports of the patients collected and with proper consultation with the pathologist. The accuracy of classification was evaluated. The accuracy is the proportion of true results (both true positives and true negatives) among the total number of cases examined. Higher the accuracy, higher the rate of truly classified classes.

6. Results and Discussion

Results for the multi-experiment classification, as described in Section 4 are presented in Table 1. The table presents classification accuracy performances across each experiment along with the details of training and test data in each experiment. From the results, it can be observed that in general GoogLeNet-V3 [37] can learn visual features and classification margin from different training samples during transfer learning and achieves the smear level classification accuracy of 96.25%. The fine-tuned VGG16 achieved 90.09%, fine-tuned VGG19 achieved 72.53%, and ResNet50 achieved 89.15%. This accuracy is
comparable with the accuracy of 89.7% mean accuracy in 8-fold validation from [34]. However, it is less than the classification accuracies achieved with conventional approach of breast cancer diagnosis, which achieves up to 99.80% accuracy [2]. This proves that GoogLeNet-V3 is by far the best deep learning method for FNAC cell image classification.

7. Conclusion

In this paper, we presented an FNAC breast cytopathology samples image dataset along with the performance of VGG16, VGG19, ResNet50 and GoogLeNet-V3 CNN architecture in breast FNAC cell sample diagnosis in malignant or benign categories. It was observed that in general GoogLeNet-V3 can learn visual features and classification margin from different training samples included in the dataset and achieves the smear level classification accuracy of 96.25%, which is less than the classification accuracies achieved with conventionally defined ‘Computed Feature Dataset’ and statistical classifiers. It is known, computed feature dataset is slave of segmentation technique used so the results used for one dataset may not show consistent results on other dataset, but deep learning technique learn the features itself and user do not have to bother and spend too much time on significant feature selection. This proves that GoogLeNet-V3 is by far the best deep learning method for FNAC cell image classification. The proposed scheme can be considered as a baseline for future research. Data augmentation by adding more samples and data replication, transfer learning along with better CNN models and hyper parameter tweaking can be used to improve results.

References:
[34] Garud et al, High-magnification Multi-views Based Classification of Breast Fine Needle Aspiration Cytology Cell Samples using Fusion of Decisions from Deep Convolutional Networks, DOI: 10.1109/CVPRW.2017.115