

Exploring Convolutional Neural Networks for the Classification of Acute Lymphoblast Leukemia Blood Cell Images

Andrey Trubnikov^{1*} and Dmitry Savelyev^{1,2}

¹Samara National Research University, 34 Moskovskoye shosse, Samara 443086, Russia

²Image Processing Systems Institute, NRC "Kurchatov Institute", 151 Molodogvardeyskaya str., Samara 443001, Russia

*e-mail: antei3mer@gmail.com

Abstract. This paper introduces a novel approach to blood cell classification using convolutional neural networks (CNNs). Our emphasis lies in methodological insights derived from comparative analyses across various CNN architectures, implemented in Python with PyTorch such as ResNet152V2, Xception, EfficientNetB5, and EfficientNetV2M. In exploring model capabilities and limitations, we observe that the simple and shallow architectures is not sufficient to learn patterns compared to more complex networks. During research we have discovered that EfficientNetV2M demonstrate stable results with mean $F_1=0.891$ score, which is higher compare to other models. Saliency maps are applied to reveal significant regions within or near cells, offering nuanced insights into morphological influences like cell shape. Extensive data augmentation allows one effectively mitigate overfitting, aligning train learning curves with validation and test splits. Our study is methodologically rich, comprising a meticulous data-specific overview, a robustness analysis through multiple model training, a systematic architecture comparison, and an in-depth examination using saliency maps. © 2024 Journal of Biomedical Photonics & Engineering.

Keywords: deep learning; acute lymphoblastic leukemia; convolutional neural network; blood cell classification; saliency maps; Python.

Paper #9045 received 15 Dec 2023; revised manuscript received 20 Dec 2023; accepted for publication 26 Dec 2023; published online 20 Feb 2024. doi: [10.18287/JBPE24.10.010302](https://doi.org/10.18287/JBPE24.10.010302).

1 Introduction

Acute lymphoblastic leukemia (ALL) is a malignant disease of the hematopoietic system, which is caused by extensively proliferating lymphoblasts (immature lymphocytes). It is part of a large group of leukemias [1]. Unlike other leukemias, it is a rapidly progressive disease and in the absence of appropriate and timely treatment, it can lead to death in a few months [1]. About 85% of all cases are among children and adolescents under age 15 [2].

Literally there are several organs in the human body producing blood cells: bone marrow, lymph nodes, and spleen. ALL starts in the bone marrow, which, in case of dysfunction, can release immature leukocytes into the blood, which can sometimes affect other human organs (liver, spleen, brain, and others). The process of formation

and maturation of lymphocytes is called lymphocytopoiesis and is shown in Fig. 1.

A distinctive feature of ALL is a strong morphological similarity between normal lymphoid cells and abnormalities [3]. Early diagnostic methods were based on morphological examination of peripheral blood smears, but now, cytogenetic examination, bone marrow biopsy, and lumbar puncture have become the main diagnostic tools [2]. Such approaches require highly qualified doctors and pathologists, special conditions and equipment. Therefore, such tests are usually not available to everyone.

The use of neural networks makes it possible to build automatic classification systems for digital images of blood cells without an expert knowledge. The possibilities of using machine learning are widely studied in related fields including this particular problem [4, 5].

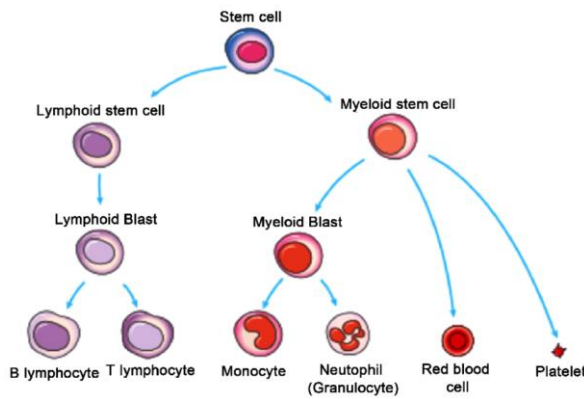


Fig. 1 Hematopoietic stem cell differentiation [2].

The purpose of this study is to analyze blood cell data with respect to its specifics and to conduct a comparative study of various neural network architectures.

The scientific novelty of this study lies in the fact that during the work, the distinctive features of cellular images were taken into account and the results of the trained models were interpreted, which was not carried out in similar works [6].

2 Methods

2.1 Microscopic Images

Data exploration is the first stage of classification pipeline, it includes calculations of numerical characteristics and revealing distinctive features that can be used in the future. When working with cellular images, it is worth paying attention to several features:

- small amount of available data,
- dataset imbalance,
- specific errors and noise,
- pre-segmentation,
- requires deep domain knowledge to manually construct features,
- high intra-class variability,
- variation among subjects.

2.1.1 Data Availability

Usually, medical data falls under the category of personal data, so some preliminary administrative work is required to work with them. Therefore, medical data is usually collected only for a specific and pre-approved purpose, which greatly affects their overall availability. Nevertheless, medical slowly become more and more available [7].

2.1.2 Dataset Imbalance

The main sources of medical data are various diagnostic procedures and tests, such as ECG, CT, Echo, biochemical blood tests, and others. The purpose of diagnosis is to identify the disease. All this leads to a high percentage of healthy subjects in the total sample [8].

2.1.3 Specific Errors

Depending on the subject area, the data may contain specific errors. For example, radiographic images inevitably contain impulse noise, which can be effectively eliminated by a median filter [9]. When working with digital images of microscopic photographs of cells, several sources of noise can be identified [10–12]:

- lightning,
- the optical system of the microscope and camera,
- staining.

There are many ways to eliminate the errors introduced by these factors [13]. In the analyzed data, lighting and coloring errors were minimized by applying a special color normalization method.

2.1.4 Preliminary Segmentation

Often digital microscopic photographs require prior processing in the form of segmentation. This step is required due to the fact that initially the whole cells ensembles are presented in the images, from which it is necessary to isolate individual cells. The error introduced by this procedure depends on whether manual segmentation is performed (then the error depends on the qualifications and expertise of the person conducting it) or automatic (in this case, the error value will depend on the specific algorithm) segmentation [14, 15]. Typical pre-segmentation errors include: the presence of more than one cell in the resulting image, violation of cell integrity, selection of an empty area.

2.1.5 High Level of Required Qualifications

Manually constructing features from microscopic photographs of cells will require one or even several specialists with knowledge of the subject area (e.g., an oncologist and a cytologist).

2.1.6 High Intra-Class Variability

Depending on many factors (the developmental process, the etiology of the disease, anomalies, etc.), cell characteristics for every class may vary (for example, morphological features such as size, shape, color, granularity, cytoplasmic volume, etc.) [16].

2.1.7 Variation Among Subjects

Given the original format of most medical samples (the presence of patient's id feature), this factor deserves special attention. The values of cell features within the same class can also vary significantly for two separate subjects due to the individual distinctions of the organism [3].

2.2 Data Augmentation

There are many different approaches to augmenting cellular data [17, 18], but most of them are based on utilization data specific features. So, in addition to the

features listed above that are important for building an adequate model, we can highlight characteristics of images, through the use of which it is possible to increase the size of the training set. Main feature of cells is their axial and point symmetry (in this regard). Thus, there is an invariance of the image class to the use of random rotation and mirror reflection.

In addition, to increase the resistance of the classifier to random noise and for regularization, a random error can be introduced into the training data [19].

When using image augmentation techniques, it is also worth to consider the ratio of sizes of the final and initial samples and the amplitude values of the characteristics used.

Data distinctive features (class invariance to image rotation and introduction of small noises) can be used for their augmentation (rotation by a random angle, introduction of random noise). This approach will partially overcome the initial limitation associated with the availability of data. These values should be selected empirically so as not to make significant changes to the distribution of original features.

2.3 Training Parameters

The data set consists of 10,581 digital photographs of lymphoid blood cells, among which two classes are represented: normal cells and abnormal ones

(lymphoblasts) [20]. The files have a .bmp extension and a resolution of 600×600 .

Different machine learning or deep learning tools can be used for cells classification. Previous studies have compared various machine learning models with convolutional neural networks and found out that deep learning methods achieve better results [9].

Also, a big advantage of deep learning methods is the absence of the need for manual feature engineering, which, as mentioned above, is a significantly difficult when working with cell images.

Several architectures of convolutional neural networks were selected for the further experiments. All of them represent different approaches to building deep networks: VGG16 (wide networks), ResNet152V2 (deep networks), EfficientNetB5 (result of NAS), EfficientNetV2M (second iteration, also result of NAS) [21, 22]. Also, a hypothesis was formulated and tested that a relatively narrow and shallow network, without preliminary training, would not able to learn from the initial data and carry out a successful classification for the test set.

As all of selected models were pretrained, suitable strategy for changing the learning rate parameter was chosen [23].

Each model was trained 6 times with different random seeds to further estimate how stochastic nature of the training process affect results.

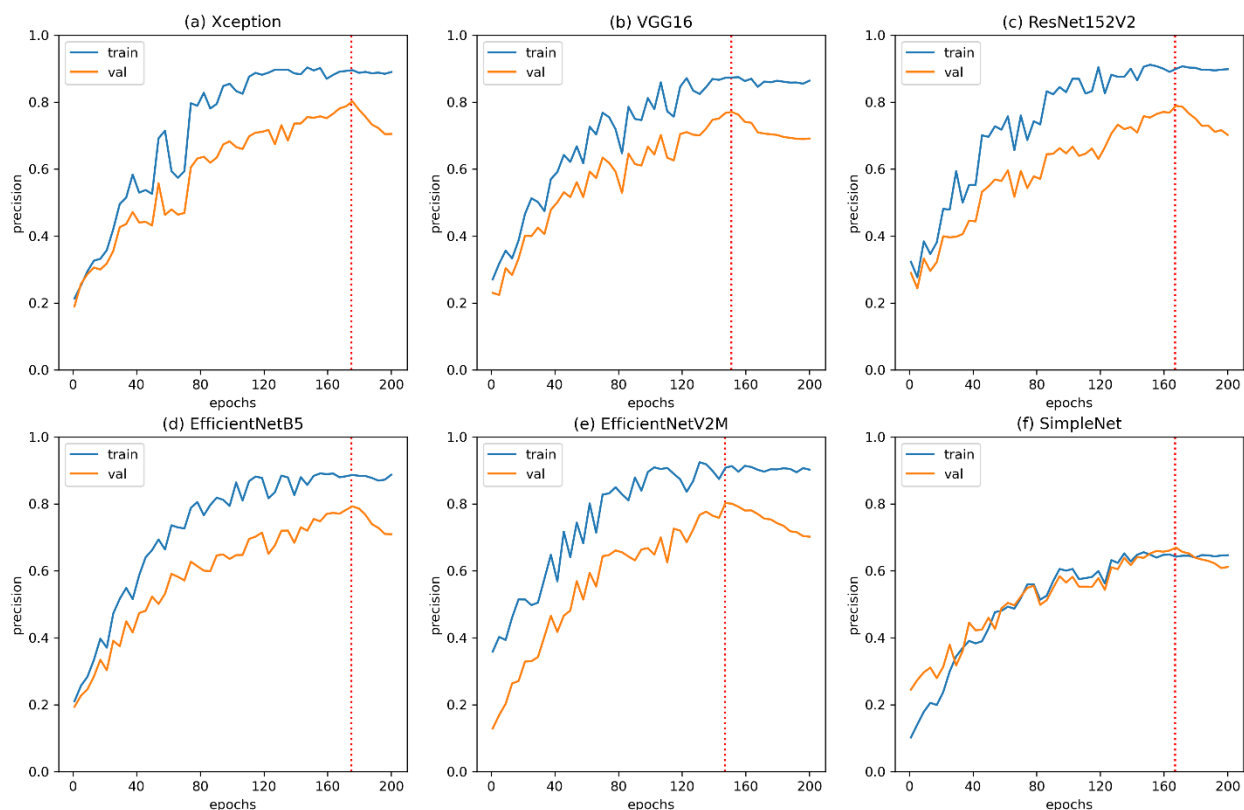


Fig. 2 Precision curves for selected models: (a) Xception, (b) VGG16, (c) ResNet152V2, (d) EfficientNetB5, (e) EfficientNetV2M, and (f) SimpleNet.

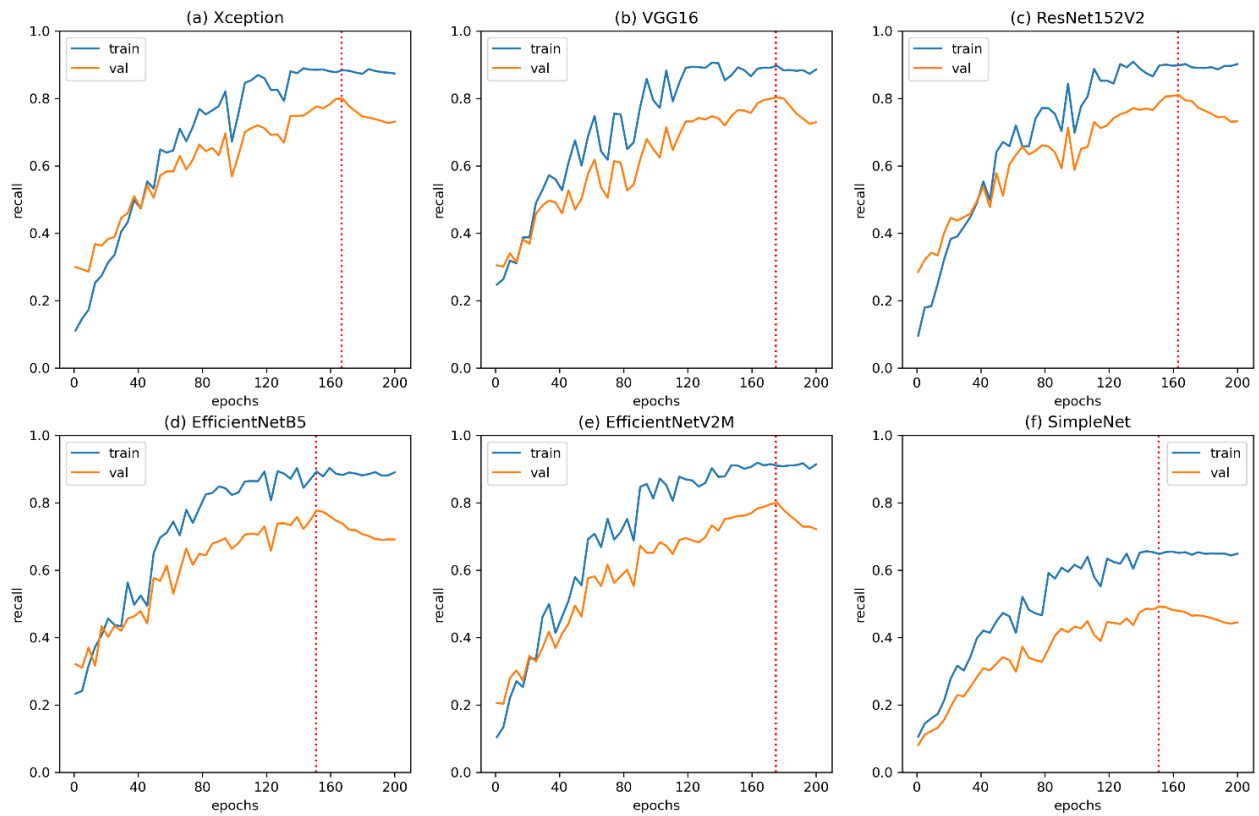


Fig. 3 Recall curves for selected models: (a) Xception, (b) VGG16, (c) ResNet152V2, (d) EfficientNetB5, (e) EfficientNetV2M, and (f) SimpleNet.

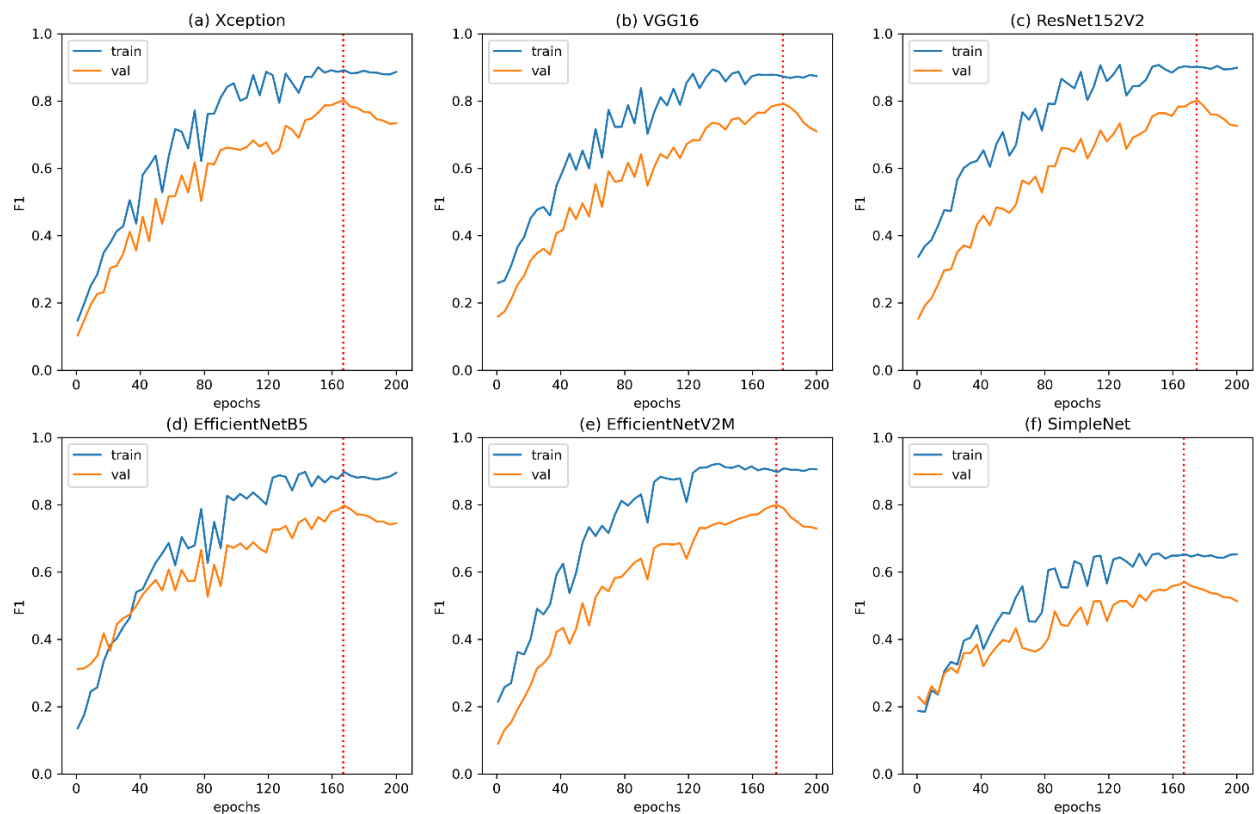


Fig. 4 F_1 score curves for selected models: (a) Xception, (b) VGG16, (c) ResNet152V2, (d) EfficientNetB5, (e) EfficientNetV2M, and (f) SimpleNet.

The original dataset was split in a ratio of 90:5:5 (training, validation, and test sets). The partitioning was performed with respect to the fact that cell properties, in particular morphological properties, can vary greatly among subjects. Thus, data splits have no intersection by subjects (patients).

Several metrics were chosen for evaluation by the following Eqs. (1–3):

$$precision = \frac{TP}{TP+FP}, \quad (1)$$

$$recall = \frac{TP}{TP+FN}, \quad (2)$$

$$F_1 = \frac{2*precision*recall}{precision+recall}. \quad (3)$$

Since the original dataset is not balanced, an *adequate* and *interpretable* estimate can be obtained based only on all three values.

At the same time, saliency maps were used to interpret the results of the model operation [24–27].

3 Results and Discussion

Each model was trained 6 times for 200 epochs. Learning curves for different metrics presented in Figs. 2–4.

Vertical red line on each plot depicted epoch with best F_1 value. For each plot it is can be seen that models start gradually decline or stagnate after reaching their best F_1 value across all epochs. For further studies weights from these epochs will be taken.

Results of models evaluation on training, validation, and test splits are presented in Tables 1–3. Metric values presented in the following format: mean $\pm$ std.

It is easy to observe plateau on the right side of each plot in Figs. 2–4. This is because, each model has same initial learning rate and scheduler (exponential decay), according to this approximate value of learning rate after 150 epochs is $4.55e-9$ (therefore plateau is to be expected).

Table 1 Results of models evaluation on the training set.

Model	Time, s	Precision	Recall	F_1	dF_1 , %
Xception	417 $\pm$ 7	0.891 $\pm$ 0.009	0.88 $\pm$ 0.004	0.886 $\pm$ 0.005	4.186
VGG16	557 $\pm$ 6	0.864 $\pm$ 0.008	0.888 $\pm$ 0.014	0.875 $\pm$ 0.008	2.974
ResNet152V2	550 $\pm$ 12	0.901 $\pm$ 0.012	0.897 $\pm$ 0.009	0.899 $\pm$ 0.008	5.738
EfficientNetB5	777 $\pm$ 12	0.885 $\pm$ 0.009	0.887 $\pm$ 0.008	0.886 $\pm$ 0.006	4.221
EfficientNetV2M	671 $\pm$ 13	0.904 $\pm$ 0.006	0.911 $\pm$ 0.009	0.907 $\pm$ 0.006	6.727
SimpleNet	402 $\pm$ 9	0.645 $\pm$ 0.008	0.65 $\pm$ 0.014	0.647 $\pm$ 0.009	–23.846

Table 2 Results of models evaluation on the validation set.

Model	Time, s	Precision	Recall	F_1	dF_1 , %
Xception	408 $\pm$ 6	0.847 $\pm$ 0.009	0.848 $\pm$ 0.007	0.847 $\pm$ 0.007	5.057
VGG16	561 $\pm$ 5	0.824 $\pm$ 0.008	0.862 $\pm$ 0.009	0.843 $\pm$ 0.004	4.474
ResNet152V2	542 $\pm$ 9	0.843 $\pm$ 0.011	0.863 $\pm$ 0.005	0.851 $\pm$ 0.004	5.528
EfficientNetB5	779 $\pm$ 8	0.847 $\pm$ 0.008	0.833 $\pm$ 0.012	0.842 $\pm$ 0.007	4.152
EfficientNetV2M	646 $\pm$ 8	0.855 $\pm$ 0.01	0.852 $\pm$ 0.009	0.854 $\pm$ 0.007	5.825
SimpleNet	439 $\pm$ 4	0.716 $\pm$ 0.012	0.524 $\pm$ 0.008	0.605 $\pm$ 0.008	–25.036

Table 3 Results of models evaluation on the test set.

Model	Time, s	Precision	Recall	F_1	dF_1 , %
Xception	403 $\pm$ 4	0.805 $\pm$ 0.012	0.882 $\pm$ 0.011	0.841 $\pm$ 0.009	1.122
VGG16	562 $\pm$ 15	0.833 $\pm$ 0.004	0.86 $\pm$ 0.008	0.846 $\pm$ 0.005	1.723
ResNet152V2	550 $\pm$ 15	0.832 $\pm$ 0.01	0.88 $\pm$ 0.012	0.854 $\pm$ 0.005	2.685
EfficientNetB5	797 $\pm$ 7	0.847 $\pm$ 0.009	0.915 $\pm$ 0.007	0.881 $\pm$ 0.006	5.932
EfficientNetV2M	633 $\pm$ 9	0.868 $\pm$ 0.011	0.911 $\pm$ 0.013	0.891 $\pm$ 0.006	6.774
SimpleNet	430 $\pm$ 3	0.692 $\pm$ 0.110	0.669 $\pm$ 0.165	0.680 $\pm$ 0.098	–18.236

All models were trained 6 times to estimate variance created by stochastic nature of the training process (with different random seeds). Kruskal-Wallis test was performed to determine whether there is significant difference between model's results. With significance level of 0.05 null-hypothesis was rejected indicating that there is statistical difference in results.

The EfficientNetV2M showed the best results, while its inference time differs little from other models (86.5 ms more than the average). The SimpleNet showed worst results and demonstrated the impossibility of applying small and primitive architectures on the given dataset.

Saliency maps depicted in Fig. 5 contains two examples: a normal cell and an anomaly. It can be seen that the attention of the model is distributed unevenly over the area of the cell (points outside the boundary are acceptable, since the used method for constructing saliency maps is quite noisy [24]). This means that some parts of the cell had a greater influence on the process of making decision by the model than others, so the information presented in these parts of the image may be used in a decision support system for a medical specialist.

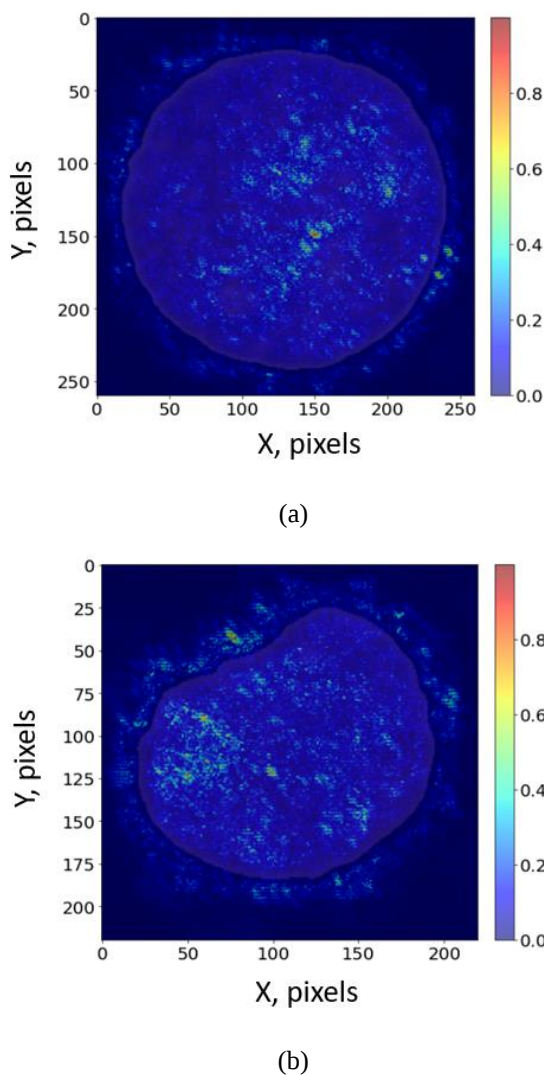


Fig. 5 Saliency maps for EfficientNetV2M: a) normal cell, b) abnormal cell.

As we can see, examined simple model (SimpleNet) clearly does not have ability to learn patten from our data comparable with bigger and complex networks.

Exploring saliency maps for the best model (EfficientNetB5) also gives some insight. All important regions bounded are located within a cell or its vicinity, therefore we can conclude, that model has learned patterns possessed by a cell itself. Important regions are located both inside and on the border of a cell indicating that learned patterns influenced by some morphological cell features such as its shape.

One more observation can be made. Extensive usage of data augmentation techniques allowed us to overcome overfitting problem, lowering train learning curves and make it closer to those for validation and test splits.

While our study has provided valuable insights into the performance and interpretability of different neural network architectures for blood cell classification, it is essential to highlight the distinct contributions and differences of our work compared to existing literature.

In contrast with other researches we started our research from analyzing data and it is specific [28, 29]. We developed set of techniques to artificially increase sample size based on determined data characteristics and existing studies [30, 31].

Our investigation goes beyond model accuracy and includes a comprehensive analysis of model robustness what is absent in some existing works [30, 32]. By training each model multiple times with different random seeds, we aimed to estimate the variance introduced by the stochastic nature of the training process [33, 34]. This nuanced approach enhances the reliability of our findings and provides a more comprehensive understanding of the models' behavior.

We systematically compared the performance of multiple architectures, including EfficientNetV2M, SimpleNet, and EfficientNetB5. The use of both ANOVA and Kruskal-Wallis tests allowed us to rigorously assess the statistical differences among these models, providing a robust foundation for our conclusions [35].

The inclusion of saliency maps in our analysis adds an interpretability dimension to our study. We not only present model predictions but also explore where the models focus their attention. This qualitative analysis sheds light on the features influencing the decision-making process, offering potential applications in decision support systems for medical specialists [36–38].

In the future, it is planned to study the possibility of using auxiliary information to address intraclass and interclass variability [39–40]. It is also planned to extend comparison list of models and include transformer-based architectures (ViT-based models) [41].

4 Conclusion

The COVID-19 epidemic had a great impact on the development of information technologies in medicine and showed the importance of a developed information infrastructure. The emergence of models that provide an acceptable level of accuracy will make it possible to build

a large number of innovative systems: diagnostic systems, decision systems or decision support systems. In addition, it will be possible to centrally process the results of medical diagnostics, without being tied to the specific medical facilities.

In the course of the work, the features of the analysis of cellular images were identified, the data were augmented and divided into training and test sets, taking into account the distribution of subjects, experiments were carried out and model's outputs were interpreted. It was found that the best results on the given dataset were shown by the EfficientNetV2M architecture (the value of the target metric F_1 is 6.8% higher than the average

among the architectures under consideration, while the difference between the worst and best value of the F_1 metric is 0.208).

Acknowledgements

This work was funded by the government project of the NRC "Kurchatov Institute".

Disclosures

The authors declare no conflict of interest.

References

1. I. Milevsky, Acute leukemias – history, causes, MedUniver, 18 March 2021 (accessed 23 May 2022). [https://meduniver.com/Medical/gematologia/ostrie_leikozi.html].
2. Overview Acute lymphoblastic leukaemia, UK (accessed 15 September 2022). [<https://www.nhs.uk/conditions/acute-lymphoblastic-leukaemia/>].
3. W. Ladines-Castro, G. Barragan-Ibanez, M. Luna-Perez, A. Santoyo-Sánchez, J. Collazo-Jaloma, E. Mendoza-García, and C.O. Ramos-Peñafiel, "Morphology of leukaemias," *Revista Médica del Hospital General de México* 79(2), 107–113 (2018).
4. Y. Gu, A. Chen, X. Zhang, C. Fan, K. Li, and J. Shen, "Deep Learning based Cell Classification in Imaging FlowCytometer," *ASP Transactions on Pattern Recognition and Intelligent Systems* 1(2), 18–27 (2021).
5. S. Shafique, S. Tehsin, "Acute lymphoblastic leukemia detection and classification of its subtypes using pretrained deep convolutional neural networks," *Technology in Cancer Research & Treatment* 17, 153303381880278 (2018).
6. T. Pansombut, S. Wikaisuksakul, K. Khongkraphan, and A. Phon-On "Convolutional Neural Networks for Recognition of Lymphoblast Cell Images," *Computational Intelligence and Neuroscience* 2019, 7519603 (2019).
7. R. Cuocolo, M. Caruso, T. Perillo, L. Ugga, and M. Petretta, "Machine learning in oncology: a clinical appraisal," *Cancer Letters* 481, 55–62 (2020).
8. M. Khushi, K. Shaikat, T. M. Alam, I. A. Hameed, S. Uddin, S. Luo, X. Yang, and M. C. Reyes, "A comparative performance analysis of data resampling methods on imbalance medical data," *IEEE Access* 9, 109960–109975 (2021).
9. C. Lu, M. Chen, J. Shen, L. L. Wang, and C. C. Hsu "Removal of salt-and-pepper noise for X-ray bio-images using pixel-variation gain factors," *Computers & Electrical Engineering* 71, 862–876 (2018).
10. N. J. Everall, "Confocal Raman microscopy: common errors and artefacts," *Analyst* 135(10), 2512–2522 (2010).
11. J. Ludzik, C. Lee, A. Witkowski, and J. Hetts, "Minimizing sampling error using a unique ink-stained reflectance confocal microscopy-guided biopsy technique to diagnose a large lentigo maligna," *JAAD Case Reports* 24, 118–120 (2022).
12. W. Yu, W. K. Dodds, M. K. Banks, J. Skalsky, and E. A. Strauss, "Optimal staining and sample storage time for direct microscopic enumeration of total and active bacteria in soil with two fluorescent dyes," *Applied and Environmental Microbiology* 61(9), 3367–3372 (1995).
13. R. Laine, G. Jacquemet, and A. Krull, "Imaging in focus: An introduction to denoising bioimages in the era of deep learning," *The International Journal of Biochemistry & Cell Biology* 140, 106077 (2021).
14. T. Parag, A. Chakraborty, S. Plaza, and L. Scheffer, "A context-aware delayed agglomeration framework for electron microscopy segmentation," *PloS One* 10(5), e0125825 (2015).
15. S. U. Akram, J. Kannala, L. Eklund, and J. Heikkilä, "Cell segmentation proposal network for microscopy image analysis," in *Deep Learning and Data Labeling for Medical Applications*, G. Carneiro, D. Mateus, L. Peter, A. Bradley, J. M. R. S. Tavares, V. Belagiannis, J. P. Papa, J. C. Nascimento, M. Loog, Z. Lu, J. S. Cardoso, and J. Cornebise (Eds.), Springer International Publishing, 10008, 21–29 (2016).
16. A. Venkataramanan, M. Laviale, C. Figus, P. Usseglio-Polatera, and C. Pradalier, "Tackling inter-class similarity and intra-class variance for microscopic image-based classification," in *Computer Vision Systems*, M. Vincze, T. Patten, H. I. Christensen, L. Nalpantidis, and M. Liu (Eds.), Springer International Publishing, 12899, 93–103. (2021).
17. H. Rizk, A. Shokry, and M. Youssef, "Effectiveness of data augmentation in cellular-based localization using deep learning," in *IEEE Wireless Communications and Networking Conference*, 1–6 (2019).

18. S. Yu, S. Zhang, B. Wang, H. Dun, L. Xu, X. Huang, and X. Feng, “[Generative adversarial network based data augmentation to improve cervical cell classification model](#),” *Mathematical Biosciences and Engineering* 18, 1740–1752 (2021).
19. M. Momeny, A. M. Latif, M. A. Sarram, R. Sheikhpour, and Y. D. Zhang “[A noise robust convolutional neural network for image classification](#),” *Results in Engineering* 10, 100225 (2021).
20. “Leukemia Classification,” US (accessed 16 February 2024) [<https://www.kaggle.com/datasets/andrewmvd/leukemia-classification>].
21. M. Tan, Q. V. Le, “[EfficientNet: rethinking model scaling for convolutional neural networks](#),” *International Conference on Machine Learning* 10, 6105–6114 (2019).
22. G. Kyriakides, K. Margaritis, “[An introduction to neural architecture search for convolutional networks](#),” *arXiv preprint arXiv:2005.11074v1* (2020).
23. X. Yin, W. Chen, X. Wu, and H. Yue, “[Fine-tuning and visualization of convolutional neural networks](#),” *12th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, 1310–1315 (2017).
24. G. Chelnlei, L. Zhang, “[A Novel Multiresolution Spatiotemporal Saliency Detection Model and Its Applications in Image and Video Compression](#),” *IEEE Transactions on Image Processing* 19(1), 185–198 (2010).
25. R. Monroy, S. Lutz, T. Chalasani, and V. Smolic, “[Salnet360: Saliency maps for omni-directional images with CNN](#),” *Signal Processing: Image Communication* 69, 26–34 (2018).
26. K. Nagasubramanian, S. Jones, A. K. Singh, B. Ganapathysubramanian, and S. Sarkar “[Explaining hyperspectral imaging based plant disease identification: 3D CNN and saliency maps](#),” *arXiv preprint arXiv:1804.08831* (2018).
27. A. Alqaraawi, M. Schuessler, P. Weiß, E. Costanza, and N. Berthouze, “[Evaluating saliency map explanations for convolutional neural networks: a user study](#),” *Proceedings of the 25th International Conference on Intelligent User Interfaces*, 275–285 (2020).
28. A. Rehman, N. Abbas, T. Saba, S. I. Rahman, Z. Mehmood, and H. Kolivand, “[Classification of acute lymphoblastic leukemia using deep learning](#),” *Microscopy Research and Technique* 81(11), 1310–1317 (2018).
29. A. Genovese, M. S. Hosseini, V. Piuri, K. N. Plataniotis, and F. Scotti, “[Acute lymphoblastic leukemia detection based on adaptive unsharping and deep learning](#),” in *ICASSP 2021–2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1205–1209 (2021).
30. P. Chlap, H. Min, N. Vandenberg, J. Dowling, L. Holloway, and A. Haworth, “[A review of medical image data augmentation techniques for deep learning applications](#),” *Journal of Medical Imaging and Radiation Oncology* 65(5), 545–563 (2021).
31. F. Garcea, A. Serra, F. Lamberti, and L. Morra, “[Data augmentation for medical imaging: A systematic literature review](#),” *Computers in Biology and Medicine* 152, 106391 (2023).
32. L. H. S. Vogado, R. M. S. Veras, F. H. D. Araujo, R. R.V. Silva, and K. R.T. Aires, “[Leukemia diagnosis in blood slides using transfer learning in CNNs and SVM for classification](#),” *Engineering Applications of Artificial Intelligence* 72, 415–422 (2018).
33. D. Picard, “[Torch. manual_seed \(3407\) is all you need: On the influence of random seeds in deep learning architectures for computer vision](#),” *arXiv preprint arXiv:2109.08203* (2021).
34. P. Madhyastha, R. Jain, “[On model stability as a function of random seed](#),” *arXiv preprint arXiv:1909.10447* (2019).
35. J. Luengo, S. García, and F. Herrera, “[A study on the use of statistical tests for experimentation with neural networks: Analysis of parametric test conditions and non-parametric tests](#),” *Expert Systems with Applications* 36(4), 7798–7808 (2009).
36. N. Arun, N. Gaw, P. Singh, K. Chang, M. Aggarwal, B. Chen, K. Hoebel, S. Gupta, J. Patel, M. Gidwani, J. Adebayo, M. D. Li, and J. Kalpathy-Cramer, “[Assessing the trustworthiness of saliency maps for localizing abnormalities in medical imaging](#),” *Radiology: Artificial Intelligence* 3(6), e200267 (2021).
37. N. T. Arun, N. Gaw, P. Singh, K. Chang, K. V. Hoebel, J. Patel, M. Gidwani, and J. Kalpathy-Cramer, “[Assessing the validity of saliency maps for abnormality localization in medical imaging](#),” *arXiv preprint arXiv:2006.00063* (2020).
38. M. Aggarwal, N. Arun, S. Gupta, A. Vaswani, B. Chen, M. Li, K. Chang, J. Patel, K. Hoebel, M. Gidwani, J. Kalpathy-Cramer, and P. Singh, “[Towards trainable saliency maps in medical imaging](#),” *arXiv preprint arXiv:2011.07482* (2020).
39. C. Andreou, D. Rogge, and R. Müller, “[A New Approach for Endmember Extraction and Clustering Addressing Inter- and Intra-Class Variability via Multiscaled-Band Partitioning](#),” in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9(9), 4215–4231 (2016).
40. D. Lee, K. Kim, “[Improved Noise-filtering Algorithm for AdaBoost Using the Inter-and Intra-class Variability of Imbalanced Datasets](#),” *Journal of Intelligent & Fuzzy Systems* 43(4), 5035–5051 (2022).
41. Y. Liu, Y. Zhang, Y. Wang, F. Hou, J. Yuan, J. Tian, Y. Zhang, Z. Shi, J. Fan, and Z. He, “[A survey of visual transformers](#),” *IEEE Transactions on Neural Networks and Learning Systems* (2023).