Evaluations of Deep Convolutional Neural Networks for Automatic Identification of Malaria Infected Cells

Yuhang Dong[1], Zhuocheng Jiang[1], Hongda Shen[1], W. David Pan[1]
Lance A. Williams[2], Vishnu V. B. Reddy[2], William H. Benjamin, Jr.[2], Allen W. Bryan, Jr.[2]

***Abstract*— This paper studied automatic identification of malaria infected cells using deep learning methods. We used whole slide images of thin blood stains to compile an dataset of malaria-infected red blood cells and non-infected cells, as labeled by a group of four pathologists. We evaluated three types of well-known convolutional neural networks, including the LeNet, AlexNet and GoogLeNet. Simulation results showed that all these deep convolution neural networks achieved classification accuracies of over 95%, higher than the accuracy of about 92% attainable by using the support vector machine method. Moreover, the deep learning methods have the advantage of being able to automatically learn the features from the input data, thereby requiring minimal inputs from human experts for automated malaria diagnosis.**

I. INTRODUCTION

Malaria is a life-threatening disease caused by parasites that are transmitted to people through the bites of infected female *Anopheles* mosquitoes. According to the report released by World Health Organization (WHO), there were 214 million cased of malaria in 2015 and 438,000 deaths [1]. In most cases, malaria can only be diagnosed by manual examination of the microscopic slide. Whole slide imaging, which scans the conventional glass slides in order to produce digital slides, is the most recent imaging modality being employed by pathology departments worldwide. Fig. 1 shows four human red blood cell samples obtained using whole slide images. In order to provide a reliable diagnosis, necessary training and specialized human resource are required. Unfortunately, these resources are far from being adequate and frequently often unavailable in underdeveloped areas where malaria has a marked predominance. Therefore, an automated diagnostic system can provide an efficient solution to this problem.

Fig. 1: Red blood cell samples: the two cells on the left are malaria infected, and two cells on the right are non-infected.

Recently, machine learning algorithms have gained an increasing attention from researchers for its great capability in building automated diagnostic system for malaria [2], [3]. In [2], SVM and Naive Bayes Classifier were utilized to achieve accuracies of 84% and 83.5% respectively. In contrast to the supervised learning, unsupervised learning, for instance, K-Nearest Neighbor (KNN) has also been proposed to recognize malaria infected cells in [3]. In [4], a three-layer Neural Network (NN) was designed as a classifier to detect malaria infected cells at an accuracy of 85%. Although these supervised learning algorithms had some success with good infection detection accuracies, their performances are very sensitive to the feature extraction used in their work. Thus building a discriminant feature vector with minimal redundancy is very crucial. Plenty of work has been done to extract features for the malaria infected cells [2], [3], [5]. In [6], a survey of feature extraction and optimization for malaria cells has been discussed in detail. Although good feature extraction method can improve the detection accuracy, this type of infection detection cannot achieve fully automated diagnosis because it still requires trained experts to manually extract feature vectors according to the specific datasets. In this work, we propose to employ deep learning algorithms for malaria cell detection, with the ultimate goal of building a truly automated diagnostic platform without any manual feature extraction. To this end, deep machine learning methods can provide a good solution. Deep learning methods can extract from the input a hierarchical representation of the data, with higher layers representing increasingly abstract concepts, which are increasingly invariant to transformations and scales. For example, in [7], a deep convolutional neural network (CNN) was applied to diagnose malaria in thick blood smear. However, to distinguish infected and non-infected samples in thick films is essentially difficult for pathologists, as the difference is not as clear as those individual red blood cells cropped from whole slide images based on thin films.

In this work, three well-known deep convolutional neural networks, including LeNet-5 [8], [9], AlexNet [10] and GoogLeNet [11] were used to learn the inherent features of the malaria infected cells and the non-infected cells. For comparison, a support vector machine (SVM) was trained on pre-selected features extracted from the same dataset.

II. CONVOLUTIONAL NEURAL NETWORKS

Convolutional neural network is an artificial neural network inspired by the animal visual system [12]. Convolutional layer, pooling layer and fully connection layer are the three main types of layers used to construct the CNN

[1]Y. Dong, H. Shen, Z. Jiang and W. D. Pan are with the Dept. of Electrical and Computer Engineering, University of Alabama in Huntsville, Huntsville, AL 35899, USA. Email: pand@uah.edu.

[2]L. A. Williams, V. V. B. Reddy, W. H. Benjamin, Jr. and A. W. Bryan, Jr. are with the Dept. of Pathology, University of Alabama at Birmingham, Birmingham, AL 35233, USA.

978-1-5090-4179-4/17/$31.00 ©2017 IEEE

architecture. Compared to traditional neural networks, CNNs can extract features without losing much spatial correlations of the input. Each layer consists of neurons that have learnable weights and biases. The optimal model is achieved after feeding data into the network and minimizing the loss function at the top layer. Several different architectures of CNN have been proposed. In this work, three well-known CNN models, LeNet-5, AlexNet and GoogLeNet, were evaluated on the same dataset and compared to SVM. Comparison of details of each CNN is given in Table I.

CNN	LeNet-5	AlexNet	GoogLeNet
Year Proposed	1998	2012	2014
# of Layers	4	8	22
Top 5 errors on ILSVRC	?	16.4%	6.7%
# of Convolutional Layers	3	5	21
Kernal Size	5	11,5,3	7,1,3,5
# of Fully Connected Layers	1	3	1
# of Parameters	3628072	20176258	5975602
Dropout	No	Yes	Yes
Data Augmentation	No	Yes	Yes
Inception	No	No	Yes
Local Response Normalization	No	Yes	Yes

TABLE I: Comparison of LeNet-5, AlexNet and GoogLeNet. ILSVRC stands for "ImageNet Large Scale Visual Recognition Competition".

LeNet-5 [8] was first used in handwritten digit recognition and achieved an impressive error rate as low as 0.8%. AlexNet [10] won ImageNet ILSVRC in 2012 by decreasing the top-5 errors to less than 20% for the first time. It popularized the use of convolutional neural networks in computer vision tasks. With the help of the so-called "Inception Module", GoogLeNet [11] was able to go deeper (to 22 layers), albeit with many fewer parameters than AlexNet. GoogLeNet won the ILSVRC in 2014. Since LeNet-5 was proposed more than 10 years ahead of AlexNet and GoogLeNet, its result on ILSVRC dataset was not available in Table I. Its simple architecture (without the features of dropout, data augmentation, inception, as well as local response normalization) makes its number of parameters far less than the other two methods. Among the three networks, AlexNet has the most parameters, while the GoogLeNet is the "deepest".

III. Cell Image Dataset of Malaria Infection

A. Image Source

Whole slide images refer to the digital image of high magnification created by scanning an entire microscopic slide. Whole slide images scanned by a $40\times$ objective give rise to substantially large file size, for instance, approximately 2 Gigabytes. Therefore, whole slide images are mostly represented using a tiled pyramid file format such as DeepZoom pyramid [13]. The whole slide images used in this work were similar to those found in the PEIR-VM repository built by the University of Alabama at Birmingham (UAB) [14]. The images contained in the repository were originally acquired in the form of TIFF images that were uncompressed. In order to reduce the arising demand on storage space, the images were converted to JPEGs without significant loss of quality. The image libraries were re-organized using the MediaWiki publication platform. The images in the repository were subject to Wright-Geimsa staining which would highlight components such as white blood cells, platelets and infected cells. The images were scanned using the Aperio™ Whole Slide Image Scanner and converted to digital format for processing through a computer using the scanner's built-in digital camera [15].

B. Data Preprocessing

The original whole slide image data contain significant amount of redundant information. In order to achieve good classification accuracy, image segmentation and de-noising have to be done to extract only blood cells and remove those redundant image pixels simultaneously. First, each image tile was converted into a grayscale image from the color space followed by the thresholding operation. However, noise as the byproduct of thresholding can degrade the quality of the acquired images and lower the classification performance. Therefore, the isolated noise pixels were eliminated by applying image morphological operations in Matlab.

Another common issue is that segmented whole slide images almost inevitably have many overlapped red blood cells (RBCs), which may cause inaccurate classification. The Hough Circle transform [16] was applied to detect disk-like overlapped RBCs and then separate them.

C. Image Dataset Compilation

After the data preprocessing, we randomly selected a large number of cell images and provided them to pathologists at the University of Alabama at Birmingham. The entire whole slide image dataset have been divided into four segments evenly. Each of four pathologists are assigned with two segments so that each cell image will be viewed and labeled by at least two experienced pathologists. One cell image can only be considered as infected and included in our final dataset if all the reviewers mark it positively whereas it will be excluded otherwise. The same selection rule also applies to the non-infected cells in our dataset. After this data curation, we collected 1,034 infected cells and 1,531 non-infected cells. Next we divide this dataset into two sets of approximately equal size: training set and testing set. Table II shows the number of infected and non-infected cells for training set and testing set.

Label	Training set	Testing set
Infected	517	517
Non-infected	765	766

TABLE II: Distribution of the two types of cells in the training and testing sets [17].

IV. Simulation Results and Analysis

A. Simulation Platform

The Nvidia DIGITS™ DevBox provides an efficient platform of using CNN for classification. LeNet-5 has been integrated into and optimized for the DevBox. The DIGITS

software accelerated by up to four Nvidia Titan X GPUs were used for the neural network training.

In DIGITS, the training epoch number was set to 30 for this work. In other words, all the training samples will run the forward pass and backward pass through the network for 30 times. The Stochastic Gradient Descent algorithm [18] was used as the solver to optimize the Mean Square Error (MSE) cost function.

B. Support Vector Machine Method

The SVM used in this work involved feature ranking and selection [6]. Among 76 features, seven best features were selected using the Kullback-Leibler (KL) distance. They are Hu's moments (7, 5, 3, 6), MinIntensity, Shannon's entropy and Hu's moment 2 (ranked from high to low) [19].

C. Simulation Results

All three neural networks were trained using this training set and the trained models were then tested on the testing set. 25% of images in the entire training set were randomly collected to build a validation set. After training, validation was performed, with the loss of each algorithm being shown in Fig. 2. We can readily see that the loss converged at the 9^{th} epoch for LeNet-5, and then went back up. This is indicative of overfitting. GoogLeNet seemed to have a similar problem – the loss reached a local minimum at the 11^{th} epoch and then slowly increased. Overfitting might be due to the dataset being too small for training and validation. Dropout [20] can be a very effective way to prevent overfitting. As shown in Table I, among the three CNNs, LeNet does not have dropout, thereby yielding the worst performance.

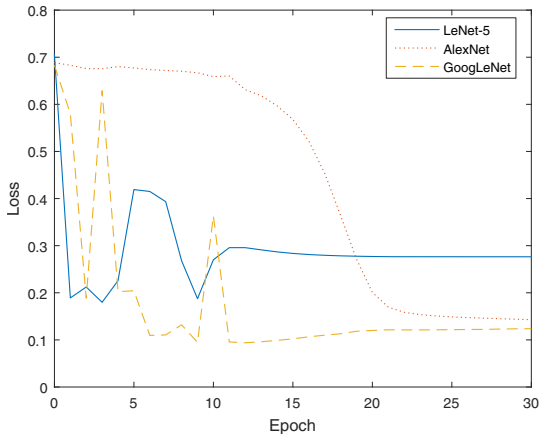

Fig. 2: Validation loss of the three convolution neural networks.

In addition to the validation, the trained neural networks have been tested on the predefined testing set to demonstrate their capability to handle unknown data. The confusion matrix is given in Table III.

While all four methods achieved classification accuracies above 90%, the SVM method has a lower accuracy than the deep convolutional neural networks, as the SVM method misclassified 90 out of 776 cells. The GoogLeNet outperformed the other two CNNs, probably owning to its depth

		Ground Truth: Positive	Ground Truth: Negative	Accuracy
LeNet-5	Positive	493	25	96.18%
	Negative	24	741	
AlexNet	Positive	502	39	95.79%
	Negative	15	727	
GoogLeNet	Positive	503	10	98.13%
	Negative	14	756	
SVM	Positive	500	90	91.66%
	Negative	17	676	

TABLE III: The confusion matrix of the testing results.

that may have extracted higher order of information. The CNNs also have a clear advantage over the SVM method, since neither manual feature extraction nor feature selection is needed using domain knowledge.

Since the SVM algorithm was run on MATLAB, with feature generation, selection and ranking requiring very long computational time, we compared only the three CNN models regarding the running time of training-validation and testing (see Table IV). Note that all input images for LeNet and AlexNet have the size of 60 $\times$ 60, while GoogLeNet only accept 256 $\times$ 256, so up-sampling was applied first.

CNN	LeNet	AlexNet	GoogLeNet
Training-Validation	7	28	141
Testing	5	5	19

TABLE IV: Comparison of the running times (in seconds).

The rank of training-validation time almost follows the statistics in Table I – more parameters and larger size of network require longer time to train. In terms of testing time, AlexNet had a similar performance as LeNet. Considering the input image size of GoogLeNet is more than 16 times larger than that of the other two CNNs, the running time could also be competitive, not to mention the highest classification accuracy it achieved as shown in Table III.

D. Case Study

To investigate the features automatically extracted by the deep learning methods from the training data, we randomly selected one image from each of the two categories of infected and normal cells (as shown in the first row of Fig. 3). We fed these images into LeNet-5, which has the simplest architecture among the three CNNs. Visualization of the features at the two convolutional layers of LeNet-5 in Fig. 3 demonstrates how the extracted features respond to inputs from the two distinct categories. Comparison of the learned features shows the excellent automatic feature extraction capability of this deep learning method. For example, the existence of a colored nucleus in the infected cell can be caused by the malaria parasite, which is a distinct feature of an infected cell compared to a non-infected cell.

In addition, histograms can be generated from feature maps for each layer, providing statistical summaries of all the features learned as shown in Fig. 4. This case study shed some light on why automated malaria diagnosis would benefit from the deep learning techniques.

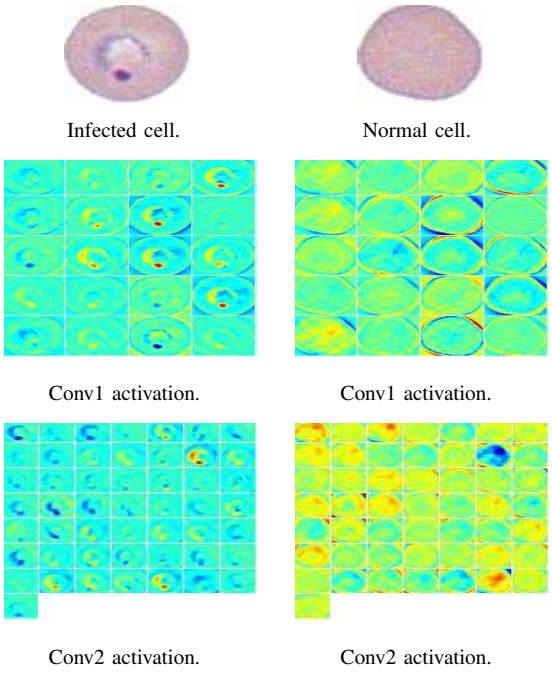

Fig. 3: Malaria infected cell visualization of example: two convolutional layers.

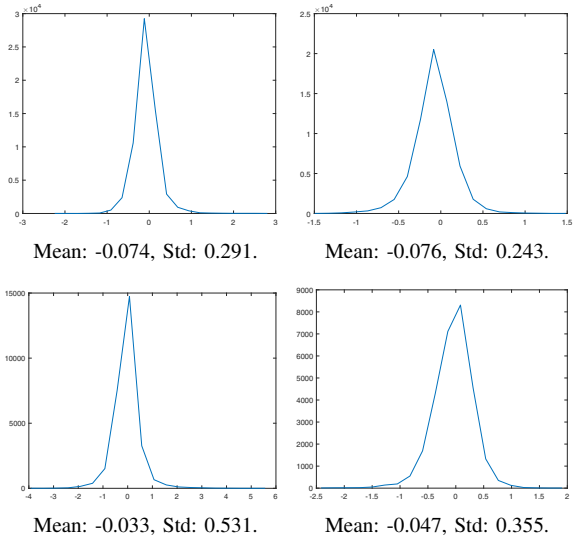

Fig. 4: Histogram of the features at two convolutional layers (left for the infected cell, right for the non-infected cell).

V. CONCLUSIONS

In view of lack of publicly available, high-resolution image datasets to support patter recognition research for automated malaria diagnosis, we built an image dataset of malaria infected human red blood cells extracted from high-resolution whole slide images. The cell images in the datasets were labeled by a group of pathologists from the Medical School of the University of Alabama at Birmingham. We then used the dataset to train and test several well-known deep convolutional networks. Simulation results showed that very high recognition accuracy could be achieved by these deep learning techniques. They not only achieved higher accuracy than the support vector machine method using hand-crafted features, but also have the advantage of being able to automatically extract multiple layers of features from the input data. We plan to expand the dataset by including more pathologist-curated cell images. Our ultimate goal is to build a reliable and accurate automated detection system for malaria diagnosis.

REFERENCES

[1] World Health Organization, "Disease burden of malaria," http://www.who.int/mediacentre/factsheets/fs094/en/.

[2] D. K. Das, et al. , "Machine learning approach for automated screening of malaria parasite using light microscopic images," *Journal of Micron*, vol. 45, pp. 97–106, Feb 2013.

[3] F. B. Tek, A. G. Dempster and I. Kale, "Parasite detection and identification for automated thin blood film malaria diagnosis," *Journal of Computer Vision and Image Understanding*, vol. 114, no. 1, pp. 21–32, January 2010.

[4] N. E. Ross, et al., "Automated image processing method for the diagnosis and classification of malaria on thin blood smears," *Medical and Biological Engineering and Computing*, vol. 44, no. 5, pp. 427–436, May 2005.

[5] F. B. Tek, "Computerised diagnosis of malaria," Ph.D thesis, University of Westminster, 2007.

[6] V. Muralidharan, Y. Dong, and W. D. Pan, "A comparison of feature selection methods for machine learning based automatic malarial cell recognition in wholeslide images," in *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, Feb 2016, pp. 216–219.

[7] J. A. Quinn, R. Nakasi, P. K. Mugagga, P. Byanyima, W. Lubega, and A. Andama, "Deep convolutional neural networks for microscopy-based point of care diagnostics," *arXiv preprint arXiv:1608.02989*, 2016.

[8] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[9] Y. LeCun, K. Kavukvuoglu, and C. Farabet, "Convolutional networks and applications in vision," in *Proc. International Symposium on Circuits and Systems (ISCAS'10)*. IEEE, 2010.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[11] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.

[12] D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture of monkey striate cortex," *The Journal of physiology*, vol. 195, no. 1, pp. 215–243, 1968.

[13] "Visualizing the images and annotations," http://grand-challenge.org/site/camelyon16/data/.

[14] "Whole slide image for malaria infected red blood cells," http://peir-vm.path.uab.edu/debug.php?slide=IPLab11Malaria.

[15] University of Alabama at Birmingham, "PEIR-VM," available: http://peir-vm.path.uab.edu/about.php.

[16] R. O. Duda and P. E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Communications of the ACM*, vol. 15, no. 1, pp. 11–15, 1972.

[17] Univ. of Alabama in Huntsville, "Link to the dataset used," http://www.ece.uah.edu/~dwpan/malaria_dataset/.

[18] V. P. Plagianakos and G. D. Magoulas, "Stochastic gradient descent," *Advances in Convex Analysis and Global Optimization: Honoring the Memory of C. Caratheodory (1873–1950)*, vol. 54, p. 433, 2013.

[19] M.-K. Hu, "Visual pattern recognition by moment invariants," *information Theory, IRE Transactions on*, vol. 8, no. 2, pp. 179–187, 1962.

[20] 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.