Published in final edited form as:

IEEE EMBS Int Conf Biomed Health Inform. 2017 February ; 2017: . doi:10.1109/bhi.2017.7897190.

Predicting Heart Rejection Using Histopathological Whole-Slide Imaging and Deep Neural Network with Dropout

Li Tong¹, Ryan Hoffman¹, Shriprasad R. Deshpande², May D. Wang^{1,*}

¹Dept. of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, GA 30332

²School of Medicine, Emory University, Atlanta, GA, 30307, USA

Abstract

Cardiac allograft rejection is one major limitation for long-term survival for patients with heart transplants. The endomyocardial biopsy is one gold standard to screen heart rejection for patients that have heart transplantation. However, manual identification of heart rejection is expensive and time-consuming. With the development of imaging processing techniques and machine learning tools, automatic prediction of heart rejection using whole-slide images is one promising approach to improve the care of patients with heart transplants. In this paper, we first develop a histopathological whole-slide image processing pipeline to extract features automatically. Then, we construct deep neural networks with and without regularization and dropout to classify the patients into nonrejection and rejection respectively. Our results show that neural networks with regularization and dropout can significantly reduce overfitting and achieve more stable accuracies.

I. INTRODUCTION

Cardiac allograft rejection is one of the main factors limiting long-term survival for patients with heart transplants, especially in the early days of transplantation [1]. Thus, screening for rejection is an important part of the care for patients with heart transplantation [2]. The endomyocardial biopsy is a gold standard for screening of rejection. Based on the biopsy, heart rejection can be classified into acute cellular rejection (ACR) and antibody-mediated rejection (AMR) [2]. ACR and AMR are mainly caused by T-cell-mediated mechanisms and host antibody responses respectively.

However, the identification of rejection using endomyocardial biopsies requires the expertise of the pathologist after manually examining histopathological whole-slide images under the microscope, which can be expensive, time-consuming, and also error-prone. With the development of imaging processing and machine learning techniques, the value of conventional histopathology is increasingly questioned [3]. Although conventional histopathology is the gold standard for allograft monitoring, the digital image processing pipeline provides valuable and unique information regarding spatial-temporal relationships, rare events/cells, and complex structural context [3]. It can also integrate with other

*Corresponding author contact: maywang@bme.gatech.edu.

information like "-Omics" analysis of tissues, peripheral blood, and body fluids [3]. The application of digital imaging processing enables digital pathology, which will significantly advance the accuracy and efficiency of diagnosis of cardiac allograft rejection.

The recent advancement of imaging processing techniques has enabled automatic quality control of raw images and comprehensive feature extraction for whole-slide images [4], [5]. After feature extraction from raw images, most classification algorithms in machine learning including random forest, support vector machine, logistic regression, and neural networks can be applied to classify a new image based on labeled training data. In this paper, we propose an automated prediction pipeline to provide decision support for the diagnosis of heart transplant rejection using image processing for feature extraction and deep neural network for classification. The rest of this paper is structured as follows. In Section II, we describe the details of the construction of neural networks with dropout. In Section III, we present the results of binary classification for heart rejection and nonheart rejection using whole-slide images. In Section IV, we draw a conclusion for the application of neural network with dropout to heart transplant rejection prediction and discuss the limitations and future directions of this study.

II. METHODS

A. Dataset

Our dataset contains whole-slide images collected from patients undergone heart transplants at Children's Healthcare of Atlanta (CHOA). Endomyocardial biopsies were captured prospectively and annotated by an expert cardiologist as no rejection, different grades of ACR, and different grades of AMR. The description and statistics of these images are summarized in Table I.

Because of the limitation of sample size (for example, we only have two images for ACR patients), we perform binary classification in this study (No Rejection versus Rejection) instead of multi-class classification.

B. Feature Extraction

We summarize the process of the feature extraction in Figure 1. The details of feature extraction for whole-slide images has been described in previous works [4], [5]. Briefly, in the first step, we construct a region-of-interest mask. The mask is generated from the megapixel scale thumbnail. This step identifies large blank areas of a slide, folded tissue, and pathologist markings, and excludes them from downstream analysis. Then, for each gigapixel-scale full resolution slide image, we break it into 512×512 pixel tiles for more manageable processing. The tiles identified as being inside the region of interest (ROI) are shown in Figure 1C. A selection of 461 features identified as relevant to WSI processing is extracted from each tile. Then for each selected tiles, we extract features including nuclear density (nuclei/tile) (Figure 1D), grey level co-occurrence matrix (GLCM) correlation coefficient of neutrophilic regions (Figure 1E), and the correlation between eosinophilic and basophilic areas (Figure 1F). Finally, we merge the features from each tile to construct the final features for each patient. After merging features of all valid tiles using a histogram

method [6], we have 4610 features for each patient. We apply principle component analysis (PCA) to reduce the dimension for data visualization. Combined patient-level features were also analyzed using the minimum-redundancy maximum-relevance mutual information algorithm (mRMR) [7] to determine which of the features are most informative to the biological problem.

C. Deep Neural Networks

In this paper, we first construct a multi-layer neural network with M hidden layers (Figure 2):

$$h_i^k = s^k \left(\sum_{j=1}^{d^k} W_{ij}^k h_j^{k-1} + b_i^k \right), k = 1, \dots, M+1 \quad (1)$$

where $f_i = h_i^{M+1}$ are output units used for prediction, s^k is a non-linear activation function (such as sigmoid and ReLU) of the k th layer, and $x_j = h_j^0$ are input values. Sigmoid function is one of the most popular activation functions for neural network models, the value of which usually stands for binary probability:

$$s(x) = \frac{1}{1 + e^{-x}} \quad (2)$$

The implemented multi-layer network is trained by minimizing supervised loss function:

$$\sum_{l=1}^L L(y_l, f_l(x)) \quad (3)$$

where L is the number of labels, y_l is the 1-of- L code of the label, f_l is the network output for l th label, x is input vector. Because we use the sigmoid function for the output unit, we decide to use the cross entropy as our loss function:

$$L(y_l, f_l(x)) = -y_l \log f_l - (1 - y_l) \log(1 - f_l) \quad (4)$$

So the loss function to be minimized is the summation of $L(y_l, f_l)$ over all labels:

$$L(\Theta) = \sum_{l=1}^L L(y_l, f_l) \quad (5)$$

D. Regularization

Large parameter values often allow models to match training data well. However, they also reduce the generalizability of the model. For improving the generalizability of the neural network, we add L_1 or L_2 regularization to the cost function to penalize large parameter values. L_1 regularization adds the linear weights to the cost function:

$$L_{reg1}(\Theta) = L(\Theta) + \lambda \sum_{k=1}^{M+1} \sum_{i=1}^{d_k+1} \sum_{j=1}^{d_k} |\Theta_{ij}^{(k)}| \quad (6)$$

On the other hand, L_2 regularization adds the squared weights to the cost function:

$$L_{reg2}(\Theta) = L(\Theta) + \lambda \sum_{k=1}^{M+1} \sum_{i=1}^{d_k+1} \sum_{j=1}^{d_k} (\Theta_{ij}^{(k)})^2 \quad (7)$$

L_1 and L_2 regularization have been extensively compared in a previous study [8]. L_1 regularization tends to produce a sparse model, while L_2 gives no preference to zero weights. Since we do not intend to train a sparse model in this study, we choose to apply L_2 regularization for our neural network.

E. Dropout

Dropout is a technique to prevent overfitting in deep neural networks, and it applies to both supervised learning [9] and unsupervised learning [10]. When using dropout, we train a different sub-model by omitting a half of hidden units in each weights update. The training procedure for dropout is similar to bagging [11], which is an ensemble method usually applied to decision trees. In the previous study [9], when applying dropout to the deep neural network, the authors randomly omit hidden units with a probability of 0.5 and input units with a probability of 0.2.

$$h_i^k = \text{drop} \left(s^k \left(\sum_{j=1}^{d^k} W_{ij}^k h_j^{k-1} + b_i^k \right) \right), k = 1, \dots, M \quad (8)$$

F. Implementation

In this study, we construct a deep neural network with one input layer, three hidden layers and one output layer with and without dropout. The input layer has 4610 units for all features we extracted from whole-slide images. Each hidden layer is consist of 256 perceptrons with sigmoid activation function as we described in previous sections. The output layer has two units for binary classification. When adding dropout and regularization to the neural network, we apply a random dropout rate of 0.8 for input layer and a random dropout rate of 0.5 for hidden layers. The L_2 regularization is also applied to the cost with a coefficient $\lambda = 10^{-5}$ to balance the training accuracy and penalty for a large number of parameters. Both neural networks are trained with 500 epochs using optimizer implemented with Adam algorithm [12]. To compare the performance of neural networks with and without dropout, we perform a 5-fold cross-validation for the 43 images. Quality control of whole-slide images and feature extraction are performed with MATLAB. The PCA and 5-fold cross-validation are performed with Python package Scikit-learn [13]. Both neural networks with and without dropout are implemented with TensorFlow [14] in Python.

III. RESULTS

The result of PCA (Figure 5) indicates that the two classes (non-rejection versus rejection) are not linearly separable, which is suitable to apply neural networks for this classification task.

The algorithms mutual information quotient variant was applied with binary sample labeling. These results are shown in Table II. The top ten ranked features are related to the shapes extracted from the basophilic stain areas as well as nuclear topology, suggesting that the shape and distribution of nuclei in the tissue images dominate the prediction.

The plot of cost, training and testing accuracy for neural networks with and without regularization and dropout are shown in Figure 6a and Figure 6b respectively. For neural networks without regularization and dropout, the cost approaches zero rapidly in the first several epochs of training. Similarly, the training accuracy also rapidly approaches to 1.0. However, the accuracy for the testing dataset is much lower than the training accuracy. For neural networks with regularization and dropout, the cost has much more fluctuations when it gradually approaches zero. We also observe similar behaviors for training and testing accuracy. In contrast to the overfitting of training accuracy for neural networks without regularization and dropout, the training and testing accuracy are comparable after applying regularization and dropout. When we run the experiments for multiple times, the neural network without regularization and dropout (training accuracy: 1.00 ± 0.00 , testing accuracy: 0.562 ± 0.145) achieves larger variation regarding testing accuracy compared to the neural network with regularization and dropout (training accuracy: 0.620 ± 0.068 , testing accuracy: 0.582 ± 0.088).

IV. CONCLUSIONS AND DISCUSSIONS

Based on the observation of extensive experiments, we conclude that neural networks with regularization and dropout can significantly reduce overfitting and achieve more stable accuracy compared to neural networks without regularization and dropout.

One limitation of this study is the small sample size. With only 43 whole-slide images available, training a comprehensive model with desired training, validations, and testing sets is difficult. Also, because of the small sample size, we are unable to perform multi-class classification. As a next step, we plan to include more patients under heart transplant and collect their whole slide images to enlarge the cohort of this study.

In this study, we demonstrated the feasibility of applying deep neural network for the prediction of heart rejection. However, optimization of the network configuration is yet to be further explored to improve the accuracy and reduce overfitting. To construct a deep neural network, we need to determine the number of layers, the number of perceptrons in each layer, activation function, and the cost function for optimization. And when we further apply regularization and dropout to the neural network, we need to decide the regularizations we want to add to the cost function and the probability of dropout in each layer. With all these factors taken into consideration, it's especially difficult to find an optimal network for

different situations. One future direction of this study is to establish a guideline for determining values of large number of hyperparameters.

ACKNOWLEDGEMENT

This work was supported by the grants from National Institutes of Health (NCI Transformative R01 CA163256, and National Center for Advancing Translational Sciences UL1TR000454), Microsoft Research and Hewlett Packard. This work was also supported in part by the scholarship from China Scholarship Council (CSC) under the Grant CSC NO. 201406010343. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or China Scholarship Council.

REFERENCES

- [1]. Patel JK, Kittleson M, and Kobashigawa JA, "Cardiac allograft rejection," *the surgeon*, vol. 9, no. 3, pp. 160–167, 2011. [PubMed: 21550522]
- [2]. Stewart S, Winters GL, Fishbein MC, Tazelaar HD, Kobashigawa J, Abrams J, Andersen CB, Angelini A, Berry GJ, Burke MM et al., "Revision of the 1990 working formulation for the standardization of nomenclature in the diagnosis of heart rejection," *The Journal of heart and lung transplantation*, vol. 24, no. 11, pp. 1710–1720, 2005. [PubMed: 16297770]
- [3]. Isse K, Lesniak A, Grama K, Roysam B, Minervini MI, and Demetris AJ, "Digital transplantation pathology: combining whole slide imaging, multiplex staining and automated image analysis," *American Journal of transplantation*, vol. 12, no. 1, pp. 27–37, 2012.
- [4]. Kothari S, Phan JH, Osunkoya AO, and Wang MD, "Biological interpretation of morphological patterns in histopathological whole-slide images," in *Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine*. ACM, 2012, pp. 218–225.
- [5]. Kothari S, Phan JH, Stokes TH, and Wang MD, "Pathology imaging informatics for quantitative analysis of whole-slide images," *Journal of the American Medical Informatics Association*, vol. 20, no. 6, pp. 1099–1108, 2013. [PubMed: 23959844]
- [6]. Kothari S, Phan JH, Young AN, and Wang MD, "Histological image classification using biologically interpretable shape-based features," *BMC medical imaging*, vol. 13, no. 1, p. 1, 2013. [PubMed: 23289431]
- [7]. Peng H, Long F, and Ding C, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005. [PubMed: 16119262]
- [8]. Ng AY, "Feature selection, l_1 vs. l_2 regularization, and rotational invariance," in *Proceedings of the twenty-first international conference on Machine learning*. ACM, 2004, p. 78.
- [9]. Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, and Salakhutdinov RR, "Improving neural networks by preventing coadaptation of feature detectors," *arXiv preprint arXiv:1207.0580*, 2012.
- [10]. Lee D-H, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Workshop on Challenges in Representation Learning, ICML*, vol. 3, 2013, p. 2.
- [11]. Breiman L, "Bagging predictors," *Machine learning*, vol. 24, no. 2, pp. 123–140, 1996.
- [12]. Kingma D and Ba J, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [13]. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, and Duchesnay E, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [14]. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M, Ghemawat S, Goodfellow I, Harp A, Irving G, Isard M, Jia Y, Jozefowicz R, Kaiser L, Kudlur M, Levenberg J, Man D, Monga R, Moore S, Murray D, Olah C, Schuster M, Shlens J, Steiner B, Sutskever I, Talwar K, Tucker P, Vanhoucke V, Vasudevan V, Viégas F, Vinyals O, Warden P, Wattenberg M, Wicke M, Yu Y, and Zheng X, "TensorFlow: Large-scale machine

learning on heterogeneous systems,” 2015, software available from [tensorflow.org](https://www.tensorflow.org). [Online]. Available: <http://tensorflow.org/>

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 1:

Visualization of the histopathological whole-slide image processing pipeline. (A) A region of the raw WSI, showing a sample of H&E stained cardiac tissue. (B) Region-of-interest mask, generated from the megapixel-scale thumbnail. (C) The gigapixel-scale, full-resolution slide image is broken into 512×512 pixel tiles for more manageable processing. (D-F) Tile-level visualization of selected features, respectively: nuclear density (nuclei/tile), GLCM correlation coefficient of neutrophilic regions, and the correlation between eosinophilic and basophilic areas.

Fig. 2:
Multilayer neural network with M hidden layers.

Fig. 3:
Diagram for a single neuron.

Fig. 4:
Diagram for a neural network with dropout. Left: a neural network with two hidden layers.
Right: the same neural network after applying dropout.

First Three PCA Directions

Fig. 5:
Visualization of the features after applying principle components analysis (PCA).

(a) Cost and accuracy for neural network without dropout

(b) Cost and accuracy for neural network with dropout

Fig. 6:
Visualization of the cost, training error, testing error for neural network with and without dropout

TABLE I:

Statistics for Heart Transplant Rejection Dataset

Category	Number of Images
No Rejection	18
Acute Cellular Rejection (ACR)	2
Antibody-Mediated Rejection (AMR)	9
AMR and ACR	14

TABLE II:

Top 10 Features Ranked by mRMR

Feature #	ID	Category
1	BS_convex_area_minimum_9	Basophilic stain object convex area
2	NTto_closeness_median_3	Closeness of nuclei
3	GgT_GrayLevel_maximum_5	Grayscale texture
4	GgT_Gabor_entrop_s_0.35_theta_1.5708.1	Grayscale Gabor entropy
5	GgT_multiwav_ghm_12_L2L1_entrop_10	Grayscale multiwavelet texture features
6	BS_e_minor_mean_1	Basophilic stain object eccentricity
7	BS_convex_area_maximum_3	Basophilic stain object convex area
8	GC_G_bin15_10	Normalized high green intensity
9	BS_boundary_fractal_std_2	Basophilic stain object boundary energy
10	PT_GrayLevel_median_1	Eosinophilic stain texture