Classification of Leukocytes: Comparison of Different Feature Extraction and Machine Learning Approaches

Fernanda D'Amico Silva, Milena A. Cardoso, Viviane R. Sommer, Evair B. Severo, Ramon G. da Silva, Valeria Tafoya-Martinez, Carolina Q. Cardoso, Ivan L. R. Silva, Victor H. A. Ribeiro, Gilberto Reynoso-Meza

ICPR 2022
AMERICAS
11th International Conference on Production Research
Curitiba - Brazil

INTRODUCTION

- Blood cells can be separated into three types: erythrocytes, leukocytes, and platelets;
- Complete Blood Count (CBC) provides information for diagnosing numerous diseases (Silva et al. 2022);
- The traditional method is performed manually, and it is time-consuming and susceptible to errors;
- One of the challenges is to automate the classification of leukocytes or white blood cells (WBC), these cells are part of the immune system and protect the body against infections.

OBJECTIVE

- Compare different approaches to execute feature extraction and machine learning techniques to leukocyte images;
- The images were acquired using a novel hematological platform that performs CBC, and it was developed by Hilab, a healthtech company based on Curitiba;
- The main goal is to improve the reliability of the classification of leukocytes, reduce the time spent evaluating CBC, and consequently support laboratories, physicians, and patients.

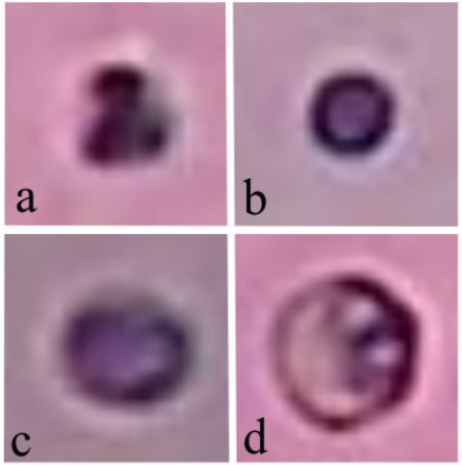
REVIEW OF LITERATURE

- The papers reviewed were from the past four years. The papers selected focused on the classification of WBC;
- Tiwari et al., 2018 proposed a double layer neural network and compared it with the application of Naive Bayes and Support Vector Machine;
- Hedge et al., 2019 proposed a comparison of image processing and deep learning techniques;
- Liang et al., 2018 combined a convolutional neural network (CNN) with a recursive neural network (RNN) to classify WBC;
- Macawile et al. 2018 applied three neural networks to classify WBC.

METHODS

- The images used in the experiments were extracted from CBC with the equipment developed by Hilab;
- The whole image with cells was labeled by experts and then cropped;
- Two image processing techniques were applied to extract the features, and three machine learning techniques were applied to the images.

METHODS



Crops of WBC from the dataset: (a) neutrophil, (b) lymphocyte, (c) monocyte, and (d) eosinophil.

IMAGE PROCESSING

- Image processing resides in removing noise and irregularities present on digital images (Chitradevi and Srimathi, 2014);
- Here, two image processing techniques were applied to extract the features from WBC individual images: Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP);
- HOG was proposed by Dalal and Triggs in 2005 and calculates a local 1-D histogram of gradient directions for each cell present in the image;
- LBP was developed by Ojala et al. in 1996, and it is a technique that describes the texture and shape of an image.

MACHINE LEARNING

- Classification is a machine learning technique to predict labels for data instances (Soofi and Awan, 2017);
- Here, three techniques were applied: Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), and a Convolutional Neural Network (CNN), which is a deep learning technique;
- SVM is a technique that separates classes with the largest gap between the support vectors. It was first developed for two class problems that were linearly separated, but it was extended to multi-class problems and non-linearly separable using kernels (Chauhan et al., 2019).

MACHINE LEARNING

- XGBoost was developed by Chen et al., 2016, and it consists of an ensemble technique that adds new models recursively until no improvements are detected. To reduce loss, a gradient descent algorithm is used;
- CNN are deep learning algorithms that, unlike classical ML approaches mentioned previously, can extract multiple features independently and provide several levels of abstraction. The algorithms convolve filters on input images, and this operation can extract features from the entire image (Chauhan et al., 2018).

EXPERIMENTS

- The dataset used in the experiments consists of the four most common types of leukocytes: neutrophils, lymphocytes, monocytes, and eosinophils;

Quantity of cell images by cell type

Cell Type	Train	Test
Neutrophil	797	341
Lymphocyte	496	213
Monocyte	183	79
Eosinophil	35	15

EXPERIMENTS

- The dataset used in the experiments consists of the four most common types of leukocytes: neutrophils, lymphocytes, monocytes, and eosinophils;
- The classifiers were optimized using a grid search;
- The CNN used was trained from scratch, it was based on the All Convolutional Net (Springenberg et al., 2015);
- The experiments were performed using hold-out, with 70% of the data dedicated to training and 30% dedicated to validating.

EXPERIMENTS

- The dataset used in the experiments is imbalanced, so metrics as precision, recall and F1-score were used.

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 - score = \frac{2.\,Precision\,.\,Recall}{Precision + Recall}$$

ICPR 2022 AMERICAS
11th International Conference on Production Research
Curitiba - Brazil

PUCPR
GRUPO MARISTA
SCIENTIA, VITA ET FIDES

RESULTS AND DISCUSSION

- The results were calculated for the four types of cells, and the mean was calculated. The best results for each metric are highlighted in bold.

	SVM		XGBoost		CNN
	HOG	LBP	HOG	LBP	
Precision	**0.8224**	0.4786	0.7843	0.7969	0.732
Recall	**0.7183**	0.4744	0.6273	0.5266	0.608
F1-score	**0.7486**	0.4126	0.6338	0.527	0.632

RESULTS AND DISCUSSION

- The application of HOG, followed by the SVM classifier, obtained the best results among the selected techniques, with a precision of 0.8224, a recall of 0.7183, and a F1-score of 0.7486;
- For all the methods, the recall was significantly lower than the precision;
- The performance was better applying HOG over LBP, which might be due to the characteristics of the techniques. For cell images, the format is more relevant than the texture, also as the images have low resolution, the texture extraction is limited;
- SVM performed better than XGBoost and CNN, which can be explained by the small quantity of samples. XGBoost and CNN require a large amount of data to perform better, as SVM is a large margin classifier, it performs better with a smaller amount of samples.
- CNN results could be improved by increasing the model size or using a hyperparameter search.

CONCLUSIONS

- This paper compared different techniques for feature extraction and machine learning classifiers to classify leukocytes, aiming to reduce the time spent evaluating CBC;
- HOG combined with SVM obtained the best results amidst the proposed techniques;
- A suggestion for future works is implementing a class balancing technique to evaluate if the results improve;
- Further experiments with the CNN, increase the model's size, apply data augmentation techniques, and implement a hyperparameter optimization.

REFERENCES

1. Gasparin, A.T., Araujo, C.I.F., Schmitt, P. et al. Hilab system, a new point-of-care hematology analyzer supported by the Internet of Things and Artificial Intelligence. Sci Rep 12, 10409 (2022) https://doi.org/10.1038/s41598-022-13913-8

2. Silva, F., Severo, E., Ribeiro, V., Reynoso-Meza, G. (2022). Detection and classification of abnormal red blood cells with computational intelligence techniques: a review. Revista Principia - Divulgação Científica E Tecnológica do IFPB, 0. doi:http://dx.doi.org/10.18265/1517-0306a2021id6456

3. Chitradevi, B., Srimathi, P. (2014). An overview on image processing techniques. International Journal of Innovative Research in Computer and Communication Engineering, 2(11), 6466-6472.

4. Nazlibilek, S., Karacor, D., Ercan, T., Sazli, M. H., Kalender, O., Ege, Y. (2014). Automatic segmentation, counting, size determination and classification of white blood cells. Measurement, 55, 58-65.

5. Theerapattanakul, J., Plodpai, J., Pintavirooj, C. (2004, November). An efficient method for segmentation step of automated white blood cell classifications. In 2004 IEEE Region 10 Conference TENCON 2004. (pp. 191-194). IEEE.

REFERENCES

6. Su, M. C., Cheng, C. Y., Wang, P. C. (2014). A neural-network-based approach to white blood cell classification. The scientific world journal, 2014.

7. Dalal, N., Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR’05) (Vol. 1, pp. 886-893). Ieee.

8. Ojala, T., Pietikäinen, M., Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. Pattern recognition, 29(1), 51-59.

9. Rahim, M. A., Azam, M. S., Hossain, N., Islam, M. R. (2013). Face recognition using local binary patterns (LBP). Global Journal of Computer Science and Technology.

10. Haralick, R. M., Shanmugam, K., Dinstein, I. H. (1973). Textural features for image classification. IEEE Transactions on systems, man, and cybernetics, (6), 610-621.

REFERENCES

11. Tiwari, P., Qian, J., Li, Q., Wang, B., Gupta, D., Khanna, A., de Albuquerque, V. H. C. (2018). Detection of subtype blood cells using deep learning. Cognitive Systems Research, 52, 1036-1044.

12. Hegde, R. B., Prasad, K., Hebbar, H., Singh, B. M. K. (2019). Comparison of traditional image processing and deep learning approaches for classification of white blood cells in peripheral blood smear images. Biocybernetics and Biomedical Engineering, 39(2), 382-392.

13. Liang, G., Hong, H., Xie, W., Zheng, L. (2018). Combining convolutional neural network with recursive neural network for blood cell image classification. IEEE access, 6, 36188-36197.

14. Macawile, M. J., Quiñones, V. V., Ballado, A., Cruz, J. D., Caya, M. V. (2018, April). White blood cell classification and counting using convolutional neural network. In 2018 3rd International conference on control and robotics engineering (IC-CRE) (pp. 259-263). IEEE.

15. Soofi, A. A., Awan, A. (2017). Classification techniques in machine learning: applications and issues. Journal of Basic & Applied Sciences, 13, 459-465.

REFERENCES

16. Chauhan, V. K., Dahiya, K., Sharma, A. (2019). Problem formulations and solvers in linear SVM: a review. Artificial Intelligence Review, 52(2), 803-855.

17. Chen, T., Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).

18. Ogunleye, A., Wang, Q. G. (2019). XGBoost model for chronic kidney disease diagnosis. IEEE/ACM transactions on computational biology and bioinformatics, 17(6), 2131-2140.

19. Chauhan, R., Ghanshala, K. K., Joshi, R. C. (2018, December). Convolutional neural network (CNN) for image detection and recognition. In 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC) (pp. 278-282). IEEE.

20. Srinivas, S., Sarvadevabhatla, R. K., Mopuri, K. R., Prabhu, N., Kruthiventi, S. S., Babu, R. V. (2016) “A taxonomy of deep convolutional neural nets for computer vision.”

REFERENCES

20. Srinivas, S., Sarvadevabhatla, R. K., Mopuri, K. R., Prabhu, N., Kruthiventi, S.S., Babu, R. V. (2016) “A taxonomy of deep convolutional neural nets for computervision.”

21. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A. (2014) “Object detectors emerge in deep scene cnns.”

22. Sharma, N., Jain, V., Mishra, A. (2018). An analysis of convolutional neural networks for image classification. Procedia computer science, 132, 377-384.

23. Tigner, A., Ibrahim, S. A., & Murray, I. (2020). Histology, white blood cell.

24. Ribeiro, V. H. A., Reynoso-Meza, G. (2020). Ensemble learning by means of a multi-objective optimization design approach for dealing with imbalanced data sets. Expert Systems with Applications, 147, 113232.

25. Springenberg, J., Dosovitskiy, A., Brox, T., Riedmiller, M. (2015). Striving for Simplicity: The All Convolutional Net. In ICLR (workshop track).

CONTACT

Fernanda D'Amico Silva

fernanda.damico@pucpr.edu.br

linkedin.com/in/fernanda-d-amico-silva-4b1604129/

orcid.org/0000-0002-7871-202X

"Nature is written in mathematical language"
— Galileo Galilei, Italian astronomer, physicist and engineer

ICPR 2022 AMERICAS
11th International Conference on Production Research
Curitiba - Brazil