Breast Cancer detection Using Convolutional Neural Networks

Zahirul Islam
Dept. of CSE
United International University
Dhaka, Bangladesh
ID: 011181211

Khorshed Alam
Dept. of CSE
United International University
Dhaka, Bangladesh
ID: 011181173

***Abstract*—Breast cancer is one of the major public health issues and is considered a leading cause of cancer-related deaths among women worldwide. Its early diagnosis can effectively help in increasing the chances of survival rate. However, the histopathological analysis of breast cancer is non-trivial, labor-intensive, and may lead to a high degree of disagreement among pathologists. This paper presents a** Convolutional Neural Networks **deep learning approach for the definite classification of non-carcinoma and carcinoma breast cancer histopathology images using our collected image dataset. These experimental results demonstrated that our proposed deep learning approach is effective for the automatic classification of complex-natured histopathology images of breast cancer, more specifically for carcinoma images.**

Keywords—Breast cancer, Histopathological images, Image classification, Convolutional Neural network, Image Dataset

I. Introduction

Breast cancer is still one of the top leading causes of death and common cancer among women worldwide, which is ranked the second-place following lung cancer in the survey about cancer mortality in recent years. Among the breast cancer subtypes, invasive ductal carcinoma is the most common type of breast cancer and accounts about 70- 80% of all breast cancer diagnoses. In every year, more than 180,000 women in the United States find out they have invasive breast cancer. Nonetheless, breast cancer is the most prevalent type of cancer and early diagnosis and its treatment highly improves the chances of survival. Breast cancer develops from breast tissue identified by lump in the breast and there are some changes in normal conditions. Clinical screening includes mammography, breast ultrasound, biopsy and other methods. A biopsy is the only diagnostic procedure that can definitely determine if the suspicious area is cancerous. The pathologists diagnose by visual inspection of histological slides under the microscope, which is considered as the confirmatory gold standard for diagnosis. However, the traditional manual diagnosis needs intense workload by experts with expertise. Diagnostic errors are prone to happen with the pathologists that have not enough diagnostic experience. It is shown that the use of Computer-aided diagnosis to automatically classify histopathological images can not only improve the diagnostic efficiency, but also provide doctors with more objective and accurate diagnosis results.
Recently, deep learning methods have made considerable progress and achieved remarkable performance in the field of computer vision and image processing, which has inspired many scholars to apply this technique to histopathological image classification. Convolutional neural networks are the most widely used type of deep learning network, and they perform equally well on image classification and image feature extraction. These results have laid the foundation for the application of the CNN in histopathological image classification.

In this paper, we propose a method that extracts richer multilevel features and integrates the advantages of the CNN thus, the short-term and long-term spatial correlations between patches are preserved. We first split the high-resolution pathology images into small patches. Then, the CNN is used to extract the richer multilevel image features of each patch. Finally, the RNN is used to fuse the patch features to make the final image classification. we obtained an average accuracy of 86.3%.

II. Proposed

A. Dataset

Invasive Ductal Carcinoma is the most common subtype of all breast cancers. To assign an aggressiveness grade to a whole mount sample, pathologists typically focus on the regions which contain the IDC. As a result, one of the common pre-processing steps for automatic aggressiveness grading is to delineate the exact regions of IDC inside of a whole mount slide.
The original dataset consisted of whole mount slide images of Breast Cancer specimens scanned at 40x.

From that, 277,524 patches of size 50 x 50 were extracted (198,738 IDC negative and 78,786 IDC positive). Each patch's file name is of the format: uxXyYclassC.png — > example 10253idx5x1351y1101class0.png . Where u is the patient ID (10253idx5), X is the x-coordinate of where this patch was cropped from, Y is the y-coordinate of where this patch was cropped from, and C indicates the class where 0 is non-IDC and 1 is IDC.

.

B. Diagram

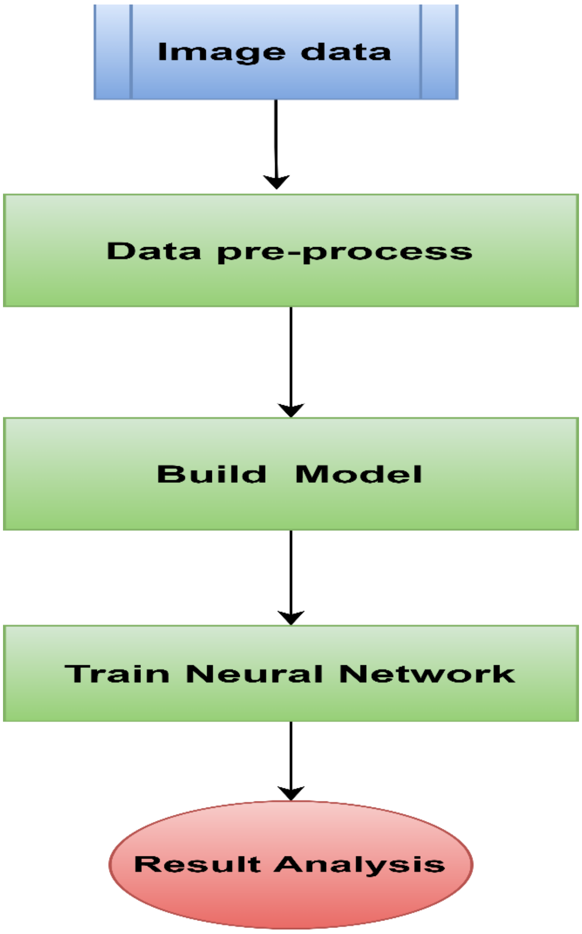

Fig-1

III. METHOD

All the patches are in BGR pixel format and are scaled from 0 to 255. We intended to apply machine learning classification methods to these images. Therefore, we made the scale between 0 and 1 to be compatible with the methods.

In this data preprocessing we use a sequential model. When an image with high resolution 50×50 Learning from class-imbalanced data continues to be a common problem in supervised learning and poses a significant challenge for the training of a deep learning model. Most of the standard classification algorithms are designed to handle balanced class distributions. However, in many real world applications, imbalance data sets are common in many practical applications such applications include fraud detection in banking, rare medical diagnoses, and oil spill recognition in satellite images. Moreover, in the medical applications, the correct classification for the minority class samples is more valuable than that for the majority class samples. For example, in a binary classification within the medical domain, if we are interested in classifying a patient's condition, the diagnosis of cancer of a patient, the minority class is more interested than the majority class of cancer free patients.

Many techniques have been proposed to mitigate the bias towards the plentiful classes in the problem of class imbalance. Techniques aimed at improving classification in the presence of class imbalance, the machine learning community has addressed the issue of class imbalance in two ways: sampling-based approaches and cost function-based approaches. Sampling based can be divided into three broad categories: a) over sampling b) under sampling c) hybrid of oversampling and under-sampling.

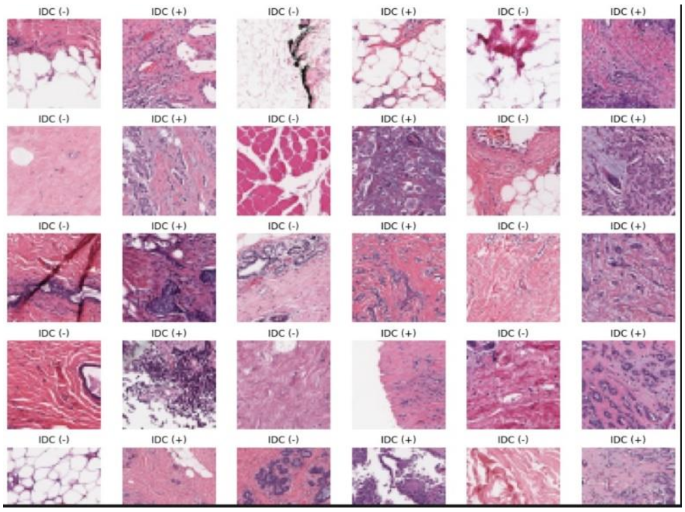

Fig-2

The network consists of two Convolution2D layers and a MaxPooling2D layer following after the second convolution. To regularize the model, a dropout layer was applied after the pooling layer, and after the first Dense layer. Afterwards, the output of the second pooling layer is flattened to 1D and passed through two fully connected (Dense) layers. ReLU activations will once again be used for all layers except the output dense layer, which will use a softmax activation. The remainder of the model specification, the cross-entropy loss function as the objective to optimize, and Adadelta optimiser for gradient descent. In addition, the model includes ImageDataGenerator API in Keras for generating existing training images and create many altered versions of the same image. This provides more images to train on, but can also help expose our classifier to a wider variety of lighting and coloring situations so as to make our classifier more robust.

A. *Related work*

Breast cancer histopathological image classification using a hybrid deep neural network in this paper they have used hybrid convolutional and recurrent deep neural network for breast cancer histopathological image classification and the dataset is breakhis. Accuracy is 85% .

Breast Cancer detection Using Convolutional Neural Networks for Mammogram Imaging System have used a new hybrid convolutional and recurrent deep neural network and dataset is histopathological image and their accuracy is 91.3%.

Breast cancer detection, Deep learning, CNN, MRI have used LSTM and dataset is TCIA public dataset. Accuracy is 97.20% .

Histology images,CN N,Enginee red Features, A bag of words uses dataset is BreaKHis and uses algorithms aresCNN using BVLC caffe . Accuracy is 96.15%.

IV. RESULTS

In this section, we present the performance of our proposed algorithm on our released dataset. Our proposed algorithm accuracy is 90.05%

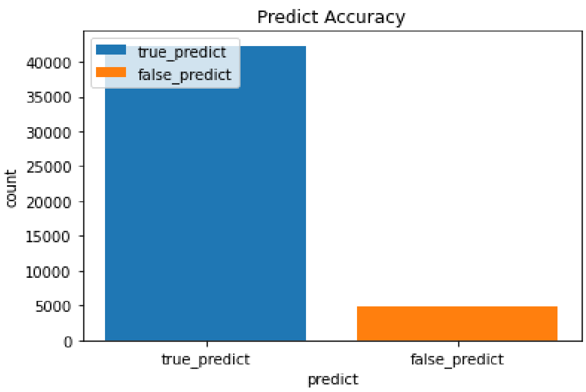

Fig-3

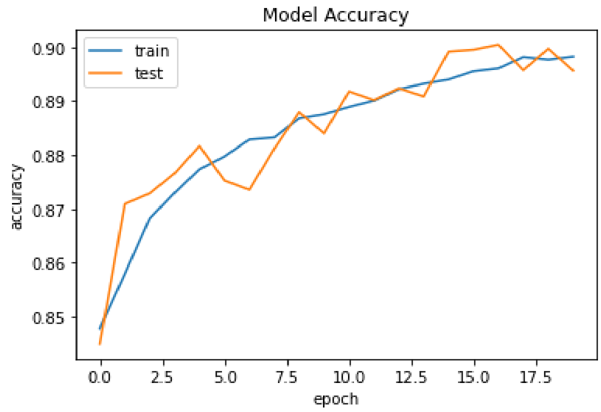

Fig-4

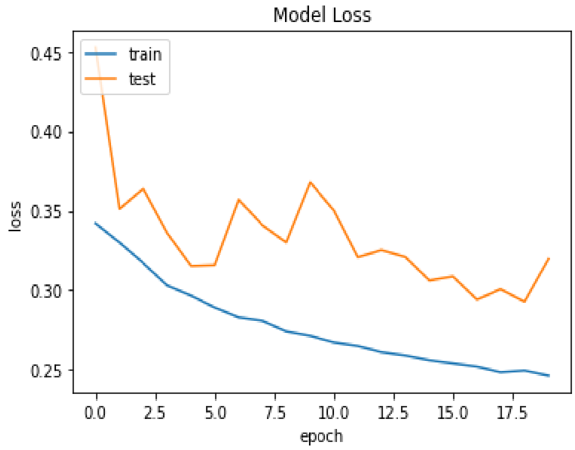

Fig-5

A. *Conclusion*

In this paper, we proposed a new method for breast cancer image classification using Convolutional Neural Networks. We have examined the impact of class imbalance on classification performance of convolutional neural networks and investigated the effectiveness of addressing the issue. Using CNN Model achieves above 90.05% accuracy. We hope that the dataset can serve as a benchmark to facilitate a broader study of deep learning in the field of breast cancer pathologic images.

In conclusion, breast cancer detection by using Convolutional Neural Network has been successfully developed and tested with 277000 Histopathology images. This method provides a fast diagnosis time and high accuracy system.