

Automatic Lymphocyte Detection on Gastric Cancer IHC Images using Deep Learning

Emilio Garcia, Renato Hermoza, Cesar Beltran Castanon
GRPIAA - Departamento de Ingeniería
Pontificia Universidad Católica del Perú
San Miguel, Lima, Peru

Luis Cano, Miluska Castillo, Carlos Castañeda
Departamento de Investigación
Instituto Nacional de Enfermedades Neoplásicas
San Borja, Lima, Peru

Abstract—Tumor-infiltrating lymphocytes (TILs) have received considerable attention in recent years, as evidence suggests they are related to cancer prognosis. Distribution and localization of these and other types of immune cells are of special interest for pathologists, and frequently involve manual examination on Immunohistochemistry (IHC) Images. We present a model based on Deep Convolutional Neural Networks for Automatic lymphocyte detection on IHC images of gastric cancer. The dataset created as part of this work is publicly available for future research.

Keywords—cell detection; deep learning; immunohistochemistry; gastric cancer

I. INTRODUCTION

Gastric Cancer is one of the five most common types of cancer among men and women according to the World Health Organization. It represents 7% of all cases and 9% of all deaths among cancer patients.[1]. With 12 months of disease-specific survival and 90% of all cases dying within the first five years, it is one of the most aggressive and deadliest types of cancer. Therefore it is of interest for medical professionals to accurately estimate patients' prognoses.

Current evidence suggests that TILs may have an association with the prognoses and clinical features of cancer patients. In some scenarios, high concentrations of Cytotoxic T Cells and Memory T cells (T-Mem) are associated with favorable results, whereas elevated levels of Regulatory T Cells (T-Reg) may contribute to the pathogenesis[2].

However, there are other scenarios in which this relation changes drastically. For example, in the specific case of gastric cancer of the cardia, high levels of tumor-infiltrating macrophages are favorable for carcinogenesis, whereas elevated levels of T-Reg lead to better clinical results.

Moreover, not only the type of immune cells but also their location is of interest for pathologists, as the concentration of tumor-infiltrating T-Reg is only relevant to the prognosis if found in the tumor stroma[3].

*Supplementary material is available at <https://github.com/grpiaa-pucp/prisma-public/tree/master>

Over the last years, increasing computer power has been enabling considerable improvements in different areas of machine learning. Multilayered Neural Networks are making impressive progress in computer vision tasks[3]. There are even cases where convolutional neural networks have achieved better performance than humans[4, 8].

In this paper, we present an approach to automatically detect and count TILs on IHC images using Deep Convolutional Neural Networks. We also describe an innovative approach to collect images for the training set using a piece of software built in-house. The main contribution of this work is its expert-validated results. Another significant contribution is the lymphocyte images dataset which can be downloaded and used for further research and improvements in the area. The proposed software has a great potential to be used as a tool for helping medical professionals and researchers in further cancer studies.

We organized the rest of this paper as follows: we introduce Deep Learning and Convolutional Neural Networks along with related work in Section 2, the Methodology applied is described in Section 3, and Experiments are presented in Section 4. Finally, we provide conclusions in Section 5.

II. PREVIOUS AND RELATED WORK

Deep Learning appeared as a new area in Machine Learning in 2006. In the last few years, it has changed the way of working in signal processing considerably[5]. Deep Learning techniques allow machines to identify patterns and recognize images and voices. Its goal is for machines to be able to learn without needing prior preparation of training data[6].

Advances in signal processing research, big data, and the drastic increment of computing power in CPU and GPU have made popular the use of Deep Learning. Consequently, there are also substantial improvements in the applications of Deep Learning for computer vision and object recognition. Examples of successful applications include supervised and non-supervised feature extraction and classification tasks[7].

In 2011, *Cirean et al.* achieved better performance than humans for the first time with an object recognition task[8].

Later, in February 2012, he reported a new error rate of 0.23% in the MNIST handwritten digit recognition problem[7]. The same year *Krizhevsky et al.* won the ImageNet competition by combining convolutional neural networks and max-pooling with graphic processing units[9].

Later in 2015, *Delahunt et al.* built a device named Autoscope that enables users to diagnose malaria using a digital microscope and a combination of algorithms for computer vision and image classification[10].

In biological and medical fields, *Pan et al.* presented an average accuracy of 78.6% for detection of lung cancer cells, whereas *Chen and Chefd'hotel* reported to achieve a coefficient correlation between manual and automated counting of up to 99.49% using image decomposition and convolutional neural networks[11, 12].

Recently, in March 2016, *Abdel-Zaher and Eldeib* achieved 99.68% of accuracy in breast cancer classification. These results were very promising compared to previous work[13]. Similarly, other works that have used Deep Learning techniques outperformed previous studies in the areas of prediction and detection in medicine[14, 15, 16].

III. METHODOLOGY

All IHC images in this study were acquired using an Olympus BX63 Microscope and CD3 stains in cancerous gastric tissue.

For training, pathologists extracted and annotated 70x70 pixel patches containing individual cells from IHC images. Then, we trained a convolutional deep neural network to make binary classifications for inputs of the same size.

After training the network, we evaluated individual patches from test images using a sliding window algorithm. Next, we calculated the final output by applying non-maximum suppression to all single classifications in the previous step. We present our framework in Figure 1.

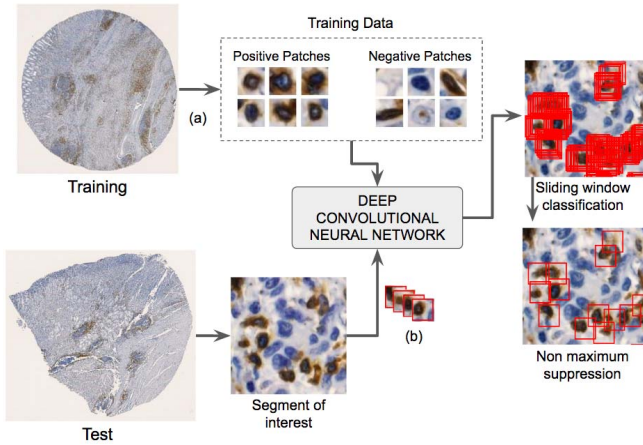


Figure 1. Overview of the proposed model.

- In contrast to other studies, we did not ask experts to mark the center of individual lymphocytes in full-sized IHC images. Instead, we provided a web application that allows for image visualization with a frame-like cropping tool and asked them to crop the cells by themselves. The process of cropping individual cell patches was as simple as identifying a cell and double-clicking on it.
- We extracted 70x70 pixel patches from test images using a sliding window algorithm. Then passed all extracted patches to the model and collected all individual classification results before applying non-maximum suppression.

A. Deep Convolutional Neural Network

Our Deep Convolutional Neural Network (DCNN) has nine layers, excluding input, which are arranged in three convolutional layers (C1, C2, and C3), three max-pooling layers (MP1, MP2, and MP3), and two fully connected layers (FC1 and FC2) with dropouts. The output layer is a fully-connected layer using the softmax function as the activation function; it outputs two classes (lymphocyte or non-lymphocyte). We show the complete network architecture in Figure 2.

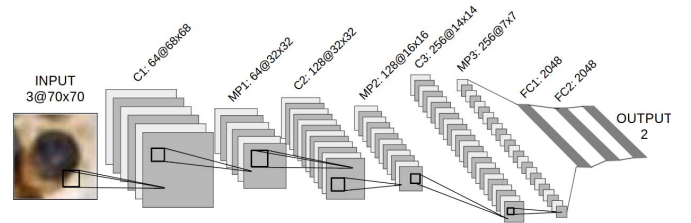


Figure 2. Deep Convolutional Neural Network Architecture.

The convolutional layers C1, C2, and C3 use 3x3 kernels with strides = [1, 1, 1, 1] and their outputs are passed to the max-pooling layers MP1, MP2, and MP3 which have kernel size and strides = [1, 2, 2, 1]. The number of features in each pair of convolutional and max-pooling layers are 64, 128, and 256 respectively. We later reshape the output of the 3rd max-pooling layer to a vector. Finally, in both fully connected layers, FC1 and FC2, we have 2048 feature vectors.

We used rectified linear units (ReLU), dropouts and batch normalization as it is known that they are useful for overfitting reduction and performance improvement with unseen data[17, 18]. For optimization, we used the ADAM algorithm, which was originally proposed by *Kingma and Ba* in [19].

IV. EXPERIMENTS

In this section, we describe the dataset and experiments setting. We also validate the proposed algorithm and compare it to medical professional output.

A. Data Set and Experiment Setting

A clinical dataset containing gastric cancer tissue samples was used to test the proposed model. We extracted the data from a set of 10 full-sized micrographs of cancerous gastric tissue scanned at 40x magnification using a software utility developed as part of this work. Figure 3 shows the software utility. Another 35 additional 600x500 pixel images were used to compare against the results provided by human pathologists.

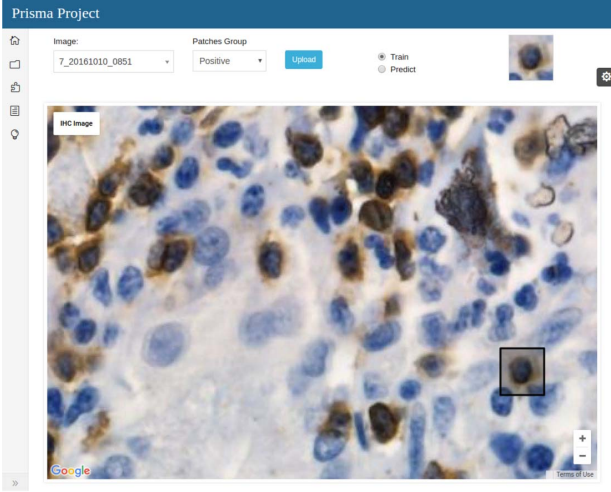


Figure 3. Patches Extraction for Training and Test

After extracting individual cell patches, the database consisted of 3,257 images (70 x 70 pixel each). Then, we used data augmentation techniques such as rotation and reflection to increase the dataset size and got a total of 10868 individual cell patches. The distribution of the dataset for training and testing is shown in Table I.

Table I
DESCRIPTION OF THE DATA SET

Set	Quantity
Before data augmentation	3257
After data augmentation	10868
Positive Train Samples	4437
Negative Train Samples	4257
Positive Test Samples	1143
Negative Test Samples	1031

B. Performance on single-cell images

For training, our best model achieved 96.88% accuracy after 1,193 iterations. As the training was done using GPU-enabled hardware, each configuration took about 1 hour to train. We tested more than 200 different configurations. Figure 4 shows the Accuracy (a) and Loss (b) for our best model.

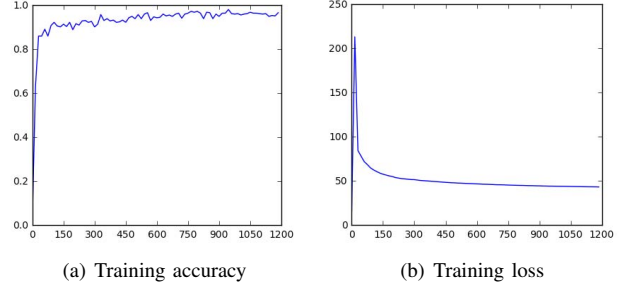


Figure 4. Model Accuracy and Loss

C. Comparison against human experts

We used 35 images, 600 x 500 pixel each, to compare the model results against human pathologists. These validation images were randomly selected from new full-sized IHC micrographs. Samples of classification outputs are shown in Figure 5.

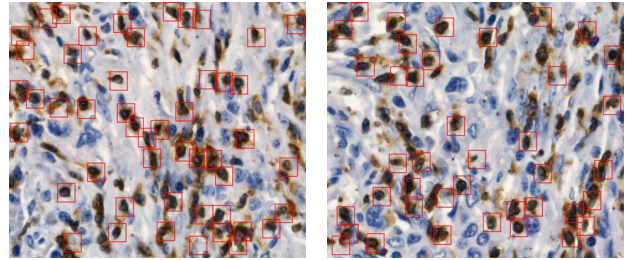


Figure 5. Classification on 600 x 500 pixel images

When analyzing results using data augmentation, 17 images out of 35 had an output difference in terms of detected cells equal to or lower than 5, whereas a total of 29 had a count difference equal to or lower than 11. Therefore, only six images had a count difference higher than 11.

If we compare our best model using data augmentation (Model A) against our best model without data augmentation (Model B), better results are seen in Model A.

However, both models present the same behavior: as the difference between numbers of human and algorithm outputs increase, the number of images exhibiting those differences decrease. Figure 6 explains this relationship in a more visual fashion.

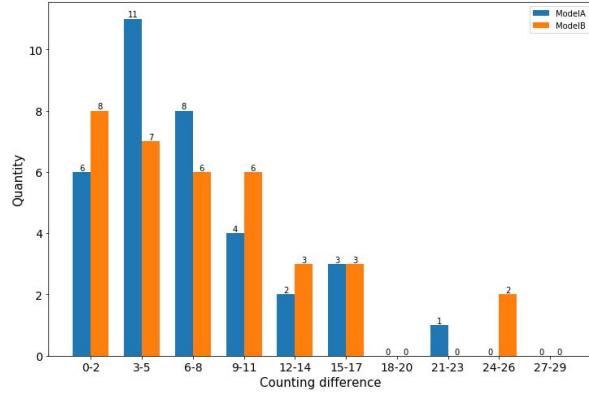


Figure 6. Behaviour of counting differences

V. CONCLUSION

We present an approach for automatic TILs detection and counting on IHC images of gastric cancer using Deep Convolutional Neural Networks. Our experiments produced an acceptable 96.88% accuracy, but the most valuable contribution of our work is the exhaustive validation with the pathologist. A similar work was found in the literature [12], but we enhanced the segmentation and labeling process. While *Chen and Chefhotel* applied a semi-automatic segmentation of cells, we implemented a fully automated segmentation process. We also developed an innovative approach to collect images for the training set using a piece of software built in-house. Another significant difference is the number of instances for the training set, while *Chen and Chefhotel* extracted manually 491 positives and 539 negative samples, we extracted 1395 positive and 1322 negative samples, this new dataset is a significant contribution to the scientific community and is available as supplementary material. Other works and methods can also use the new dataset created as part of this work to continue improving the accuracy in lymphocyte detection. Finally, the provided software, which made the patches extraction easier, can also be used to create datasets for other types of cells or images. While there is room for improvement, the proposed method has potential to be used as a resource for helping medical professionals and researchers in further cancer studies.

ACKNOWLEDGMENT

The authors would like to thank Joselyn Sánchez and Luis Bernabé for their helpful contribution while creating the lymphocytes dataset. We are also thankful to Corrado Daly for his contribution getting our computing infrastructure ready, and to all of our friends and colleagues for their valuable comments and discussions that helped to better shape this work.

REFERENCES

- [1] S. W and W. P, *World Cancer Report 2014 (PDF)*, ser. IARC Nonserial Publication. International Agency for Research on Cancer (I A R C) (UN), 2014.
- [2] A. Amedei, C. D. Bella, E. Silvestri, D. Prisco, and M. M. D’Elios, “T Cells in Gastric Cancer: Friends or Foes.” *Clinical & Developmental Immunology*, pp. 1–10, Jan. 2012.
- [3] C. Edwards, “Growing Pains for Deep Learning.” *Communications of the ACM*, vol. 58, no. 7, pp. 14–16, Jul. 2015.
- [4] K. He, X. Zhang, S. Ren, and J. Sun, “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification,” in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.
- [5] L. Deng, D. Yu *et al.*, “Deep learning: methods and applications,” *Foundations and Trends® in Signal Processing*, vol. 7, no. 3–4, pp. 197–387, 2014.
- [6] “How Deep Learning Will Change Our World.” *Trends Magazine*, no. 156, pp. 27–30, 2016.
- [7] D. C. Cirean, U. Meier, and J. Schmidhuber, “Multi-column Deep Neural Networks for Image Classification,” *CoRR*, vol. abs/1202.2745, 2012.
- [8] D. Cirean, U. Meier, J. Masci, and J. Schmidhuber, “Multi-column deep neural network for traffic sign classification,” *Neural Networks*, vol. 32, pp. 333 – 338, 2012, selected Papers from {IJCNN} 2011.
- [9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105.
- [10] C. B. Delahunt, C. Mehanian, L. Hu, S. K. McGuire, C. R. Champlin, M. P. Horning, B. K. Wilson, and C. M. Thompson, “Automated microscopy and machine learning for expert-level malaria field diagnosis,” in *Global Humanitarian Technology Conference (GHTC), 2015 IEEE*. IEEE, 2015, pp. 393–399.
- [11] H. Pan, Z. Xu, and J. Huang, “An Effective Approach for Robust Lung Cancer Cell Detection.” *Patch-Based Techniques in Medical Imaging*, p. 87, Jan. 2015.
- [12] T. Chen and C. Chefhotel, “Deep Learning Based Automatic Immune Cell Detection for Immunohistochemistry Images.” *Machine Learning in Medical Imaging: 5th International Workshop, MLMI 2014, Held in Conjunction with MICCAI 2014, Boston, MA, USA, September 14, 2014. Proceedings*, p. 17, Jan. 2014.
- [13] A. M. Abdel-Zaher and A. M. Eldeib, “Breast cancer classification using deep belief networks,” *Expert Systems With Applications*, vol. 46, pp. 139–144, Mar.

- 2016.
- [14] J.-Z. Cheng, D. Ni, Y.-H. Chou, J. Qin, C.-M. Tiu, Y.-C. Chang, C.-S. Huang, D. Shen, and C.-M. Chen, "Computer-aided diagnosis with deep learning architecture: Applications to breast lesions in us images and pulmonary nodules in ct scans," *Scientific reports*, vol. 6, p. 24454, 2016.
 - [15] Guo-Ping Liu, Jian-Jun Yan, Yi-Qin Wang, Wu Zheng, Tao Zhong, Xiong Lu, and Peng Qian, "Deep Learning Based Syndrome Diagnosis of Chronic Gastritis." *Computational & Mathematical Methods in Medicine*, pp. 1–8, Jan. 2014.
 - [16] P.-P. Ypsilantis, M. Siddique, H.-M. Sohn, A. Davies, G. Cook, V. Goh, and G. Montana, "Predicting Response to Neoadjuvant Chemotherapy with PET Imaging Using Convolutional Neural Networks." *PLoS ONE*, vol. 10, no. 9, pp. 1–18, Sep. 2015.
 - [17] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting." *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.
 - [18] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *CoRR*, vol. abs/1502.03167, 2015.
 - [19] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014.