Journal of Electronics, Electromedical Engineering, and Medical Informatics
Multidisciplinary: Rapid Review: Open Access Journal **Vol. 6, No. 1, January 2024, pp: 84-91; eISSN: 2656-8632**

RESEARCH ARTICLE **OPEN ACCESS**

Manuscript received January 10, 2024; revised January 26, 2024; accepted January 26, 2024; date of publication January 30, 2024
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeemi.v6i1.354

How to cite: Ahmad Badruzzaman and Aniati Murni Arymurthy, A Comparative Study of Convolutional Neural Network in Detecting Blast Cells for Diagnose Acute Myeloid Leukemia, Journal of Electronics, Electromedical Engineering, and Medical Informatics, vol. 6, no. 1, pp. 84-91, January 2024.

A Comparative Study of Convolutional Neural Network in Detecting Blast Cells for Diagnose Acute Myeloid Leukemia

Ahmad Badruzzaman and Aniati Murni Arymurthy
Faculty of Computer Science, Universitas Indonesia, Depok, Indonesia

Corresponding author: Ahmad Badruzzaman (e-mail: ahmad.badruzzaman@ui.ac.id).

ABSTRACT Understanding blood plays a crucial role in obtaining information for monitoring health conditions and diagnosis of hematologic diseases such as acute myeloid leukemia. It is characterized by irregular expansion of immature white blood cells called blast cells in the blood and bone marrow. To diagnose acute myeloid leukemia, a sample of bone marrow is necessary to be examined under a microscope through bone marrow examination. As for minimizing human subjectivity and automating medical screening, this study performed image classification for detecting blast cells in leukocytes from microscopic images. The main objective of this study is to examine an established convolution neural network structure in detecting blast cells. We compared well-known architecture such as ResNet, ResNeXt, and EfficientNetV2. The model's performance assessment was done by two evaluation levels which are at a macro level and per class level. The experiment results show ResNet architecture with 18 layers (ResNet 18) outperforms the remaining models at both levels. Furthermore, as the architecture utilizes residual learning, ResNet and ResNeXt models converge faster than EfficientNetV2 at the training phase. In addition, ResNet architecture with 50 layers (ResNet 50) outperforms the remaining models specifically at blast cell identification in case of medical screening. However, EfficientNetV2 shows a promising potential at a macro level to classify leukocytes in general while maintaining a competitive performance to ResNet and ResNeXt in the same numbers of parameter. Therefore, this study concludes that residual learning shows an outstanding performance in a few numbers of iteration. In addition, a model with shallow layer is the best model for classifying leukocyte in general and a model with deeper layer is the best model for detecting blast cells in leukocyte specifically.

INDEX TERMS acute myeloid leukemia, blast cells, convolution neural network, image classification

I. INTRODUCTION

In medical investigation, blood cells are observed under microscopic examination [1]. Within certain conditions, such as diagnosing leukemia, blood cells are also observed through bone marrow examination. Leukemia is a hematologic disease where immature leukocyte called blast cells proliferate irregularly, filling up the bone marrow, and preventing the production of erythrocytes and thrombocytes [2], [3]. Depending on cell proliferation rate, leukemia can be categorized as acute leukemia (more aggressive) and chronic leukemia (slow growing). In addition, affected leukocyte cells also classify leukemia as myeloid leukemia (myeloid cells) and lymphocytic leukemia (lymphocyte cells) [4], [5]. Acute myeloid leukemia refers to an aggressive leukemia which affected myeloid cells such as myeloblast cells which are immature precursors of leukocyte that will mature into granulocyte and monocyte [6] - [8]. As a clonal and malignant disease of hematopoietic tissues, acute myeloid leukemia characterized by accumulation of irregular blast cells (mainly in the bone marrow) and impaired production of normal blood cells [7], [8]. According to the World Health Organization (WHO), a criterion for indicating an acute myeloid leukemia patient is 20% or more of leukemic blast cells either in blood peripheral or bone marrow [9], [10]. In general, this hematologic disease commonly occurs in elderly people with symptoms including anemia, fever, ulcer of mucous

membranes, and granulocytic insufficiency. In fact, a study of leukemia incidence trends on a global scale conducted by Dong et al. (2020) [11] showed that the incident rate of acute myeloid leukemia with age was increased exponentially. Furthermore, the 5-years survival rate of acute myeloid leukemia in younger patients is between 40% - 50%, while in older patients (age > 60) between 20% - 30% [12], [13]. Therefore, the diagnosis of acute myeloid leukemia plays a crucial part in order to give a treatment as soon as possible.

When a diagnosis of acute myeloid leukemia is suspected, bone marrow examination is conducted to observe a sample of myeloid blast cells in bone marrow [7]. However, blood cells observation is labor intensive and time consuming that leads to inefficiency [1]. Furthermore, subjectivity of physicians and experts sometimes leads to discrepancies of diagnosis [14], [15]. In that case, machine learning algorithm becomes a logical option for solving those problems as machine learning can learn a pattern from observational data by itself during training phase while adapt without explicit instruction [16]. In addition, a machine learning based on artificial neural network algorithm which mimicking human brain and encouraging pattern recognition to extracting high levels of features of data [17], [18].

Automated blood cells classification using machine learning algorithms has been conducted for several years. Generally, blood marrow images were classified to distinguish blood cells. In case of leukemia, lymphoblastic leukemia has received more attention than myeloid case due to its less diversity of cytomorphology [19]. As for myeloid leukemia case, the early study that utilize deep learning in case of acute myeloid leukemia were conducted using ResNeXt architecture [20], [21]. Those studies claim the dataset used on their respective study were the largest expert-annotated single blood cell images dataset prior to publication time. There are also studies focused on further diagnosing which classify leukemic cells subtypes using support vector machine [22], [23]. As deep learning requires a large number of data while at some blood cells there are limited numbers of images, study for generating synthetic blood cells image progressed to enhance performance of deep learning model on specific blood class scheme [24] - [26]. While generating synthetic image possibly enhance deep learning performance, it also led to another problem such as bias to some class due to imbalance dataset problem and unsolved the limited data problem due to the large number of samples that required for learn to generating synthetic images. An alternative method through transfer learning also taken into consideration when blood cells images are considered low [27] - [29]. However, the limited number of myeloid leukemia cytomorphology dataset requires another task domain to utilize transfer learning which is sometimes the selected learned domain lacking justification. In addition, those previous studies focused on image classification tasks without taking into account the deep learning architecture. Therefore, a fundamental understanding of suitable deep learning structures has become a crucial part of progressing myeloid blast cells recognition research. While the previous research focused on implementing deep learning architecture on their respective study case, we suspect that examining the suitability of deep learning structures in blast cells recognition will be useful for a basis to designing a high-performance model on this study case.

Here, we are conducting a comparative study for convolution neural networks based deep learning in case of blast cells detection in leukocytes. The main objective of this study is to analyze the best structures for blast cells detection in blood cells classification. Mainly, our contribution is evaluating a well-known convolution neural network structure as basis to develop a high-performance model for detecting myeloid blast cells.

II. MATERIAL AND METHODS

As a comparative study, the experiment was conducted under the same dataset, environment, and hyperparameter settings. Related studies have their own classification scheme [30] while focusing on small number of classes to build a high-performance model [31] - [33]. Hence, this study focuses on leukocyte classification in addition to detecting blast cells as a screening for acute myeloid leukemia diagnosis.

A. BONE MARROW DATASET

Microscopic images of blood cells in this study are secondary data which provided by MLL Munich Leukemia Laboratory [21] through The Cancer Imaging Archive [34]. Bone marrow cytologic images were obtained from 961 patients diagnosed with a variety of hematological diseases between 2011 and 2013. Bone marrow smears are digitized with Zeiss Axio Imager Z2 and annotated into 21 classes. However, this study was using the subset of those datasets which contain 12 classes of leukocyte including blast cells. From the annotated region, the image observed were cropped into 250 × 250-pixel images as shown in FIGURE 1.

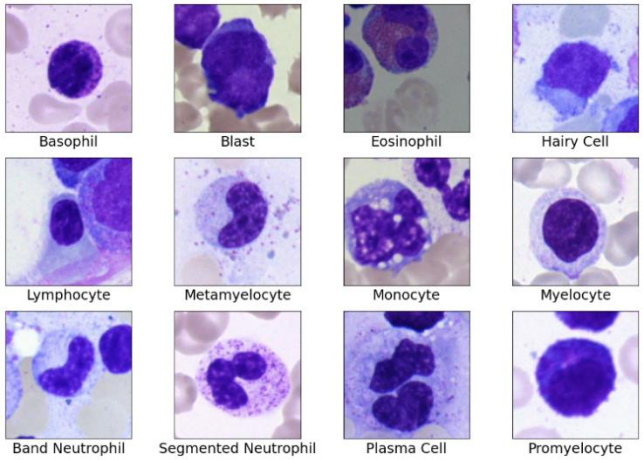

FIGURE 1. Sample image of bone marrow cytologic datasets

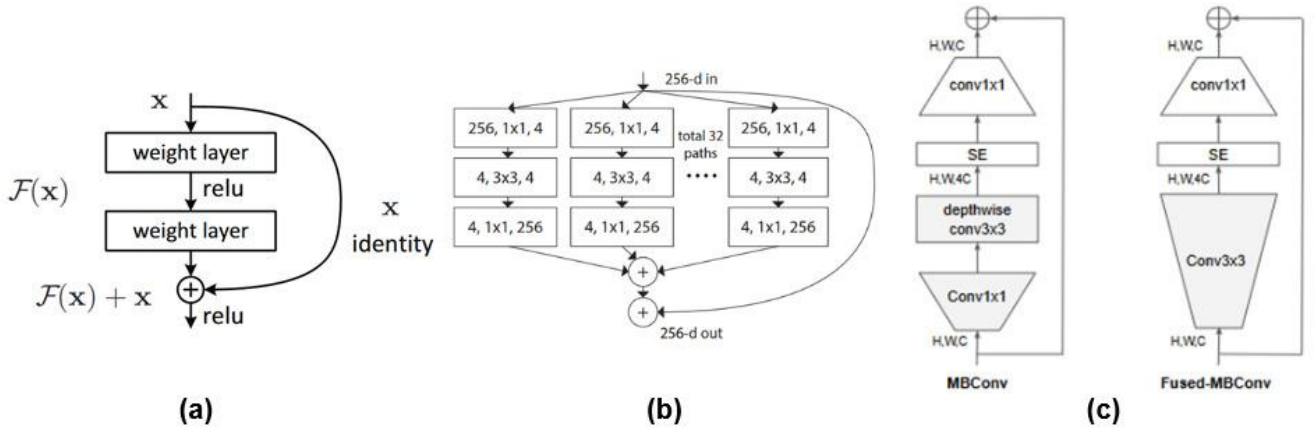

FIGURE 2. CNN based architecture. (a) Residual block of ResNet model, (b) Aggregation of residual block with cardinality of ResNeXt model, (c) MBConv block and Fused-MBConv block of EfficientNetV2

B. CNN BASED ARCHITECTURE

The convolution neural network (CNN) has a well-known reputation in image recognition amongst various domain knowledge. Within CNN based architecture models, research mainly led to model performance and parameter efficiency. Thus, we choose ResNet as representations of the best performance and EfficentNetV2 as representations of the best efficient CNN model. In addition, we also used ResNeXt as a benchmark which also used in previous study [20], [21].

ResNet architecture was built to tackle degradation problems which occurred in very deep models. ResNet introduced residual learning as shown in FIGURE 2(a) which explicitly let the layers approximate the residual function so that a very deep model can avoid degradation in training phase [35]. When ResNet was compared to architecture without residual learning with the same number of layers, ResNet outperformed those architectures and converged earlier.

As a variant of ResNet, ResNeXt also uses a residual learning within this architecture. As shown in FIGURE 2(b), ResNeXt is utilized an aggregation in residual block named cardinality [36]. This architecture shows a small improvement in performance compared to ResNet. In addition, this improvement comes from the increment of cardinality rather than number of layers which shows the effectivity of cardinality.

EfficientNetV2 was built as a parameter efficient oriented model. However, this model shows a good performance in addition to parameter efficiency and claimed in having a better performance compared to vision transformer [37]. The efficiency of this model comes from depthwise convolution, as shown in FIGURE 2(c), which makes the number of parameters and FLOPs more efficient. However, depthwise convolution has a drawback in learning phase at shallow layer which lead to MBConv block learn ineffectively. Therefore, this architecture utilizes a Fused-MBConv layer in order to speed up the learning time. Consequently, Fused-MBConv layer slightly increased the number of parameters and FLOPs. Hence, MBConv and Fused-MBConv are combined using Neural Architecture Search (NAS) to optimize this architecture which can learn faster at shallow layer and parameters efficiently.

C. EXPERIMENT SETTINGS

In this study, the experiment was conducted using PyTorch framework under DGX-1 environment. The available GPU in this setup is Nvidia V100. The models we compared are ResNet 18, ResNet 50, ResNeXt 50 with 32 cardinalities, EfficientNetV2S, and EfficientNetV2M. In order to provide a fair comparison, we conducted the inference of those models under the same environment. For the dataset, we randomly split into 3 (three) parts as training set (60%), validation set (15%), and test set (25%). In addition, we conducted data augmentation such as flipping and rotating to provide variety to dataset. Furthermore, we considered random over sampling to increase the number of minority class while maximize the variety of majority class [38] as the dataset was imbalanced as shown in FIGURE 3.

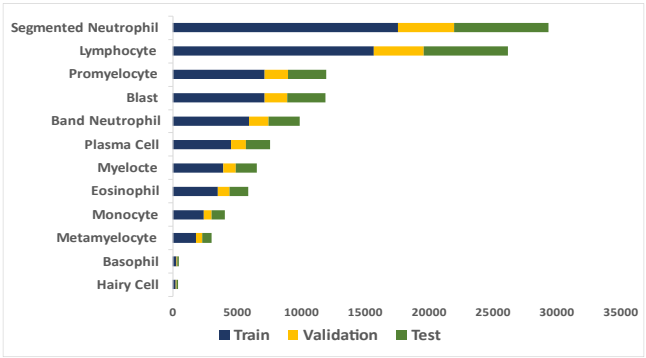

FIGURE 3. Microscopic image sample distribution for each class

Although hyperparameters also take a crucial part in the learning process, some restrictions were unavoidable to reach a trustworthy comparison. Therefore, the hyperparameters which are used in this study are the same setup for every model. The hyperparameters setup for this study is shown in TABLE 1. We constrain the learning iteration to 100 epoch

and applied early stopping on top of using validation set to avoid overfitting. In addition, the same batch size settings applied as restriction to compare the learning time of each model.

TABLE 1.
Hyperparameter settings for restricted comparison

Epoch	100
Batch size	32
Learning rate	10^{-3}
Optimizer	Adam
Loss function	Cross Entropy

Cross entropy loss function is commonly used for supervised classification [39], [40]. It coincides with the logistic loss applied to the outputs of a neural network, when the SoftMax is used. Assume dataset is defined as a set of N samples of label y that correspond to input x from data distribution $D = \{(x_1, y_1), \dots, (x_n, y_n)\}$ with number of class C. Let $\mathcal{L}$ as loss function with w as weight, cross entropy loss can be notated as in (1).

$$\mathcal{L} = -\sum_{n=1}^{N}\sum_{c=1}^{C} w_c \, log \frac{exp(x_{n,c})}{\sum_{i=1}^{C} exp(x_{n,i})} y_{n,c} \quad (1)$$

As in optimizer, we consider adaptive moment estimation (Adam) which robust and well-suited to a wide range of non-convex optimization problems [41]. Furthermore, Adam demonstrated that empirical convergence meets the expectations of theoretical analysis. In mathematical notation, consider θ_{t-1} as model parameter that computed using gradient descent become θ_t at timestep t with learning rate α. Therefore, optimization at first step can be written as in (2).

$$\theta_t = \theta_{t-1} - \alpha \nabla_\theta \mathcal{L}\big(f(\theta)\big) \quad (2)$$

Since Adam utilize momentum to accelerate gradient descent, it requires decay rate to estimate the momentum. Let β_1 and β_2 as decay rate, m_t as update biased first moment estimation at timestep t, and v_t as update biased second moment estimation at timestep t. Consider those parameter, computed bias corrected first moment $\widehat{m_t}$ and computed bias corrected second moment $\widehat{v_t}$ can be written as follows:

$$m_t \leftarrow \beta_1 m_{t-1} + (1-\beta_1)\nabla_\theta f(\theta_{t-1}) \quad (3)$$

$$v_t \leftarrow \beta_2 v_{t-1} + (1-\beta_2)\big(\nabla_\theta f(\theta_{t-1})\big)^2 \quad (4)$$

$$\widehat{m_t} \leftarrow m_t/(1-\beta_1^t) \quad (5)$$

$$\widehat{v_t} \leftarrow v_t/(1-\beta_2^t) \quad (6)$$

Intuitively, Adam adapting to the gradient descent after every iteration so that it remains controlled and unbiased throughout the process. Instead of normal weight parameters, Adam take the momentum in addition of constant ϵ to avoid division by zero. When it put in together, adam can be written as in (7).

$$\theta_t \leftarrow \theta_{t-1} - \alpha\widehat{m_t}/(\widehat{v_t} + \epsilon) \quad (7)$$

D. PERFORMANCE EVALUATIONS

Performance assessment in this study was conducted in two levels evaluation which are macro level and micro level. Metric evaluations that are used at macro level are accuracy, precision, recall, and F1 score. As N number of evaluation of predictive model f for image x and label y from evaluation sample distribution $S = \{(x_1, y_1), \dots, (x_n, y_n)\}$, those metrics can be mathematically notated as follows:

$$Accuracy = \frac{1}{N}\sum_{i=1}^{N} f(x) = y \quad (8)$$

$$Precision = \mathbb{P}(y = +|f(x) = +) \quad (9)$$

$$Recall = \mathbb{P}(f(x) = +|y = +) \quad (10)$$

$$F1\ score = 2 \times \frac{precision \times recall}{precision + recall} \quad (11)$$

At some point, accuracy metric can be affected by type I and type II statistical error. In statistical hypothesis testing, a type I error is the mistaken rejection of a null hypothesis that is actually true (overestimate). On the other hand, a type II error is the failure to reject a null hypothesis that is actually false (underestimate). Therefore, precision and recall metric come to play in considering those error respectively. In addition, F1 score metric considers both errors at the same time under assumption that both errors are weighted equally.

In micro level, the performance assessment was evaluating for every class and specifically in detecting blast cells. Metric evaluations that are used in this level are F1 score, false detection rate (FDR), and false negative rate (FNR). As we consider the model was learning to be a semi assisted medical diagnostic tool, the usage of FDR and FNR plays a crucial role to evaluate type I and type II error respectively. In mathematical equation, FDR and FNR can be written as follows:

$$FDR = 1 - Precision \quad (12)$$

$$FNR = 1 - Recall \quad (13)$$

III. RESULT

There are two things that become concern in this comparative study which are training phase comparison and model performance comparison. Training phase comparison takes a part in comparing how fast model reach convergence which consider the number of iterations for each architecture. On the other hand, model performance comparison evaluating each architecture in classifying the unseen data of leukocyte images.

A. TRAINING PHASE COMPARISON

Comparing how models reach convergence in the learning phase plays a part in selecting model which the most timesaving and effectively utilizing computational resource. As shown in FIGURE 4, ResNet and ResNeXt architecture which utilize residual block reach convergence faster than EfficientNetV2 architecture. Furthermore, ResNet 18, ResNet 50, and ResNeXt 50 reach convergence within 10 – 25

TABLE 2.
Model evaluation at macro level

Model	Parameter	Accuracy↑	Precision↑	Recall↑	F1 Score↑
ResNet 18	**11182668**	**78.11%**	**71.94%**	**75.00%**	**73.03%**
ResNet 50	23532620	75.81%	68.13%	70.87%	68.94%
ResNeXt 50	23004492	75.95%	67.37%	72.64%	68.96%
EfficientNetV2S	20192860	77.69%	70.05%	73.77%	71.40%
EfficientNetV2M	52873728	75.84%	67.91%	71.86%	68.69%

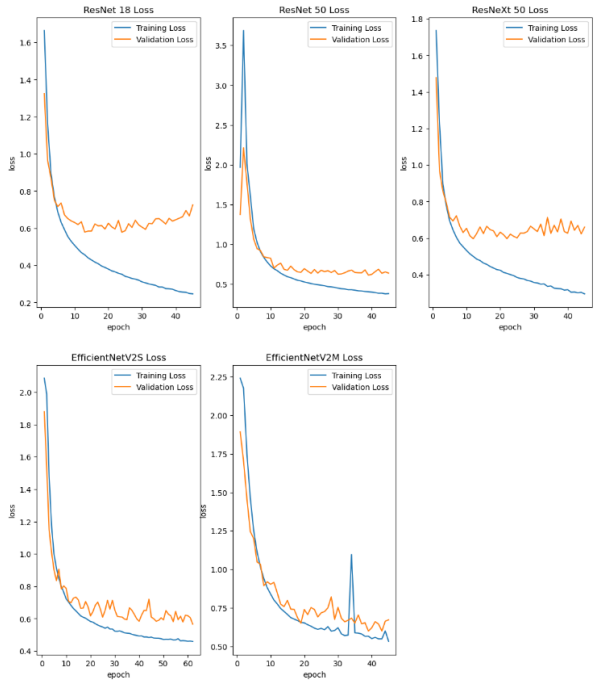

FIGURE 4. Graph of loss function during learning stage

iterations. It shows that ResNet and ResNeXt architecture can utilize resources better than EfficientNetV2. However, ResNet and ResNeXt architecture tend to overfit easily rather than EfficientNetV2. As a result, this comparison study shows that in this study case there is a tradeoff between convergence rate and overfitting potential.

B. MODEL PERFORMANCE COMPARISON

At macro level, ResNet 18 outperforms the remaining models in every metric as shown in TABLE 2. This result is consistent with degradation problem where the performance of deep learning decreases as the number of layers increases even though the analysis model complexity also increases. This also shows that residual learning has been utilized ineffectively in this study case. At the same time, aggregation of ResNeXt architecture barely outperformed ResNet architecture when those architectures using the same number of layers as shown in ResNet 50 and ResNeXt 50. However, the number of parameters in ResNeXt 50 model have less 50k (fifty thousand) parameters than ResNet 50 model so that in implementation perspectives ResNeXt outperformed ResNet architecture in resource utilization.

Although the predictive performance of EfficientNetV2 was underperformed to ResNet and ResNeXt, there are things that should be taken into consideration. First, EfficientNetV2 tends to avoid overfitting at the learning phase as shown in FIGURE 4. Second, predictive performance of EfficientNetV2 as represented by EfficientNetV2S outperformed ResNet 50 and ResNeXt 50 when compared to the model with the same number of parameters. These results raise an assumption that parameters of EfficientNetV2 can recognizing a general pattern efficiently. It is supported by predictive performance of EfficientNetV2M which has twice the number of parameters as ResNet 50 and ResNeXt 50 while maintaining similar performance. However, further study requires to authenticate this assumption.

Performance evaluation at micro level using F1 score shows that ResNet 18 predictive performance in 7 out 12 class are the highest as in TABLE 3. Therefore, ResNet 18 also outperformed the remaining models in micro level. However, if the comparison specified in ResNet 50, ResNeXt 50, and EfficienrNetV2S which have similar number of parameters, EfficientNetV2S predictive performance outperforms the remaining model in 8 out 12 class. In addition, comparison between ResNet 50, ResNeXt 50, and EfficientNetV2M based on assumption in macro level evaluation shows that ResNeXt 50 outperforms the remaining models in 7 out 12 class. The most interesting part of those comparisons is every model has a class which outperforms the other. This result was matching the No-Free-Lunch theorem which states that machine learning model cannot be used in general task but only at a specific task [16], [42]. In this study, the specific task is to semi assisted doctor to diagnose acute myeloid leukemia so that the evaluation should be focused on detecting blast cells. As the result, ResNet 18 outperforms the remaining model in this study case as shown in TABLE 3.

As F1 score metric simplify precision and recall metric, the result of TABLE 3 is less convincing when in medical diagnosis adhere to principles of zero mistakes for patient safety. Therefore, type I and type II errors become a crucial part of detecting blast cells. In addition, FDR and FNR metrics play a role in evaluating those errors in number as shown in TABLE 4. Considered in doctor point of view, the most beneficial model for semi assisted medical diagnosis is the model with low type II error which prevents a patient to diagnose as healthy due to underestimate. On contrary, models with a low statistical testing error in both type I

TABLE 3.
Model evaluation with F1 score metric at micro level

Cell	ResNet 18	ResNet 50	ResNeXt 50	EfficientNetV2S	EfficientNetV2M
Basophil	**55.00%**	44.61%	42.57%	43.75%	36.86%
Blast	**79.45%**	77.42%	73.28%	78.35%	75.85%
Eosinophil	95.09%	94.11%	**95.16%**	92.58%	92.46%
Hairy Cell	**56.25%**	42.17%	37.13%	53.18%	46.15%
Lymphocyte	84.88%	83.16%	81.75%	**85.90%**	84.13%
Metamyelocyte	48.85%	45.66%	51.14%	**52.41%**	47.81%
Monocyte	**66.84%**	62.24%	65.70%	64.39%	61.42%
Myelocyte	**66.04%**	63.33%	64.13%	64.41%	63.87%
Band Neutrophil	73.16%	72.31%	**74.69%**	73.34%	71.83%
Segmented Neutrophil	84.19%	80.74%	83.31%	**84.94%**	82.64%
Plasma Cell	**90.55%**	87.48%	84.24%	88.67%	87.29%
Promyelocyte	**76.01%**	74.09%	74.45%	74.91%	74.00%

and type II error is an ideal, but high type I error also tolerable due understanding of model's overestimation beforehand led to further medical examination. In that case, ResNet 50 model is the best model specific in detecting blast cells due to the least value in FNR metric. Moreover, ResNeXt 50 and EfficientNetV2M which having high underestimation in predicting blast cells be avoided when FDR metric overlooked. However, those models are the best model in the case of studying blast cells of patient as the least value in FDR metric.

TABLE 4.
Blast cells detection error

Model	FDR↓	FNR↓
ResNet 18	21.20%	19.88%
ResNet 50	25.81%	**19.04%**
ResNeXt 50	18.46%	33.47%
EfficientNetV2S	23.57%	19.64%
EfficientNetV2M	**15.51%**	31.20%

During comparison at micro level, contradiction arise when in F1 score metric result shows degradation of performance in architecture as the number of parameters or layers increases while in FNR metric shows ResNet 50 outperforms ResNet 18 in detecting blast cells. Furthermore, there are correlations between class with low performance and the number of samples for that class. It is specific in metamyelocyte, basophil, and hairy cell which have a bad performance and least number of samples as shown in TABLE 3 and FIGURE 3 respectively. Those facts led to the fact that the imbalance dataset is the main problem. Moreover, solving imbalance dataset problem using random over sampling also led minority class to overfit which make it even worse. Conversely, random under sampling can avoid minority class to overfit at the sacrifice of reducing observed variety in majority sample. Semi supervised learning can be an alternative to avoid minority class to overfit while sustain observed variety in majority sample [25], [26]. However, semi supervised learning has its own problem that requires a number of samples to classify unlabeled data or generate a synthetic image. Thus, the lack of dataset as a main problem is still unsolved and unpredictable even though model predictive performance enhanced.

IV. DISCUSSION

As for this study case, we consider CNN based architecture to be the most suitable deep learning model. The architectures that used in this study limited to the well-known best performance and most efficient on top of benchmark from previous study. Although this study covered mainstream convolutional blocks such as residual block, aggregation block, and MBConv block as shown in FIGURE 2, there are several modifications and CNN based architectures that are out of this comparative study's scope. Although another deep learning variant such as Vision Transformer in transformer-based architecture shows a competitive performance to CNN based architecture, there are several problems in this architecture which are lack inductive bias, high computational resource, and more data intensive than CNN based architecture [43]. Furthermore, EfficientNetV2 architecture also outperforms Vision Transformer in ImageNet scenario [37]. A fused CNN and transformer also suggested as a solution for those problems [44]. However, as this fused architecture only slightly leverages those problems, we considered CNN based architectures are more desirable rather than transformer-based or fused CNN-transformer at this point. Experiment results justify our perspective which the number of samples in this study case still a problem.

A limited number of microscopic images in this study case are related to a rate of new case of acute myeloid leukemia is estimated to 23.1% of total leukemia case worldwide in 2017 [11]. In addition, several blood cells are low in number so that imbalanced data was unavoidable. Big data or large-scale participative observations worldwide can be considered to tackle this unbalanced and lacking data problem. However, those problems cannot be solved immediately. Therefore, an alternative such as transfer learning is more practical at this point which exploit a similar domain into this study case [45]. Moreover, automated blood cells image recognition in broader

and more specific subtype can be considered through meta learning which involve training on a variety of task to learn generalize knowledge [46], [47]. In this case, a specific subtype cells for diagnosing acute myeloid leukemia subtype can also be recognized.

V. CONCLUSION

As the main objective of this study is to examine an established convolution neural network structure in detecting blast cells for semi-assisted diagnose of acute myeloid leukemia, this study presents a comparison of a well-known CNN based architectures in image recognition. The experiment shows that ResNet outperforms the selected architecture in this study especially ResNet 18 for classify leukocyte in macro and micro level. The contradiction arises when ResNet 50 outperforms the remaining model in specific study case using FNR metric. In addition, predictive performance on every metric for each model has not been able to reach automatic leukocyte recognition for medical purposes when considered a zero-mistake principle. According to our findings, we suspect an architecture combining a residual block and MBConv block can reach an early convergence point with good performance while avoiding overfitting. We also found that imbalanced dataset and lacking data sample become the significant problem to lead model underperformed at some class. Therefore, CNN based architectures are preferable than transformer-based architectures in this study case.

Although fully automated models are unreachable in this study, we conclude that semi-automated models are still tolerable at this point. Regardless of imbalanced dataset problems, CNN already shows its performance in several domain tasks. However, a comparative study between CNN based architectures and transformer-based architectures requires when imbalanced dataset and lacking data sample can be solved. As for the future work, we suggest that transfer learning and meta-learning can be considered for blast cells recognition to build a fully automated model under imbalanced dataset and lacking data sample problem.

ACKNOWLEDGMENT

This work was a part of coursework in master program at Faculty of Computer Science, Universitas Indonesia. The authors are grateful to Tokopedia AI Center for providing Nvidia V100 in supporting this study. Furthermore, the authors are also grateful to MLL Munich Leukemia Laboratory in providing bone marrow cytology dataset for public through The Cancer Imaging Archive.

REFERENCES

[1] V. V. B. Reddy and D. Morlote, "Examination of Blood and Marrow Cells," in *Williams Hematology*, 10th ed., K. Kaushanky, M. A. Lichtman, J. T. Prchal, M. Levi, L. J. Burns, and D. C. Linch, Eds., McGraw Hill, 2021, pp. 11–30.

[2] R. G. Bagasjvara, I. Candradewi, S. Hartati, and A. Harjoko, "Automated detection and classification techniques of Acute leukemia using image processing: A review," in *2016 2nd International Conference on Science and Technology-Computer (ICST)*, IEEE, Oct. 2016, pp. 35–43. doi: 10.1109/ICSTC.2016.7877344.

[3] A. Shah, S. S. Naqvi, K. Naveed, N. Salem, M. A. U. Khan, and K. S. Alimgeer, "Automated Diagnosis of Leukemia: A Comprehensive Review," *IEEE Access*, vol. 9, pp. 132097–132124, 2021, doi: 10.1109/ACCESS.2021.3114059.

[4] K. R. Kampen, "The discovery and early understanding of leukemia," *Leuk Res*, vol. 36, no. 1, pp. 6–13, Jan. 2012, doi: 10.1016/j.leukres.2011.09.028.

[5] E. J. Freireich, P. H. Wiernik, and D. P. Steensma, "The Leukemias: A Half-Century of Discovery," *Journal of Clinical Oncology*, vol. 32, no. 31, pp. 3463–3469, Nov. 2014, doi: 10.1200/JCO.2014.57.1034.

[6] G. H. Jackson and P. R. A. Taylor, "Acute Myeloid Leukaemia," *Drugs Aging*, vol. 19, no. 8, pp. 571–581, 2002, doi: 10.2165/00002512-200219080-00003.

[7] K. Riding, "Acute Myeloid Leukemias," in *Clinical Laboratory Hematology*, 4th ed., S. B. McKenzie, K. Landis-Piwowar, and J. L. Williams, Eds., Pearson, 2020, pp. 583–605.

[8] J. L. Liesveld and M. A. Lichtman, "Acute Myelogenous Leukemia," in *Williams Hematology*, 10th ed., K. Kaushanky, M. A. Lichtman, J. T. Prchal, M. Levi, L. J. Burns, and D. C. Linch, Eds., McGraw Hill, 2021, pp. 1445–1521.

[9] D. A. Arber *et al.*, "The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia," *Blood*, vol. 127, no. 20, pp. 2391–2405, May 2016, doi: 10.1182/blood-2016-03-643544.

[10] D. A. Arber *et al.*, "International Consensus Classification of Myeloid Neoplasms and Acute Leukemias: integrating morphologic, clinical, and genomic data," *Blood*, vol. 140, no. 11, pp. 1200–1228, Sep. 2022, doi: 10.1182/BLOOD.2022015850.

[11] Y. Dong *et al.*, "Leukemia incidence trends at the global, regional, and national level between 1990 and 2017," *Exp Hematol Oncol*, vol. 9, no. 1, pp. 1–11, Jun. 2020, doi: 10.1186/S40164-020-00170-6/FIGURES/5.

[12] H. Kantarjian *et al.*, "Acute myeloid leukemia: current progress and future directions," *Blood Cancer J*, vol. 11, no. 2, p. 41, Feb. 2021, doi: 10.1038/s41408-021-00425-3.

[13] C. Récher *et al.*, "Long-term survival after intensive chemotherapy or hypomethylating agents in AML patients aged 70 years and older: a large patient data set study from European registries," *Leukemia*, vol. 36, no. 4, pp. 913–922, Apr. 2022, doi: 10.1038/s41375-021-01425-9.

[14] P. Font *et al.*, "Inter-observer variance with the diagnosis of myelodysplastic syndromes (MDS) following the 2008 WHO classification," *Ann Hematol*, vol. 92, no. 1, pp. 19–24, Jan. 2013, doi: 10.1007/S00277-012-1565-4/TABLES/4.

[15] P. Font *et al.*, "Interobserver variance in myelodysplastic syndromes with less than 5 % bone marrow blasts: unilineage vs. multilineage dysplasia and reproducibility of the threshold of 2 % blasts," *Ann Hematol*, vol. 94, no. 4, pp. 565–573, Mar. 2015, doi: 10.1007/S00277-014-2252-4/TABLES/4.

[16] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning*. Cambridge University Press, 2014. doi: 10.1017/CBO9781107298019.

[17] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[18] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[19] S. Mohapatra, D. Patra, and S. Satpathy, "An ensemble classifier system for early diagnosis of acute lymphoblastic leukemia in blood microscopic images," *Neural Comput Appl*, vol. 24, no. 7–8, pp. 1887–1904, Jun. 2014, doi: 10.1007/S00521-013-1438-3/FIGURES/12.

[20] C. Matek, S. Schwarz, K. Spiekermann, and C. Marr, "Human-level recognition of blast cells in acute myeloid leukaemia with convolutional neural networks," *Nat Mach Intell*, vol. 1, no. 11, pp. 538–544, Nov. 2019, doi: 10.1038/s42256-019-0101-9.

[21] C. Matek, S. Krappe, C. Münzenmayer, T. Haferlach, and C. Marr, "Highly accurate differentiation of bone marrow cell morphologies using deep neural networks on a large image data set," *Blood*, vol. 138, no. 20, pp. 1917–1927, Nov. 2021, doi: 10.1182/blood.2020010568.

[22] S. J. Azzahra, R. Sigit, H. Yuniarti, Y. Hernaningsih, and A. Imannurohma, "Identification of Acute Myeloid Leukemia (AML)

Subtypes: M1, M2, M3 on White Blood Cells Using Microscopic Images," *IES 2023 - International Electronics Symposium: Unlocking the Potential of Immersive Technology to Live a Better Life, Proceeding*, pp. 454–459, 2023, doi: 10.1109/IES59143.2023.10242476.

[23] A. Setiawan, A. Harjoko, T. Ratnaningsih, E. Suryani, Wiharto, and S. Palgunadi, "Classification of cell types in Acute Myeloid Leukemia (AML) of M4, M5 and M7 subtypes with support vector machine classifier," in *2018 International Conference on Information and Communications Technology (ICOIACT)*, IEEE, Mar. 2018, pp. 45–49. doi: 10.1109/ICOIACT.2018.8350822.

[24] C. Jung, M. Abuhamad, D. Mohaisen, K. Han, and D. H. Nyang, "WBC image classification and generative models based on convolutional neural network," *BMC Med Imaging*, vol. 22, no. 1, pp. 1–16, Dec. 2022, doi: 10.1186/S12880-022-00818-1/TABLES/18.

[25] Z. Zhu, Z. Ren, S. Lu, S. Wang, and Y. Zhang, "DLBCNet: A Deep Learning Network for Classifying Blood Cells," *Big Data and Cognitive Computing 2023, Vol. 7, Page 75*, vol. 7, no. 2, p. 75, Apr. 2023, doi: 10.3390/BDCC7020075.

[26] S. Ansari, A. H. Navin, A. B. Sangar, J. V. Gharamaleki, and S. Danishvar, "A Customized Efficient Deep Learning Model for the Diagnosis of Acute Leukemia Cells Based on Lymphocyte and Monocyte Images," *Electronics (Basel)*, vol. 12, no. 2, p. 322, Jan. 2023, doi: 10.3390/ELECTRONICS12020322.

[27] A. Abhishek, N. Santhanam, R. K. Jha, R. Sinha, and K. Jha, "Multi Class Classification of Acute Leukemia using Transfer Learning," *2022 International Conference for Advancement in Technology, ICONAT 2022*, 2022, doi: 10.1109/ICONAT53423.2022.9726083.

[28] P. K. Das, B. Sahoo, and S. Meher, "An Efficient Detection and Classification of Acute Leukemia using Transfer Learning and Orthogonal Softmax Layer-based Model," *IEEE/ACM Trans Comput Biol Bioinform*, 2022, doi: 10.1109/TCBB.2022.3218590.

[29] T. Tamang, S. Baral, and M. P. Paing, "Classification of White Blood Cells: A Comprehensive Study Using Transfer Learning Based on Convolutional Neural Networks," *Diagnostics*, vol. 12, no. 12, p. 2903, Nov. 2022, doi: 10.3390/diagnostics12122903.

[30] A. Acevedo, S. Alférez, A. Merino, L. Puigví, and J. Rodellar, "Recognition of peripheral blood cell images using convolutional neural networks," *Comput Methods Programs Biomed*, vol. 180, p. 105020, Oct. 2019, doi: 10.1016/J.CMPB.2019.105020.

[31] M. Hosseini, D. Bani-Hani, and S. S. Lam, "Leukocytes Image Classification Using Optimized Convolutional Neural Networks," *Expert Syst Appl*, vol. 205, p. 117672, Nov. 2022, doi: 10.1016/J.ESWA.2022.117672.

[32] C. Jung, M. Abuhamad, D. Mohaisen, K. Han, and D. Nyang, "WBC image classification and generative models based on convolutional neural network.," *BMC Med Imaging*, vol. 22, no. 1, p. 94, May 2022, doi: 10.1186/s12880-022-00818-1.

[33] S. Srimahima, G. Yuvarani, and L. K. Nandhini, "White Blood Cells Classification Using Deep Learning Technique," *Lecture Notes in Networks and Systems*, vol. 520, pp. 79–86, 2023, doi: 10.1007/978-981-19-5331-6_9/COVER.

[34] K. Clark *et al.*, "The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository," *J Digit Imaging*, vol. 26, no. 6, pp. 1045–1057, Dec. 2013, doi: 10.1007/s10278-013-9622-7.

[35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[36] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, "Aggregated Residual Transformations for Deep Neural Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jul. 2017, pp. 5987–5995. doi: 10.1109/CVPR.2017.634.

[37] M. Tan and Q. V. Le, "EfficientNetV2: Smaller Models and Faster Training," *Proceedings of the 38th International Conference on Machine Learning*, pp. 10096–10106, Apr. 2021, Accessed: Dec. 21, 2022. [Online]. Available: http://arxiv.org/abs/2104.00298

[38] G. Menardi and N. Torelli, "Training and assessing classification rules with imbalanced data," *Data Min Knowl Discov*, vol. 28, no. 1, pp. 92–122, Jan. 2014, doi: 10.1007/S10618-012-0295-5/METRICS.

[39] K. P. Murphy, *Machine Learning A Probabilistic Perspective*. The MIT Press, 2012.

[40] L. Li, M. Doroslovacki, and M. H. Loew, "Approximating the Gradient of Cross-Entropy Loss Function," *IEEE Access*, vol. 8, pp. 111626–111635, 2020, doi: 10.1109/ACCESS.2020.3001531.

[41] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, Dec. 2014, Accessed: Dec. 29, 2022. [Online]. Available: http://arxiv.org/abs/1412.6980

[42] D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization," *IEEE Transactions on Evolutionary Computation*, vol. 1, no. 1, pp. 67–82, Apr. 1997, doi: 10.1109/4235.585893.

[43] A. Vaswani *et al.*, "Attention Is All You Need," *Adv Neural Inf Process Syst*, vol. 30, Jun. 2017, Accessed: Dec. 21, 2022. [Online]. Available: http://arxiv.org/abs/1706.03762

[44] X. Liu, Y. Hu, and J. Chen, "Hybrid CNN-Transformer model for medical image segmentation with pyramid convolution and multi-layer perceptron," *Biomed Signal Process Control*, vol. 86, p. 105331, Sep. 2023, doi: 10.1016/J.BSPC.2023.105331.

[45] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans Knowl Data Eng*, vol. 22, no. 10, pp. 1345–1359, 2010, doi: 10.1109/TKDE.2009.191.

[46] J. Vanschoren, "Meta-Learning," in *Automated Machine Learning: Methods, Systems, Challenges*, F. Hutter, L. Kotthoff, and J. Vanschoren, Eds., Springer, Cham, 2019, pp. 35–61. doi: 10.1007/978-3-030-05318-5_2.

[47] T. M. Hospedales, A. Antoniou, P. Micaelli, and A. J. Storkey, "Meta-Learning in Neural Networks: A Survey," *IEEE Trans Pattern Anal Mach Intell*, vol. 44, no. 9, pp. 1–1, Sep. 2021, doi: 10.1109/TPAMI.2021.3079209.

AUTHOR BIOGRAPHY

AHMAD BADRUZZAMAN **is** currently pursuing a master's degree in computer science at the Faculty of Computer Science, Universitas Indonesia. He completed his bachelor's degree in geodesy and geomatics engineering at the Faculty of Earth Science and Technology, Institut Teknologi Bandung. He is a student member of the IEEE Computer Society, IEEE Geoscience and Remote Sensing Society, and IEEE Engineering in Medicine and Biology Society. His interest is mainly in computational science in applications on earth observation and medical image processing. His skills include computer vision, feature extraction, image processing, machine learning, quantum computing, remote sensing, spatial analysis, and spectral imaging.

ANIATI MURNI ARYMURTHY is currently a professor of computer science with the Faculty of Computer Science, Universitas Indonesia. She is an Expert in image processing and spatial data. Her skills and expertise are in classifications, feature extraction, algorithms, feature selection, algebra, genetic algorithms, image enhancement, wavelet, and self-organizing maps. She has published a large number of scientific articles in her field