Grant recipient
Associate Professor Ramakrishnan Mukundan
University of Canterbury
Computer Science and Software Engineering
mukundan@canterbury.ac.nz
033692201
Private Bag 4800, Christchurch. 8014

Grant details

GRANT TYPE	Grant in Aid	FUNDING ROUND	2018 Grant In Aid
GRANT REFERENCE	GIA1 2018	GRANT AMOUNT	$5,000

Final report

1. Scientific Assessing Committee report

Advanced Deep Learning Algorithms for Automated Breast Cancer Histology Slide Evaluations

Ramakrishnan Mukundan (Principal Investigator)
Prerna Singh (Research Student)

Summary

The Canterbury Medical Research Foundation Grant in Aid helped us to carry out an extensive study of machine learning algorithms for automatic classification and HER2 grading of breast cancer histology slides using whole slide images (WSI). Specifically, we developed Neural Network based classification algorithms using both intensity and texture features for efficient classification of histopathological slides. The performance evaluation of the algorithm gave 91% accuracy when tested with a large dataset as outlined below. As part of this project, we also wrote a research paper [1] describing our algorithm and experimental analysis, which have been accepted for an oral presentation at the IEEE 20th International Workshop on Multimedia Signal Processing (MMSP-2018) to be held at the Simon Fraser University, Vancouver, Canada from 29th to 31st August 2018.

The project was carried out in stages as described in the following sections.

1. Data pre-processing and patch generation

The dataset used in this research work consisted of a total of 172 whole slide images in Nano-zoomer Digital Pathology (NDPI) format. These WSIs were obtained from 86 cases of patients with invasive breast carcinomas. The WSI dataset contained images of both Hematoxylin and Eosin (H&E) stained and IHC stained slides. The training data also included the evaluation of the tissue samples by clinical experts, which formed the ground truth for classification algorithms. In the data pre-processing and patch generation phase, each image was further subdivided into 26 tiles (image patches) of size 512x512 pixels. Thus, we obtained 1345 image patches. From this set of image patches, we discarded outliers that do not actually represent their HER2 class. For example, there could be image segments where there is only a very minimal amount of staining seen while a

majority of regions within the same WSI had higher levels of staining. Some patches may contain a large background region of uniform intensity. We used a preprocessing step to select only those patches with a minimum of 80% containing regions of interest. The filtered set of image patches contained a total of 1271 images. This set was further subdivided into a training set consisting of 900 image patches, and a cross-validation set consisting of 371 image patches.

2. Feature Analysis

A highly novel aspect of our work is the use of biomarker specific features representing both intensity and texture characteristics of membrane staining for classification. These features were derived from characteristic curves that represent variations in the intensity of membrane staining, and uniform local binary patterns that represent the texture of staining patterns. The dimensionality of the feature set was further reduced using Principal Component Analysis (PCA) and Fisher Linear Discriminant Analysis (LDA) , maximizing inter-class separability and improving classification accuracy. We conducted a large number of experiments to analyse the variation of accuracy with the number of principal components to obtain the optimal set of features. As a result of this analysis, the dimension of the feature vector obtained from characteristic curves could be reduced from 20 to 12, and the ULBP feature vector from 160 to 40 without significantly affecting the classification accuracy.

3. Classification Algorithm Development

We conducted extensive analysis of Neural Network architectures by varying the number of hidden layers and using different types of activation functions as detailed in our paper[1]. We also carried out an analysis of hyper parameters, and analysed the overall accuracy of results with changes in the learning rate. We also compared the performance using different types of classification algorithms such as support vector machines and one-vs-all logistic regression. A detailed analysis of results is given in the paper [1].

4. Current Development and Future Work

We have carried out the design and implementation of a Convolutional Neural Networks (CNN) and trained the network with the training dataset. Increasing the number of layers did not provide any noticeable increase in the accuracy of results. We will be conducting further analysis using a larger dataset for training and cross validation, and also by augmenting the feature set with higher order statistical features and morphological shape features.

5. Publication

We prepared a research paper outlining the work done in this project, including a detailed description of the feature extraction and selection processes, the analysis of classification algorithms, and the performance evaluation of the results. The paper has been accepted for oral presentation at the MMSP-2018 conference (details given below). The authors have acknowledged the CMRF grant support in the paper. The paper will be presented at the conference session SS.F1: *Multimodal Machine Learning: Advances and Applications*, on Friday, 31st August, 2018.

1. Prerna Singh, Ramakrishnan Mukundan, ***A Robust HER2 Neural Network Classification Algorithm Using Biomarker-Specific Feature Descriptors***, Accepted for Oral Presentation at IEEE 20th International Workshop on Multimedia Signal Processing (MMSP-2018), Simon Fraser University, Vancouver, Canada, 29-31 August 2018. (http://www.ece.ubc.ca/~mmsp2018/technical/)

2. Photographs

View an attachment by double clicking the icon to the left of the file name. Icons are not displayed and attachments are not accessible when this PDF is viewed in a web browser; you must open it in PDF reader software.

CMRF Fig.pdf
663.1 KiB

MMSP2018_HER2.pdf
1.1 MiB

4. Feedback