Neural Cellular Automata for Lightweight, Robust and Explainable Classification of White Blood Cell Images

Michael Deutges[1,2⋆], Ario Sadafi[1,3⋆], Nassir Navab[3,4], and Carsten Marr[1,5]

[1] Institute of AI for Health, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany
[2] Faculty of Mathematics, Technical University Munich, Munich, Germany
[3] Computer Aided Medical Procedures, Technical University of Munich, Munich, Germany
[4] Computer Aided Medical Procedures, Johns Hopkins University, USA
[5] Helmholtz AI, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany

Abstract. Diagnosis of hematological malignancies depends on accurate identification of white blood cells in peripheral blood smears. Deep learning techniques are emerging as a viable solution to scale and optimize this process by automatic cell classification. However, these techniques face several challenges such as limited generalizability, sensitivity to domain shifts, and lack of explainability. Here, we introduce a novel approach for white blood cell classification based on neural cellular automata (NCA). We test our approach on three datasets of white blood cell images and show that we achieve competitive performance compared to conventional methods. Our NCA-based method is significantly smaller in terms of parameters and exhibits robustness to domain shifts. Furthermore, the architecture is inherently explainable, providing insights into the decision process for each classification, which helps to understand and validate model predictions. Our results demonstrate that NCA can be used for image classification, and that they address key challenges of conventional methods, indicating a high potential for applicability in clinical practice.

Keywords: Neural Cellular Automata · Explainability · Single-Cell Classification · Domain Generalization

1 Introduction

The diagnosis of hematological disorders heavily relies on the microscopic examination of blood cells in laboratory settings. Among the challenges that cytologists encounter is the identification of relevant white blood cells under the microscope for diagnosing leukemia, which ranks among the most lethal hematological disorders. Certain types of leukemia, such as acute promyelocytic leukemia (APL),

[⋆] equal contribution

pose significant challenges as they demand urgent attention due to the potential for life-threatening bleeding or coagulation [5].

Identifying relevant cells, known as blast cells, under the microscope is an essential initial step in diagnosing leukemia subtypes, including APL. This task is time consuming and burdensome. Recent advancements in deep learning methods have introduced automatic tools for cytologists, enabling them to expedite the process of locating and classifying these critical cells. As a result, the diagnosis can become faster and more robust, streamlining the workflow in leukemia laboratories. Matek et al. [10,11] suggested a highly accurate classifier for bone marrow cells and cells from peripheral blood, demonstrating performance comparable with trained cytologists. Eckardt et al. proposed deep learning methods for diagnosis of acute myeloid leukemia (AML) subtypes [2,3]. In a similar work, Sidhom et al. focused on diagnosis of APL from other subtypes of AML [18].

These methods are typically trained and validated on datasets gathered from a single source, rendering them susceptible to domain shifts [6]. To enhance cross-domain performance, various domain adaptation methods have been proposed. These include extracting domain-invariant features [17] or employing continual learning techniques [16], which facilitate regular model updates to address challenges introduced by new domains.

Explaining the decisions of convolutional neural networks (CNNs) presents a significant challenge in the applicability of deep learning methods and is crucial for computer aided diagnosis systems [7,15]. This explanation is typically achieved through pixel or feature attributions, where certain parts of the image are identified as important by the explanation method [15]. The interpretation of these clues relies on the observer, who examines the explanations to form an understanding of the system's inner workings.

Neural cellular automata (NCA) are gradually emerging as a lightweight, robust and input-invariant solution for image generation or segmentation tasks. Growing NCA are able to generate, maintain and regenerate complex shapes [13]. In the medical domain, Kalkhof et al. [9] have proposed Med-NCA for segmentation of different organs in T1-weighted hippocampus and T2-MRI datasets. It demonstrates comparable performance with nnUNet [8], while having a fraction of its parameters.

In this paper, we propose a method for explainable and robust single white blood cell classification utilizing a NCA backbone. Features from the single-cell images are extracted by the NCA and used for classification with a multi-layer perceptron. Our approach inherently offers explainability and demonstrates robustness against domain shifts when evaluated on three datasets collected from different medical centers, each with its specific laboratory procedures and staining techniques.

To the best of our knowledge, this is the first application of NCA for image classification. Our contribution is thus the methodological advancement of a new high-potential method and its application to a challenging biomedical classification task. To foster reproducible research we are providing our source code at https://github.com/marrlab/WBC-NCA.

2 Methods

Conventional deep learning models for image classification involve two steps: feature extraction and classification. In our approach, we extract the features with the use of NCA. Unlike traditional methods that rely on a series of convolutions, activation functions, and pooling operations, NCA apply a local update rule iteratively over a fixed number of steps, allowing each cell to aggregate information from a wider context. This local architecture of NCA, which update cells based on their immediate surroundings, ensures lightweight storage and fast inference without compromising performance.

2.1 Neural Cellular Automata Architecture

A NCA can be defined as a tuple

$$\text{NCA} = \langle S, f \rangle \tag{1}$$

where $S \in \mathbf{R}^{64\times 64\times n}$ denotes the seed and $f : \mathbf{R}^{3\times 3\times n} \to \mathbf{R}^n$ is the transition function, which is applied iteratively starting from S. The dimensions of the seed imply the number of individual cells $c \in \mathbf{R}^n$ which, arranged in a 64 by 64 grid, comprise the domain our NCA operates on. We refer to each 64×64 slice of the domain as a channel. In our case, S consists of the RGB image $\in \mathbf{R}^{64\times 64\times 3}$ and $n - 3$ channels of zeros.

The transition function updates each cell c based on its 3×3 neighborhood denoted by N_c and can be divided into two parts: perception f_p and update f_u.

$$f : \mathbf{R}^{3\times 3\times n} \to \mathbf{R}^n, \qquad N_c \mapsto f_u(f_p(N_c)) \tag{2}$$

$$f_p : \mathbf{R}^{3\times 3\times n} \to \mathbf{R}^{3n}, \qquad N_c \mapsto (c, N_c * k_1, N_c * k_2)^T \tag{3}$$

$$f_u : \mathbf{R}^{3n} \to \mathbf{R}^n, \qquad f_p(N_c) \mapsto W_2 \max(W_1 f_p(N_c) + b_1, 0) + b_2 \tag{4}$$

We call $f_p(N_c)$ the perception vector. It consists of two channel-wise convolutions with kernels k_1 and k_2 and the identity of the cell c. The update function f_u is comprised of two fully connected layers, parameterized by W_1, W_2, b_1, and b_2, with a Rectified Linear Unit (ReLU) activation in between. The convolution kernels k_1 and k_2 and the weights W_1, W_2 and biases b_1, b_2 of the linear layers are trainable parameters.

The transition function defines the cell update at time step t according to

$$c^{t+1} = c^t + \delta f(N_{c^t}) \tag{5}$$

where δ is randomly set to 0 or 1 for each cell in each step, i.e. only approximately 50% of cells get updated in each step. This stochastic activation acts as regularization, enhancing the model's robustness and improving generalization. The architecture and NCA step including perception and cell update is illustrated in Figure 1.

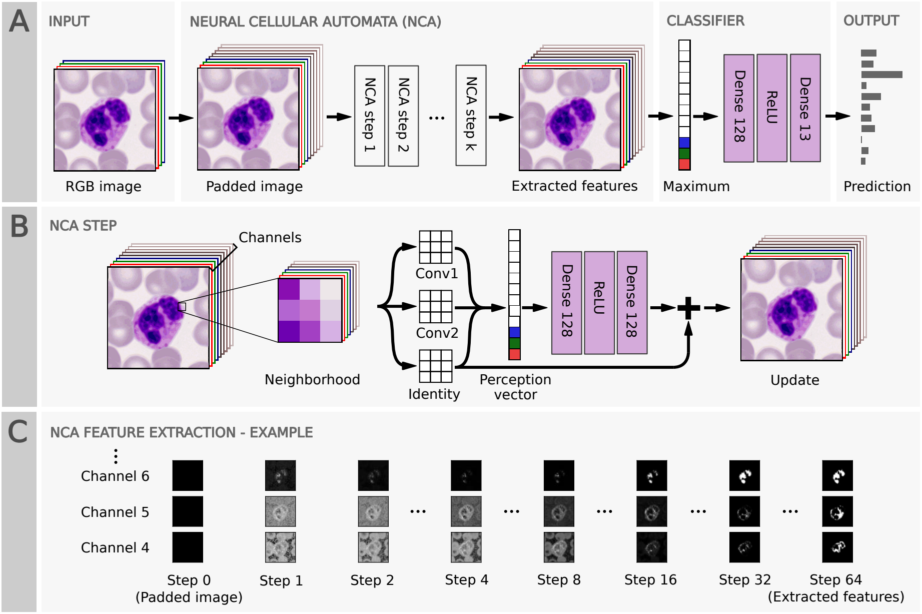

Fig. 1. Neural cellular automata (NCA) can be used for the accurate classification of single white blood cells in patient blood smears. **A:** Our approach consists of four steps: i) image padding to increase the number of channels, ii) k NCA update steps to extract features from the image that manifest in the hidden channels, iii) pooling via channel-wise maximum, and iv) a fully connected network to classify the image. **B:** The NCA step updates each cell based on its immediate surroundings according to equations 2 - 5. **C:** Training the model end-to-end allows the NCA to learn an update rule that extracts useful features.

2.2 NCA for Image Classification

To use NCA for a classification task we perform k update steps, which results in extracted features that form in the additional channels (Figure 1C). Each channel is condensed via taking the maximum, which yields a rich feature vector denoted by v. We arrive at the prediction by feeding this vector into a two layer fully connected network denoted as the classifier g.

$$g : \mathbf{R}^n \to [0,1]^{13}, \qquad g(v) \mapsto \sigma(W_4 \max(W_3 v + b_3, 0) + b_4) \tag{6}$$

with $\sigma(z) = \frac{1}{1+e^{-z}}$ and weights and biases W_3, W_4, b_3, b_4. The output $g(v)$ corresponds to the predictions of the model for the 13 different cell types in our datasets.

In summary, our model is structured into four steps (Figure 1):

1. **Image Padding:** We pad the image to the desired number of channels. Additional channels serve as memory, which enables the NCA to learn more complex patterns.

2. **NCA Update Steps:** We perform k NCA update steps to extract features which manifest in the hidden channels.
3. **Feature Aggregation:** Channel-wise maximum is taken to condense the extracted features into a vector.
4. **Classification:** This condensed vector is fed into a two layer fully connected neural network which yields the class predictions.

During training, we apply the cross-entropy loss function to the class prediction and the corresponding ground truth, backpropagating through both the classifier network and the NCA over time. This end-to-end training scheme enables the NCA to learn an update rule that, when applied for k steps, extracts features from the image which are effective for classification.

2.3 Explainability via Layer-wise Relevance Propagation

To gain insights into the model's decision making, we use layer-wise relevance propagation [12] on the fully connected layers of the classifier network to attribute relevance to each feature extracted by the NCA. We apply the epsilon rule described in [12]. Relevance values are propagated through each layer according to

$$R_j^{L-1} = \sum_k \frac{a_j w_{jk}}{\epsilon + \sum_j a_j w_{jk}} R_k^L, \tag{7}$$

where ϵ is a parameter to control sparsity of the explanations, a_j are the activations of the respective neurons and w_{jk} corresponds to the weights in the layer. If $\epsilon = 0$, this definition fulfills $\sum_j R_j^L = p(x)$ for every layer L, where $p(x)$ is the prediction of the model. Relevance values R_j correspond to the contribution of feature j to the prediction of the specified class.

3 Experiments and Results

3.1 Datasets

We use three different datasets to evaluate our method:

- **Matek-19** is a collection of over 18,000 annotated white blood cells from 200 individuals where half of the subjects are affected by AML. The data has been collected at the Munich university hospital and is publicly available [11]. It consists of 15 classes, and the images have a size of 400×400 pixels, which corresponds to 29×29 micrometers.
- **INT-20** is an in-house dataset and consists of around 42,000 images from 18 different classes with a resolution of 288×288 pixels or 25×25 micrometers.
- **Acevedo-20** is a publicly available dataset of around 17,000 single-cell images from healthy individuals collected at the Hospital Clinic of Barcelona [1]. The images are categorized in 8 different classes and have a resolution of 360×363 pixels or 36×36.3 micrometers.

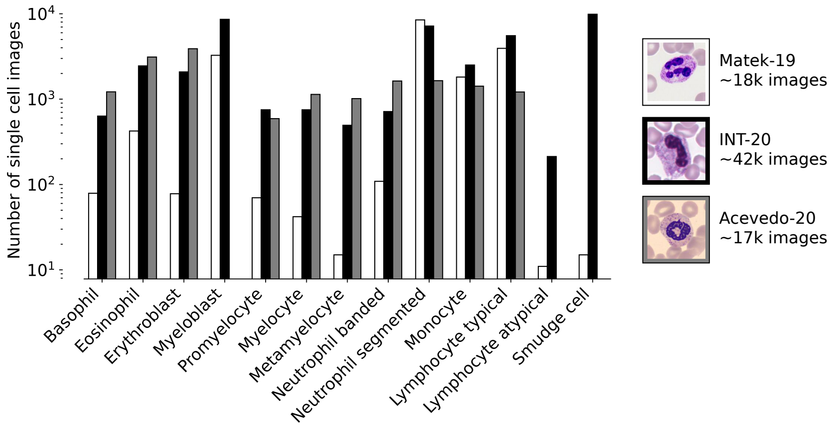

Fig. 2. Class distribution for the three datasets.

The class definitions of these three datasets are harmonized by a medical expert resulting in 13 commonly defined classes. Figure 2 shows the distribution of cell types for the three datasets.

We resampled the images to a resolution of 64×64 as a high resolution would significantly increase time and RAM requirements for training the NCA.

3.2 Implementation details

Model Choice The standard configuration of our NCA model has $n = 128$ channels. A large number of channels provides cells with a bigger memory capacity and allows the NCA to learn a more advanced update rule. We tested

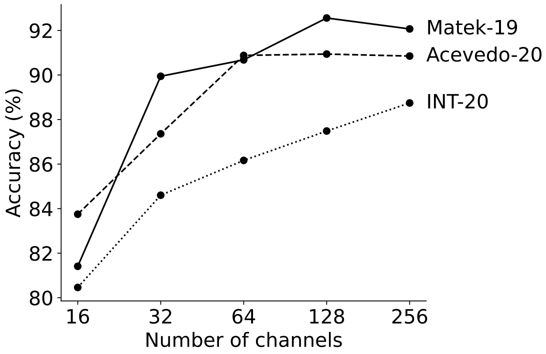

Fig. 3. NCA classification accuracy saturates at 128 channels for two out of three white blood cell datasets.

the classification accuracy on the three datasets for models with different numbers of channels to validate this choice (Figure 3). The fully connected layers in the transition function have a hidden size of 128, which is consistent with other NCA models [9,13]. A total of $k = 64$ steps are performed before the features are aggregated and propagated through the classifier network. We found that the number of steps seems to correlate positively with the accuracy. We opted for 64 steps as a compromise between performance and training time and RAM requirements, respectively.

Training We used an Adam optimizer with a learning rate of 0.0004 and a β_1 of 0.9 and β_2 of 0.999 together with an exponential learning rate decay with weight 0.9999. To treat the problems arising from an imbalanced dataset, we used random over- and undersampling together with random rotation and flip augmentations of the training images to have a uniform distribution of classes in the training process. Each model was trained with a batch size of 16 for 32 epochs, after which the validation loss did not change significantly anymore. In general, we observed our model to be notably less prone to overfitting compared to conventional methods. This might be a consequence of the small number of parameters which sums up to just about 86k.

3.3 Results

We compare the performance in terms of accuracy on the three datasets against two baselines. Matek et. al [11] proposed a method based on a ResNeXt [20] architecture for the classification of the Matek-19 dataset. The model binaries are available, which we used for testing on the three datasets. In order to compare its performance when trained on the other datasets, we are employing the same ResNeXt architecture to train and test. Additionally, we are comparing against

Table 1. When trained on data from one domain, and tested on data from a different domain, our NCA model outperforms other baselines, while only having a fraction of the parameters. The accuracy of our NCA method is reported against the baselines. Mean and standard deviation are computed from five independent runs.

		NCA	ResNeXt	AE-CFE
Trained on	Tested on	$\sim$86k pa.	$\sim$25M pa.	$\sim$3.9M pa.
Matek-19	Matek-19	92.6±0.6	**96.1**	83.7±0.5
	Acevedo-20	**43.9±1.6**	8.1	21.9±0.4
	INT-20	24.3±5.1	29.5	**48.4±0.2**
Acevedo-20	Matek-19	32.2±9.1	7.3±3.1	**45.1±0.5**
	Acevedo-20	**90.6±0.2**	85.7±2.4	65.2±0.5
	INT-20	13.6±2.8	8.1±1.4	**21.0±0.5**
INT-20	Matek-19	50.0±11.1	49.0±6.3	**73.2±0.1**
	Acevedo-20	**46.9±4.7**	16.9±1.6	31.8±0.4
	INT-20	88.0±0.3	**88.7±1.5**	65.6±0.5

AE-CFE proposed by Salehi et al. [17], which uses an autoencoder trained on all three datasets to extract domain invariant features, which are later used in a random forest to classify single white blood cell images.

Our model outperforms the ResNext architecture in 6 out of 9 experiments. Furthermore, our model outperforms AE-CFE in all experiments where the training and testing set come from the same center. The advantage of AE-CFE in the other experiments is a result of the pretraining on all three domains.

3.4 Explainability

Our NCA model offers explainability by design. Features are extracted in the channels and offer insights into the decision process. A visualization of the channels allows for easy identification of biases, e.g. if the model were to make decisions based on activations that are located not within the region of the cell but in the background. An example is shown in Figure 4. Additionally, the unique architecture of the NCA yields features which are more interpretable by humans. The extraction via local cell updates aggregated over time results in features which tend to be more connected regions of the image. In our scenario, these regions could include the entire cell, its nuclei, cytoplasm parts, or more subtle sub-cellular structures.

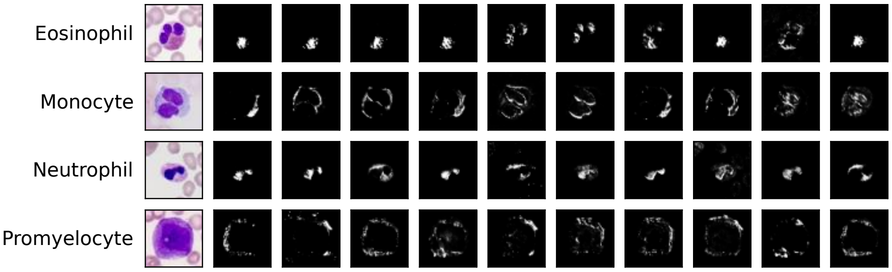

Fig. 4. Example of ten features (columns) extracted by the NCA for an exemplary eosinophil, monocyte, neutrophil, and promyelocyte. The highest relevance scores have been attributed by layer-wise relevance propagation on the fully connected layers of the classifier network. The channel activations are located within the regions of the white blood cell indicating the model's ability to correctly identify and analyze pertinent aspects of the input data.

4 Conclusion

In this study, we demonstrated the potential of neural cellular automata as a tool for image classification, addressing key challenges faced by conventional methods in clinical practice. By providing a combination of performance, robustness, and explainability, our approach provides an alternative to improved diagnostic tools in the field of hematological diseases [2,3,7,10,11,18].

The inherent explainability of NCA simplifies the process of interpreting the inner workings for observers, offering a distinct advantage compared to the noisy pixel attributions typically used for explaining CNNs [15].

Lastly, the lightweight architecture allows for application in less developed and remote areas for various diagnostic tasks without requiring access to high-end hardware to run large models.

Acknowledgements. We would like to thank John Kalkhof for the inspiring discussions. C.M. acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant Agreement No. 866411 & 101113551) and support from the Hightech Agenda Bayern.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Acevedo, A., Merino, A., Alférez, S., Molina, Á., Boldú, L., Rodellar, J.: A dataset of microscopic peripheral blood cell images for development of automatic recognition systems. Data in brief **30** (2020)
2. Eckardt, J.N., Middeke, J.M., Riechert, S., Schmittmann, T., Sulaiman, A.S., Kramer, M., Sockel, K., Kroschinsky, F., Schuler, U., Schetelig, J., et al.: Deep learning detects acute myeloid leukemia and predicts npm1 mutation status from bone marrow smears. Leukemia **36**(1), 111–118 (2022)
3. Eckardt, J.N., Schmittmann, T., Riechert, S., Kramer, M., Sulaiman, A.S., Sockel, K., Kroschinsky, F., Schetelig, J., Wagenführ, L., Schuler, U., et al.: Deep learning identifies acute promyelocytic leukemia in bone marrow smears. BMC cancer **22**(1), 201 (2022)
4. Florindo, J.B., Metze, K.: A cellular automata approach to local patterns for texture recognition. Expert Systems with Applications **179**, 115027 (2021)
5. Gill, H., Yung, Y., Chu, H.T., Au, W.Y., Yip, P.K., Lee, E., Yim, R., Lee, P., Cheuk, D., Ha, S.Y., et al.: Characteristics and predictors of early hospital deaths in newly diagnosed apl: a 13-year population-wide study. Blood Advances **5**(14), 2829–2838 (2021)
6. Guan, H., Liu, M.: Domain adaptation for medical image analysis: a survey. IEEE Transactions on Biomedical Engineering **69**(3), 1173–1185 (2021)
7. Hehr, M., Sadafi, A., Matek, C., Lienemann, P., Pohlkamp, C., Haferlach, T., Spiekermann, K., Marr, C.: Explainable ai identifies diagnostic cells of genetic aml subtypes. PLOS Digital Health **2**(3), e0000187 (2023)
8. Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P.F., Kohl, S., Wasserthal, J., Koehler, G., Norajitra, T., Wirkert, S., et al.: nnu-net: Self-adapting framework for u-net-based medical image segmentation. arXiv preprint arXiv:1809.10486 (2018)

9. Kalkhof, J., González, C., Mukhopadhyay, A.: Med-nca: Robust and lightweight segmentation with neural cellular automata. In: International Conference on Information Processing in Medical Imaging. pp. 705–716. Springer (2023)
10. Matek, C., Krappe, S., Münzenmayer, C., Haferlach, T., Marr, C.: Highly accurate differentiation of bone marrow cell morphologies using deep neural networks on a large image data set. Blood, The Journal of the American Society of Hematology **138**(20), 1917–1927 (2021)
11. Matek, C., Schwarz, S., Spiekermann, K., Marr, C.: Human-level recognition of blast cells in acute myeloid leukaemia with convolutional neural networks. Nature Machine Intelligence **1**(11), 538–544 (2019)
12. Montavon, G., Binder, A., Lapuschkin, S., Samek, W., Müller, K.R.: Layer-Wise Relevance Propagation: An Overview, pp. 193–209. Springer International Publishing, Cham (2019)
13. Mordvintsev, A., Randazzo, E., Fouts, C.: Growing isotropic neural cellular automata. In: Artificial Life Conference Proceedings 34. vol. 2022, p. 65. MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info ... (2022)
14. Randazzo, E., Mordvintsev, A., Niklasson, E., Levin, M., Greydanus, S.: Self-classifying mnist digits. Distill **5**(8), e00027–002 (2020)
15. Sadafi, A., Adonkina, O., Khakzar, A., Lienemann, P., Hehr, R.M., Rueckert, D., Navab, N., Marr, C.: Pixel-level explanation of multiple instance learning models in biomedical single cell images. In: International Conference on Information Processing in Medical Imaging. pp. 170–182. Springer (2023)
16. Sadafi, A., Salehi, R., Gruber, A., Boushehri, S.S., Giehr, P., Navab, N., Marr, C.: A continual learning approach for cross-domain white blood cell classification. In: MICCAI Workshop on Domain Adaptation and Representation Transfer. pp. 136–146. Springer (2023)
17. Salehi, R., Sadafi, A., Gruber, A., Lienemann, P., Navab, N., Albarqouni, S., Marr, C.: Unsupervised cross-domain feature extraction for single blood cell image classification. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 739–748. Springer (2022)
18. Sidhom, J.W., Siddarthan, I.J., Lai, B.S., Luo, A., Hambley, B.C., Bynum, J., Duffield, A.S., Streiff, M.B., Moliterno, A.R., Imus, P., et al.: Deep learning for diagnosis of acute promyelocytic leukemia via recognition of genomically imprinted morphologic features. NPJ precision oncology **5**(1), 38 (2021)
19. Tesfaldet, M., Nowrouzezahrai, D., Pal, C.: Attention-based neural cellular automata. Advances in Neural Information Processing Systems **35**, 8174–8186 (2022)
20. Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K.: Aggregated residual transformations for deep neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1492–1500 (2017)
21. Yeşil, Ç., Korkmaz, E.E.: A novel cellular automata-based approach for generating convolutional filters. Machine Vision and Applications **34**(3), 38 (2023)