Automatic white blood cell classification using pre-trained deep learning models: ResNet and Inception

Conference Paper · April 2018

DOI: 10.1117/12.2311282

CITATIONS
159

READS
6,845

5 authors, including:

Mehdi Habibzadeh
Concordia University Montreal
12 PUBLICATIONS **429** CITATIONS

SEE PROFILE

Zahra Rezaei
University of Kashan
11 PUBLICATIONS **187** CITATIONS

SEE PROFILE

Mehdi Totonchi
Royan Institute
120 PUBLICATIONS **1,934** CITATIONS

SEE PROFILE

Automatic white blood cell classification using pre-trained deep learning models: ResNet and Inception

Mehdi Habibzadeh, Mahboobeh Jannesari, Zahra Rezaei, Hossein Baharvand, Mehdi Totonchi

Mehdi Habibzadeh, Mahboobeh Jannesari, Zahra Rezaei, Hossein Baharvand, Mehdi Totonchi, "Automatic white blood cell classification using pre-trained deep learning models: ResNet and Inception," Proc. SPIE 10696, Tenth International Conference on Machine Vision (ICMV 2017), 1069612 (13 April 2018); doi: 10.1117/12.2311282

SPIE.

Event: Tenth International Conference on Machine Vision, 2017, Vienna, Austria

Automatic White Blood Cell Classification Using Pre-trained Deep Learning Models: ResNet and Inception

Mehdi Habibzadeh[1#], Mahboobeh Jannesari[2#], Zahra Rezaei[3], Hossein Baharvand[1]*, Mehdi Totonchi[1]*

[1]Department of Stem Cells and Developmental Biology, Cell Science Research Center, Royan Institute for Stem Cell Biology and Technology, ACECR, Tehran, Iran.

[2]Departments of Bioinformatics, Institute of Biochemistry and Biophysics (IBB), University of Tehran, Tehran, Iran.

[3]Department of Computer Engineering, University of Kashan, Kashan, Iran

ABSTRACT

This works gives an account of evaluation of white blood cell differential counts via computer aided diagnosis (CAD) system and hematology rules. Leukocytes, also called white blood cells (WBCs) play main role of the immune system. Leukocyte is responsible for phagocytosis and immunity and therefore in defense against infection involving the fatal diseases incidence and mortality related issues. Admittedly, microscopic examination of blood samples is a time consuming, expensive and error-prone task. A manual diagnosis would search for specific Leukocytes and number abnormalities in the blood slides while complete blood count (CBC) examination is performed. Complications may arise from the large number of varying samples including different types of Leukocytes, related sub-types and concentration in blood, which makes the analysis prone to human error. This process can be automated by computerized techniques which are more reliable and economical. In essence, we seek to determine a fast, accurate mechanism for classification and gather information about distribution of white blood evidences which may help to diagnose the degree of any abnormalities during CBC test. In this work, we consider the problem of pre-processing and supervised classification of white blood cells into their four primary types including Neutrophils, Eosinophils, Lymphocytes, and Monocytes using a consecutive proposed deep learning framework. For first step, this research proposes three consecutive pre-processing calculations namely are color distortion; bounding box distortion (crop) and image flipping mirroring. In second phase, white blood cell recognition performed with hierarchy topological feature extraction using Inception and ResNet architectures. Finally, the results obtained from the preliminary analysis of cell classification with (11200) training samples and 1244 white blood cells evaluation data set are presented in confusion matrices and interpreted using accuracy rate, and false positive with the classification framework being validated with experiments conducted on poor quality blood images sized 320 × 240 pixels. The deferential outcomes in the challenging cell detection task, as shown in result section, indicate that there is a significant achievement in using Inception and ResNet architecture with proposed settings. Our framework detects on average 100% of the four main white blood cell types using ResNet V1 50 while also alternative promising result with 99.84% and 99.46% accuracy rate obtained with ResNet V1 152 and ResNet 101, respectively with 3000 epochs and fine-tuning all layers. Further statistical confusion matrix tests revealed that this work achieved 1, 0.9979, 0.9989 sensitivity values when area under the curve (AUC) scores above 1, 0.9992, 0.9833 on three proposed techniques. In addition, current work shows negligible and small false negative 0, 2, 1 and substantial false positive with 0, 0, 5 values in Leukocytes detection.

Keywords: Deep learning, Inception, ResNet, transfer learning, fine-tuning, white blood cell classification.

1. MEDICAL BACKGROUND AND INTRODUCTION

The examination of peripheral thin blood smears plays the main role of hematologic diagnosis. Blood cells are categorized as Red Blood Cells (Erythrocytes), White Blood Cells (Leukocytes), and platelets. The main responsibility of leukocytes is to defend of the body organs using phagocytic activity mechanism to remove cell debris and damage in biological structures. There are five normal WBCs mature types (with typical percentage of occurrence in normal blood): Basophil (≤ 1%); Eosinophil (≤ 5%); Monocyte (3−9%); Lymphocyte (25−35%); and Neutrophil (40−75%) [1]. In this context, complete blood count (CBC) analysis is first action in diagnosis all blood-related diseases such as anemia (alpha

Tenth International Conference on Machine Vision, edited by Antanas Verikas, Petia Radeva, Dmitry Nikolaev, Jianhong Zhou, Proc. of SPIE Vol. 10696, 1069612 · © 2018 SPIE
CCC code: 0277-786X/18/$18 · doi: 10.1117/12.2311282

and beta thalassemia) and infections. Due to these difficulties, an automated, and reproducible methodology for detection of WBCs could dramatically reduce required time for microscopic observation, and reduce rate of variability in diagnosis and interpretation of thin blood slides [2]. We developed a computerized method which may allow medical physicians and patients to track cell structure to count white blood cells. Several previous reported techniques require extensive pre-processing, and extraction of specific visual features before classification. Thus, in this study, we train an Inception and ResNet techniques using a data-sets consist of imbalanced 12,444 samples with 3,120 Eosinophil, 3,103 Lymphocyte, 3,098 Monocyte, and 3,123 Neutrophil where augmentation techniques including rotations, flips, and shearing are applied (see section 3.1).

2. BACKGROUND AND LITERATURE SURVEY

This section reviews the literature concerning the usefulness of conventional laboratory medical procedures, image processing and machine learning techniques in white blood cell detection.

2.1 Literature Survey in Conventional Medical Procedures

Current hematology analyzers used in most medical laboratories are such as Siemens ADVIA 2120i [3], Sysmex XE-series [4] and also Abbott CELL-DYN [5] with the manual ground truth white blood cell deferential count. Poor resolution and leukocytes adversely affect differential count precision in manual inspection. On the other hand, the erythrocytes and leukocyte types that the current equipment are able to manage are restricted to some classes where always update of these systems are based on expensive chemicals and mechanical process [3–5].

2.2 Image Processing Background

Authors illustrated a large number of series including pre-processing, morphological operations, feature extraction and conventional classifications [2, 6–11] to WBC differntial count. However, so far, there has been little attempt to use very deep learning techniques for white blood cell recognition in given blood smear slides [12–14]. Authors evaluated certain types of well-known convolutional neural networks, including the LeNet5, AlexNet and GoogLeNet. The results of these and other studies support the idea that deep learning is ongoing in medical issues [15–20]. Following earlier work, this study examined Inception and ResNet deep learning approaches to distinguish four main dominant WBC types in blood smear slides. The empirical findings in this study provide a new understanding of feature extraction and make noteworthy contributions to computerized CBC test.

3. PROPOSED COMPUTER-AIDED DIAGNOSIS FRAMEWORK

Our framework comprises four steps: 1) Image acquisition and conversion to JPEG / RGB channel. 2) Appropriate deep learning data augmentation and pre-processing steps (see section 3.1). Next, white blood cell recognition is accomplished with very deep learning approaches: 3) Transfer learning and pre-trained models (see section 3.2). 4) Hierarchical feature extraction and classification with Inception and ResNet networks (see sections 3.3).

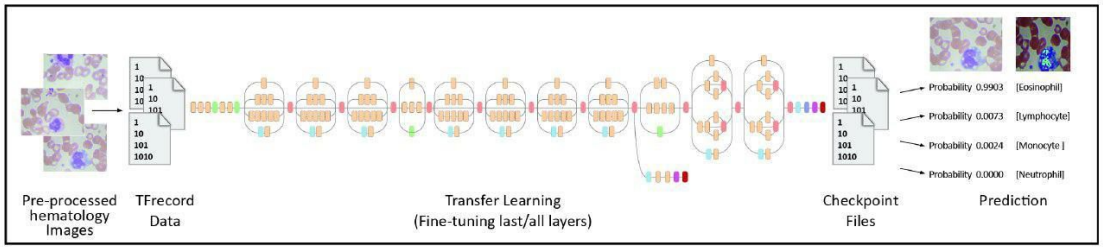

Figure 1. Framework in pipeline.

3.1 Preparing, Data Augmentation and Pre-Processing

The current study addresses date-set gathered from a given personal github [21]. Initial data-set includes of 352 images with size 320 × 240 of four main WBC types (Neutrophils, Eosinophils, Lymphocytes, and Monocytes). Data is

composed of 21 Monocyte, 33 Lymphocyte, 207 Neutrophil, 88 Eosinophil, 3 Basophil. Since the number of Basophil is very limited in current available data-set [21] and also occurrence of it is normally negligible (less than 1%), this study ignores it to generalize proposed solution. This work profits JPEG file decoder and proposes a modified image augmenter [21] that remedies drawbacks of limited data set with 352 actual samples. Overall, augmented data consists of imbalanced 12, 444 white blood cells with 3,120 Eosinophil, 3,103 Lymphocyte, 3,098 Monocyte, and 3,123 Neutrophil. Afterwards, it goes with TFrecord format conversion highly recommended at data transfer and serialization [22]. This work resizes the image to $299 \times 299 \times 3$ and $224 \times 224 \times 3$ based on model image size recommendations for Inception and ResNet architectures, respectively [23–25]. Following that, it randomly flips an input image left to right horizontally. Distorting images are also addressed with four color operations using hue, brightness, saturation and contrast adjustment. As a result, it provides an efficient mechanics for training data preparation step. On the other hand, in evaluation step, all images are normalized, cropped and resized to specific height and width based on aforementioned models input size.

3.2Transfer Learning

Data were gathered from multiple sources proved that pre-trained ConvNets along with fine-tuning policies is better or, at least, equal as well as a deep networks trained from scratch [26, 27]. In addition, fine-tuning also leads faster convergence than training from the scratch. Authors have examined transfer learning in a variety of ways [28–31].

When the target data-set is significantly smaller (12,444 for white blood cell) than the base data-set (ImageNet; with 1.2M training data [32]) transfer learning can be an efficient solution to enable Inception network training without overfitting and convergence problems [27]. This work initializes the Inception and ResNet convolutional and fully connected layers weights from ImageNet pre-trained models [32]. Admittedly, to do a comparative and comprehensive study, this work investigates full layers and last layer fine-tuning in the context of white blood cell image analysis. Indeed, current research in last layer fine-tuning uses frozen all layers ImageNet pre-trained weights before layers in logits section (see Fig. 2).

3.3 Deep Neural Networks in White Blood Cell Recognition

In recent years, there has been an increasing amount of literature on Inception [23, 33]. Inception moved from fully connected to sparsely connected architectures, even inside the convolution calculations [24]. This proposed Inception module technically leads to network dimension reduction derived from sparse connectivity among structure with factorized convolutional neural networks (see Fig. 2). It considers for example the case of a 1×1 convolutional which is followed with the rectified linear unit (ReLU) to add more non-linearity. Next, a 3×3 convolutional layer is employed. Auxiliary logits is also to solve instinct convergence problem in large deep learning layers. Several studies [24] have revealed that vanishing gradient problem is critical issue in last layers; near the end of training in which training progression identifies data details. Auxiliary logits with combination of average pool, convolutional 1×1, fully connected, and softmax activation is an efficient integrated solution to preserve the low-level detailed features most likely gone. In fact, auxiliary logits add slight weighted loss to compensate vanishing gradient problem that may exist in networks [34].

ResNet controls degradation problem via shortcut connections that bypass shallower sections to deeper networks [26]. It adjusts training error rate and identifies mappings. These shortcuts represent the residual mapping to be learned [26]. There are several possible explanations for this architecture. In this study ResNet with three variations namely; V1 50, 101 and 152 are applied [26].

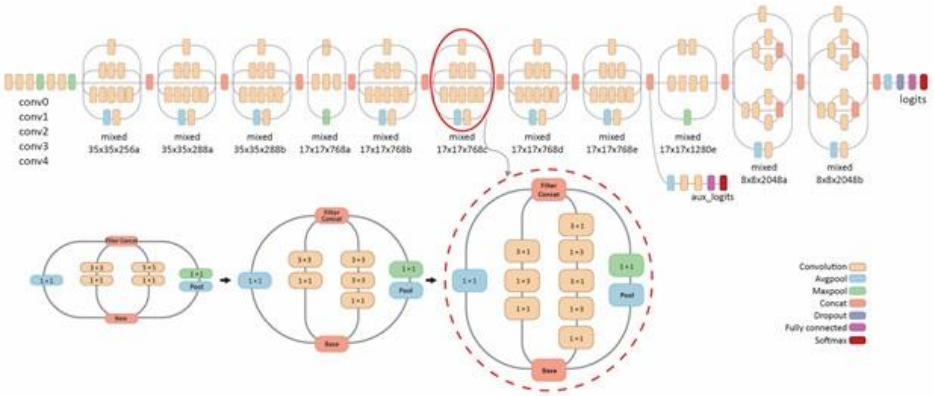

Figure 2. Inception V3 and factorized Inception module used in Inception V1 [23].

Adaptive learning rates techniques such as AdaGrad, RMSProp (Root Mean Square Propagation), and Adam in deep learning concept [36–40] are addressed. In this research adaptive learning rate between 0.001 and 0.0001 and 0.9 decay term for RMSProp optimizer is addressed [39]. Also, since the available data-set (12,444, see section 3.1) is limited compared with the number of model parameters (up to 5 millions) that need to be learned, regularization dropout and batch normalization with 32 batch size in training and 100 in evaluation process are applied. Indeed, this study goes with Inception (V1, V2, V3, V4), ResNet (V1 50, V1 101, V1 152) frameworks [23–26].

4. EXPERIMENTAL RESULTS

In this section, a set of 8-bits color scale RGB images in WBC imbalanced data-set is used. We have randomly chosen the 90% data (11200) to construct the training set after removing almost 10% of the data (1244) to be used for testing the proposed deep learning classifier. Training a deep learning model with intensive computing tasks, extreme number of parameters and on large data-sets significantly be accelerated with GPU's massively parallel architecture. In this study, single server with below specifications for computing platform is used: model: HP DL380 G9, CPU: 2x E5-2690v4 (35 MB L3 Cache, 2.6 GHz, 14C), RAM: 64 GB (8 ×8 GB) RAM DDR4 2133 MHz, HDD: 146 GB HDD 7.2k, GPU: ASUS GeForce GTX 1080, 1733 MHz, 2560 CUDA Cores, 8GB GDDR5 with CentOS 7.2 operating system. The model settings referred to official TensorFlow and TFslim documentations with NVIDIA GPUs support [22]. The GPU-enabled version of TensorFlow requires 64-bit Linux and Python 3.5. It should be noted that essential step before having TensorFlow library is to install the CUDA 8.0 Toolkit followed by cuDNN v5.1 [22, 41].

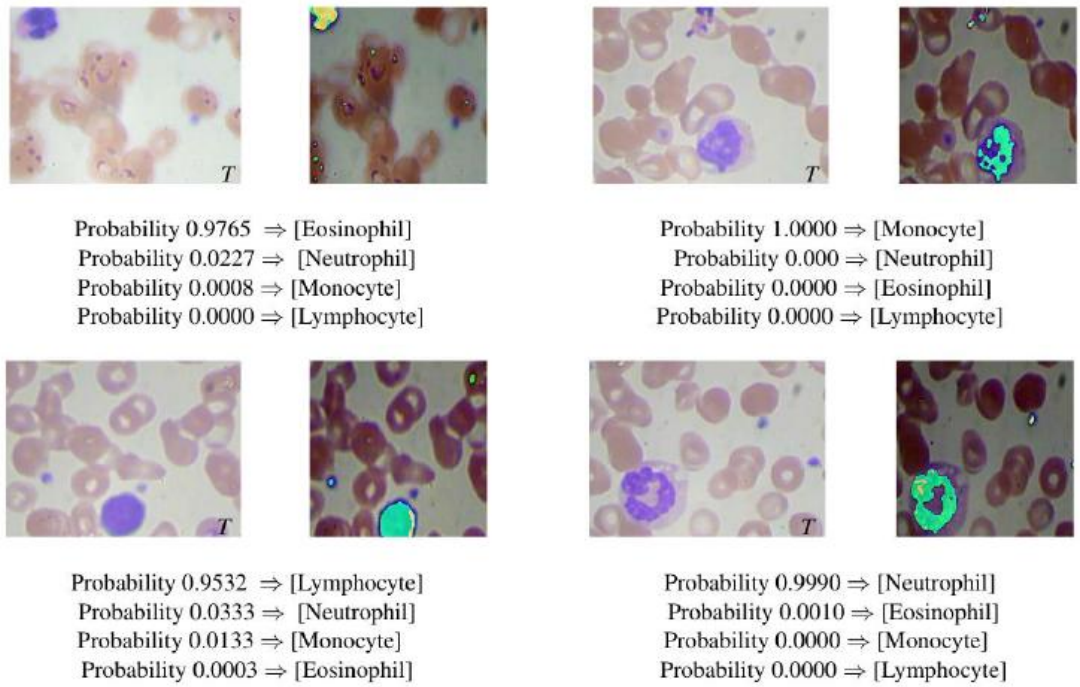

Figure 3. Prediction by inception V1, from top to bottom, left to right: Eosinophil, Monocyte, Lymphocyte, Neutrophil.

A 4 ×4 confusion matrix is used to represent the different possibilities of the set of WBC instances. The matrices are built on four rows and four columns: Neutrophils, Eosinophils, Lymphocytes, and Monocytes representing the known WBC classes whereas for each matrix, each row the values are normalized to sum to 1. Statistical performance measures [41] for each named WBC type & different deep learning frameworks (see section 3.3) summarized in tables [1, 2]. The result, as shown, indicates that for WBC samples using ResNet V1 50 & fine-tuning all layers 100% of known WBC types were classified as such, with this rate decreasing to 99.84% for ResNet V1 152 & all layers fine-tuning (see table 2) where the efficiency of ResNet V1 101 is also 99.46% with 3000 epochs and fine-tuning all layers. ResNet V1 50 with

2000 epochs and fine-tuning all layers reports 99.46% accuracy rate. As can be seen from the resultant tables the ResNet groups reported significantly more accuracy than the other Inception structures where in best scenario, Inception V2 with 3000 epochs and fine-tuning all layer revealed 96.76% which is less accurate in tests of white blood cell types. On the other hand, here is an obvious difference between false positive values associated with Inception structures and ResNets. It can be seen from the data in tables [1, 2] that the Inception models are with more false positive than the other ResNet groups with average 43.91, 9.66, respectively. Admittedly, the current study found that ResNets with aforementioned settings are more sensitive than Inception and lead more reliability in presence of fatal diseases.

The Cohens unweighted kappa coefficient of the Inception V1, V2, V3 and V4 with 3000 epochs fine-tuning all layers are 0.94, 0.94, 0.82 and 0.84, respectively where ResNet groups are above 0.97 and almost perfect in this WBCs classification. As shown in tables [1, 2] the findings would have been much more persuasive and convincing if ResNet will be run with higher epoch.

Table 1. Results of Fine-Tuning Last Layer for Different Models

Model name	Epochs	ACC	TP	TN	FP	FN	P	AUC	S
Inception V1	2000	0.4769	987	16	293	4	0.7710	0.4833	0.9959
Inception V1	3000	0.5707	894	119	190	97	0.8247	0.5334	0.9021
Inception V2	2000	0.4092	987	8	301	4	0.7663	0.4804	0.9959
Inception V2	3000	0.4484	443	251	59	557	0.8800	0.6073	0.4470
Inception V3	2000	0.4123	861	96	214	129	0.8009	0.3874	0.8688
Inception V3	3000	0.4476	625	197	112	366	0.8480	0.6281	0.6306
Inception V4	2000	0.38	849	76	233	142	0.7846	0.5025	0.8567
Inception V4	3000	0.4223	972	22	287	19	0.7720	0.2757	0.9808
ResNet V1 50	2000	0.8146	860	266	43	131	0.9523	0.8109	0.8678
ResNet V1 50	3000	0.8669	892	274	35	99	0.9622	0.8448	0.9001
ResNet V1 101	2000	0.8692	951	235	74	40	0.9278	0.7959	0.9596
ResNet V1 101	**3000**	**0.8723**	**897**	**276**	**33**	**94**	**0.9645**	**0.8580**	**0.9051**
ResNet V1 152	2000	0.8430	975	203	106	16	0.9019	0.7059	0.9838
ResNet V1 152	**3000**	**0.8746**	**956**	**239**	**70**	**35**	**0.9317**	**0.7924**	**0.9646**

Table 2. Results of Fine-Tuning All Layers for Different Models

Model name	Epochs	ACC	TP	TN	FP	FN	P	AUC	S
Inception V1	1000	0.7276	969	217	92	22	0.9132	0.8060	0.9778
Inception V1	2000	0.8730	976	276	33	24	0.9670	0.8900	0.9757
Inception V1	3000	0.9507	986	270	39	5	0.9619	0.8980	0.9949
Inception V2	1000	0.7684	764	309	1	226	0.9986	0.8848	0.7717
Inception V2	2000	0.9023	904	307	2	87	0.9977	0.9499	0.9122
Inception V2	3000	0.9676	976	307	3	15	0.9969	0.9828	0.9848
Inception V3	1000	0.7	965	188	121	26	0.8885	0.6004	0.9737
Inception V3	2000	0.65	941	186	123	50	0.8843	0.7464	0.9969
Inception V3	3000	0.8915	961	284	25	30	0.9834	0.9746	0.9330
Inception V4	1000	0.8	895	260	49	96	0.9480	0.8453	0.9031
Inception V4	2000	0.9184	946	280	29	45	0.9702	0.8830	0.9545
Inception V4	3000	0.9192	947	299	10	44	0.9895	0.9486	0.9556
ResNet V1 50	1000	0.9369	939	298	11	52	0.9884	0.9390	0.9475

ResNet V1 50	2000	0.9946	985	309	0	6	1	0.9969	0.9939
ResNet V1 50	**3000**	**1**	**991**	**309**	**0**	**0**	**1**	**1**	**1**
ResNet V101	1000	0.9692	990	273	36	1	0.9649	0.8830	0.9989
ResNet V101	2000	0.9923	991	304	0	5	0.9949	0.9923	1
ResNet V101	**3000**	**0.9946**	**990**	**304**	**5**	**1**	**0.9949**	**0.9833**	**0.9989**
ResNet V1 152	1000	0.8976	887	304	5	104	0.9943	0.9321	0.8950
ResNet V1 152	2000	0.8346	853	278	31	138	0.9649	0.8437	0.8607
ResNet V1 152	**3000**	**0.9984**	**989**	**309**	**0**	**2**	**1**	**0.9992**	**0.9979**

5. CONCLUSION

A very deep learning approach for WBC type detection is effective and reliable, while working under different and even unfavorable and adverse conditions. This work concentrates on the usefulness of proposed frameworks in connection with leukocytes recognition. In this research, various Inception and ResNet deep learning classification are presented and the use of these theories is outlined. The best results achieved where fine tunning all layers and ResNet groups settings are addressed (see table 2). The findings are expected to be persuasively supported by future work considering different deep learning segmentation algorithm, i.e., U-net: convolutional networks for biomedical image segmentation [43]. Also, in future work ResNet and Inception combination [25] with powerful distributed TensorFlow [44] to train a huge number of parameters will be addressed. Briefly, the empirical findings in this study provide a better understanding of hierarchical deep feature extraction process. One of the more significant findings to emerge from this study is that the possibility of extending the use of this framework to entire field of pathological analysis or other similar medical research.

REFERENCES

[1] H. Ramoser, V. Laurain, H. Bischof, et al., "Leukocyte segmentation and classification in blood-smear images," in 27th Annual International Conference of the Engineering in Medicine and Biology Society, IEEE-EMBS, 3371–3374, IEEE (2006).

[2] M. Habibzadeh, A. Krzyżak, and T. Fevens, "Comparative study of feature selection for white blood cell differential counts in low resolution images," in IAPR Workshop on Artificial Neural Networks in Pattern Recognition, 216–227, Springer (2014).

[3] M. Bruegel, D. Nagel, M. Funk, et al., "Comparison of five automated hematology analyzers in a university hospital setting: Abbott Cell-Dyn Sapphire, Beckman Coulter DxH 800, Siemens Advia 2120i, Sysmex XE-5000, and Sysmex XN-2000," Clinical Chemistry and Laboratory Medicine (CCLM) 53(7), 1057–1071 (2015).

[4] K. Ruzicka, M. Veitl, R. Thalhammer-Scherrer, et al., "The new hematology analyzer sysmex xe-2100: performance evaluation of a novel white blood cell differential technology," Archives of pathology & laboratory medicine 125(3), 391–396 (2001).

[5] E. Grimaldi and F. Scopacasa, "Evaluation of the abbott cell-dyn 4000 hematology analyzer," American Journal of Clinical Pathology 113(4), 497–505 (2000).

[6] S. H. Rezatofighi and H. Soltanian-Zadeh, "Automatic recognition of five types of white blood cells in peripheral blood," Computerized Medical Imaging and Graphics 35(4), 333–343 (2011).

[7] N. Ramesh, B. Dangott, M. E. Salama, et al., "Isolation and two-step classification of normal white blood cells in peripheral blood smears," Journal of pathology informatics 3 (2012).

[8] L. B. Dorini, R. Minetto, and N. J. Leite, "Semiautomatic white blood cell segmentation based on multiscale analysis," IEEE journal of biomedical and health informatics 17(1), 250–256 (2013).

[9] M. D. Joshi, A. H. Karode, and S. Suralkar, "White blood cells segmentation and classification to detect acute leukemia," International Journal of Emerging Trends and Technology in Computer Science (IJETICS) 2(3) (2013).

[10] S. Nazlibilek, D. Karacor, K. L. Ertürk, et al., "White blood cells classifications by surf image matching, pca and dendrogram," Biomedical Research 26(4) (2015).

[11] M. Sajjad, S. Khan, M. Shoaib, et al., "Computer aided system for leukocytes classification and segmentation in blood smear images," in International Conference on Frontiers of Information Technology (FIT), 99–104, IEEE (2016).

[12] Y. Dong, Z. Jiang, H. Shen, et al., "Evaluations of deep convolutional neural networks for automatic identification of malaria infected cells," in 2017 IEEE EMBS International Conference on Biomedical Health Informatics (BHI), 101–104 (2017).

[13] M. Imran Razzak and S. Naz, "Microscopic blood smear segmentation and classification using deep contour aware CNN and extreme machine learning," in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, (2017).

[14] C. L. Chen, A. Mahjoubfar, L.-C. Tai, et al., "Deep learning in label-free cell classification," Scientific reports 6, 21471 (2016).

[15] H. Chang, "Skin cancer reorganization and classification with deep neural network," arXiv preprint arXiv:1703.00534 (2017).

[16] Y. Liu, K. Gadepalli, M. Norouzi, et al., "Detecting cancer metastases on gigapixel pathology images," arXiv preprint arXiv:1703.02442 (2017).

[17] M. Aubreville, C. Knipfer, N. Oetter, et al., "Automatic classification of cancerous tissue in laserendomicroscopy images of the oral cavity using deep learning," arXiv preprint arXiv:1703.01622 (2017).

[18] A. Cruz-Roa, H. Gilmore, A. Basavanhally, et al., "Accurate and reproducible invasive breast cancer detection in whole-slide images: A deep learning approach for quantifying tumor extent," Scientific Reports 7, 46450 (2017).

[19] D. Wang, A. Khosla, R. Gargeya, et al., "Deep learning for identifying metastatic breast cancer," arXiv preprint arXiv:1606.05718 (2016).

[20] K. Sirinukunwattana, S. E. A. Raza, Y. W. Tsang, et al., "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," IEEE Transactions on Medical Imaging 35, 1196–1206 (2016).

[21] D. Parthasarathy, "WBC-classification." https://github.com/dhruvp/wbc-classification/tree/master/Original_Images (2017). [Google License, Online; accessed 23-August-2017].

[22] T. G. Developers, "Installing tensorflow on ubuntu." https://www.tensorflow.org/install/install_linux (2017). [Google License, Online; accessed 01-August-2017].

[23] C. Szegedy, V. Vanhoucke, S. Ioffe, et al., "Rethinking the inception architecture for computer vision," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2818–2826 (2016).

[24] C. Szegedy, W. Liu, Y. Jia, et al., "Going deeper with convolutions," CoRR abs/1409.4842 (2014).

[25] C. Szegedy, S. Ioffe, V. Vanhoucke, et al., "Inception-v4, inception-resnet and the impact of residual connections on learning.," in AAAI, 4278–4284 (2017).

[26] K. He, X. Zhang, S. Ren, et al., "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 770–778 (2016).

[27] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, et al., "Convolutional neural networks for medical image analysis: Full training or fine tuning?," IEEE transactions on medical imaging 35(5), 1299–1312 (2016).

[28] J. Yosinski, J. Clune, Y. Bengio, et al., "How transferable are features in deep neural networks?," in Advances in neural information processing systems, 3320–3328 (2014).

[29] H. Daume III and D. Marcu, "Domain adaptation for statistical classifiers," Journal of Artificial Intelligence Research 26, 101–126 (2006).

[30] B. Zadrozny, "Learning and evaluating classifiers under sample selection bias," in Proceedings of the twenty-first international conference on Machine learning, 114, ACM (2004).

[31] L. Gui, R. Xu, O. Lu, et al., "Negative transfer detection in transductive transfer learning," International Journal of Machine Learning and Cybernetics , 1–13 (2017).

[32] M. T. Bahadori, Y. Liu, and D. Zhang, "A general framework for scalable transductive transfer learning," Knowledge and information systems 38(1), 61–83 (2014).

[33] T. T. Authors, "Tensorflow-slim image classification library." https://github.com/tensorflow/models/tree/master/slim (2016). [Apache License, Version 2.0, Online; accessed 17-July-2017].

[34] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” CoRR abs/1502.03167 (2015).

[35] C.-Y. Lee, S. Xie, P. Gallagher, et al., “Deeply-supervised nets,” in Artificial Intelligence and Statistics, 562–570 (2015).

[36] M. D. Zeiler, “ADADELTA: an adaptive learning rate method,” arXiv preprint arXiv:1212.5701 (2012).

[37] T. Schaul, S. Zhang, and Y. LeCun, “No more pesky learning rates,” in International Conference on Machine Learning, 343–351 (2013).

[38] D. Kingma and J. Ba, “Adam: A method for stochastic optimization,” arXiv preprint arXiv:1412.6980 (2014).

[39] G. Hinton, N. Srivastava, and K. Swersky, “Rmsprop: Divide the gradient by a running average of its recent magnitude,” Neural networks for machine learning, Coursera lecture 6e (2012).

[40] I. Sutskever, J. Martens, G. Dahl, et al., “On the importance of initialization and momentum in deep learning,” in International conference on machine learning, 1139–1147 (2013).

[41] T. N. Developers, “NVIDIA GPUs - The Engine of Deep Learning.” https://developer.nvidia.com/deep-learning (2017). [Google License, Online; accessed 01-August-2017].

[42] J. R. Landis and G. G. Koch, “The measurement of observer agreement for categorical data,” biometrics , 159–174 (1977).

[43] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in International Conference on Medical Image Computing and Computer-Assisted Intervention, 234–241, Springer (2015).

[44] T. G. Developers, “Distributed tensorflow.” https://www.tensorflow.org/deploy/distributed (2017). [Google License, Online; accessed 01-August-2017].