Received: 30 August 2025 | Revised: 14 April 2026 | Accepted: 25 June 2026 | Published online: 20 July 2026

RESEARCH ARTICLE

Artificial Intelligence and Applications
2026, Vol. 00(00) 1–11
DOI: 10.47852/bonviewAIA62027495

BON VIEW PUBLISHING

VGG16-Based Multiscale Feature Fusion for Robust Blood Cell Cancer Classification

Simeon Yuda Prasetyo[1,*]

[1] *Computer Science Department, Bina Nusantara University, Indonesia*

Abstract: Leukemia, particularly acute lymphoblastic leukemia (ALL), is one of the most prevalent and fatal blood cancers, with more than 460,000 new cases and over 300,000 related deaths reported globally in 2021. Despite advances in deep learning, conventional convolutional neural network models applied to blood cell cancer classification often deliver suboptimal performance, with reported accuracies frequently below 90%. These limitations are due to difficulty in capturing subtle morphological variations, high similarity among malignant subtypes, and problems like class imbalance and noise in peripheral blood smear images. To overcome these difficulties, this study proposes a VGG16-based multiscale feature fusion (MFF) model that fuses the shallow and deep feature maps from different convolution blocks of the VGG16 backbone. Before classification, these feature maps are spatially aligned and concatenated into a single multiscale representation at a consistent resolution. The model was tested on the Blood Cell Cancer (ALL) dataset, which includes 3242 peripheral blood smear images that are classified into four categories: benign, malignant early Pre-B, malignant Pre-B, and malignant Pro-B. The experimental results showed that the proposed VGG16-MFF obtained 99.69% accuracy, precision, recall, and F1-score, which significantly outperformed baseline architectures such as ResNet50 (50.00%), ResNet101 (76.85%), VGG16 (88.27%), and VGG19 (87.34%). These results support the effectiveness of MFF in improving classification robustness, and future work may consider its extension to larger and more varied medical imaging datasets.

Keywords: blood cell cancer, convolutional neural network, medical image classification, multiscale feature fusion, VGG16

1. Introduction

Blood cancers, particularly forms such as leukemia, remain a major global health concern despite recent advancements in treatment strategies. It still accounts for a major proportion of the world's morbidity and mortality. This highlights the need for better and reliable diagnostic approaches. Leukemia ranked 13th in cancer incidence worldwide in 2021 and 10th in cancer-related mortality, with more than 461,000 new cases and a mortality rate of 3.89 per 100,000 persons [1]. While age-standardized rates have decreased, absolute case numbers are increasing due to population growth and ageing [2].

By 2050, over 519,000 leukemia cases are expected to occur worldwide, with the steepest growth expected in low- and middle-socio-demographic index (SDI) regions where healthcare resources remain scarce [2]. The elderly and areas with less robust health systems bear the brunt of this load [3].

Acute lymphoblastic leukemia (ALL) is the most common malignancy in children worldwide [4]. Despite advancements in treatment, 67,008 new cases of pediatric ALL were reported in 2020, indicating a continued worldwide health concern. Differential access to resources for diagnosis and treatment is shown by regional variations in incidence and mortality rates. For instance, leukemia-related mortality is still much greater in low-SDI nations than in high-income areas [3].

Furthermore, compared to women, men are disproportionately impacted, with increased incidence, mortality, and disability-adjusted life years [1]. Additionally, projections indicate that by 2030, there will be more than 720,000 new instances of leukemia worldwide and about 367,000 fatalities associated with the disease [5]. Notably, the most prevalent subtypes in adults are still acute myeloid leukemia and chronic lymphocytic leukemia, with regional differences in subtype frequency, even if ALL is more common in children [6].

In addition to incidence, leukemia represents a major economic and social burden, especially in low-resource settings with limited diagnostic and treatment infrastructure. The outcomes for adolescents and young adults with ALL are particularly poor in countries with limited access to healthcare and low health expenditure per capita [7].

Several recent studies on convolutional neural network (CNN)-based methods for classifying blood cancer, especially leukemia, have been done; however, most of them report suboptimal performance with accuracies below 90%, which suggests that better architectures are needed. For instance, Ahad et al. [8] employed a VGG19-based CNN with transfer learning and obtained an accuracy of only 72.29%, indicating the difficulties of traditional feature fusion in blood cancer detection.

Similarly, Hidayat et al. (2023) used a hybrid CNN-XGBoost method for leukemia detection, where the base CNN model achieved 85.32%, and after optimization, it improved slightly to

***Corresponding author:** Simeon Yuda Prasetyo, Computer Science Department, Bina Nusantara University, Indonesia. Email: simeon.prasetyo@binus.ac.id

85.43% [9]. Pooja and Megha [10] have tested a CNN for classification of ALL and multiple myeloma (MM) using images of bone marrow cells. The model could reach an accuracy of 85–88% depending on the type of cancer.

Another work uses deep CNNs with feature selection techniques for multiclass leukemia classification by Rahman et al. [11]. Their initial CNN model achieved less than 90% accuracy before optimization, indicating the challenge of distinguishing leukemia subtypes without sophisticated feature engineering. Likewise, Shourie et al. [12] proposed CNN-based classification models for blood subtypes. They found that the multiclass accuracy was below 90% due to class imbalance problems.

Another work by Imane et al. [13] proposed a CNN for early-stage leukemia diagnosis using white blood cell images. Their model, while promising, also failed to surpass the 90% accuracy mark, underscoring the challenge of early detection using raw image data alone. Additionally, Claro et al. [14] found that even with ensemble and multilevel CNN configurations on a 4-class leukemia dataset, the model only achieved 94.73% and 94.59%—better but still problematic when scaled down to less-augmented datasets.

More recently, Hamdi et al. [15] have proposed a hierarchical CNN for the classification of leukemia subtypes more recently. Their model had an accuracy of about 90%, but this deteriorates when applied to real-world data because of noise and morphological overlaps in leukemic cells. Finally, Al-Bashir et al. [16] tested various CNN models (AlexNet, DenseNet, ResNet, and VGG16) on different leukemia datasets and found that the test accuracy was usually less than 90% in multiclass classification tasks, whereas the validation accuracy was as high as 94%.

The limitations of conventional CNN-based models in blood cancer classification, particularly their inability to capture fine-grained morphological distinctions, have been effectively addressed by MFF. This approach allows the network to capture local details and global contextual information at the same time by pooling features from multiple spatial resolutions. This can greatly improve the efficiency and robustness of classification in challenging biomedical applications such as leukemia detection particularly for visually indistinguishable subtypes.

The superiority of multiscale fusion models has been recently demonstrated in many domains. Gao et al. [17] proposed a dual-branch MFF network with attention mechanisms and obtained better performance than existing methods for hyperspectral image classification. In the same way, Yang et al. [18] proposed the Enhanced Multiscale Feature Fusion Network, which has dilated convolution and parallel paths to capture spatial-spectral features better, and achieved consistently higher classification accuracy.

Other works have further emphasized the value of combining multiscale fusion with attention mechanisms. Qing et al. [19] developed a 3D self-attention multiscale network that enhanced long-range contextual learning in hyperspectral classification tasks. In the medical imaging domain, Gao et al. [20] applied dual-attention multiscale fusion to retinal vessel segmentation and achieved classification accuracies above 97%, while Wang et al. [21] combined DenseNet-121 and VGG16 in a multiscale framework for skin lesion classification, reaching over 91% accuracy.

Pan et al. [22] used cross-scale attention mechanisms in a multiscale fusion model and measured improvements over baseline methods in hyperspectral classification, providing additional support. Likewise, Liu et al. [23] introduced a multiscale feature interactive network that achieved state-of-the-art performance in multi-focus image fusion through coordinate attention and layer-wise integration of multiscale features. Together, these findings provide strong empirical support for adopting MFF in deep learning models for blood cancer detection using cell image data.

Hence, this paper introduces a VGG16-based MFF model to enhance the blood cell cancer classification performance. Different from traditional multiscale fusion methods that usually adopt complex attention mechanisms or computationally expensive architectures, the proposed method uses a lightweight fusion method by selectively fusing the feature maps of block3_conv3, block4_conv3, and block5_conv3 of VGG16. This design is customized to maintain fine-grained morphological details while sustaining high-level semantic representation, which is essential for distinguishing visually similar ALL subtypes. Furthermore, the use of channel-wise concatenation enables the model to retain complementary information across feature levels without introducing significant computational overhead.

2. Research Methodology

This section outlines the research methods utilized in this paper, including the evaluation measures used to gauge model performance, the dataset used, the underlying deep learning architectures, and the suggested MFF strategy.

2.1. Dataset

The Kaggle Blood Cell Cancer (ALL) dataset, which was created for the purpose of detecting ALL, was used in this investigation [24]. In this study, 3242 peripheral blood smear (PBS) pictures from 89 individuals were acquired at the bone marrow laboratory of Taleqani Hospital in Tehran, Iran. Expert laboratory personnel processed and dyed blood samples, and experts used flow cytometry analysis to confirm the ground truth labels. All images were acquired with a Zeiss microscope camera at 100× magnification and stored in JPG format.

The dataset is classified as Malignant—Early Pre-B (979 photos), Malignant—Pre-B (955 photos), Malignant—Pro-B (796 photos), and Benign (512 photos). The hierarchical structure allows for both binary (benign vs malignant) and multi-class classification over the three malignant subtypes. The diversity of malignant lymphoblast subtypes in this dataset is well suited for evaluating the robustness of deep learning models for ALL diagnosis. Figure 1 shows an example of images from the dataset.

2.2. VGG16 as feature extraction backbone

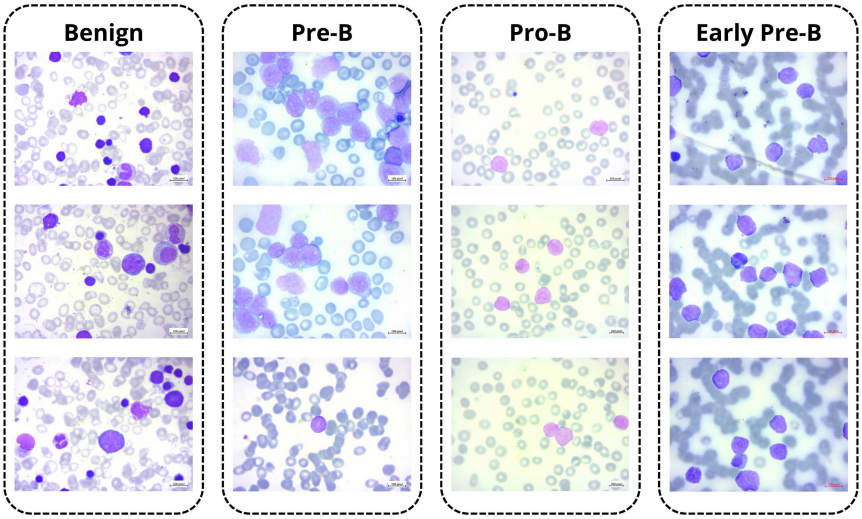
VGG16 is a deep CNN architecture composed of sequential 3 × 3 convolutional layers and max-pooling operations, enabling hierarchical feature learning from low-level edges to high-level semantic representations. As a CNN-based model, VGG16 performs feature extraction through convolution operations, where an input image $\boldsymbol{I}$ is convolved with a kernel $\boldsymbol{K}$ as:

$$(I \times K)(x, y) = \sum_{m} \sum_{n} \boldsymbol{I}(m, n) \cdot \boldsymbol{K}(x - m, y - n) \quad (1)$$

This operation allows the network to capture localized spatial patterns that are critical in medical image analysis [25, 26].

The PBS images have very small inter-class variations regarding nucleus shape, chromatin texture, cytoplasmic granularity, and nucleus-to-cytoplasm ratio when compared to natural image

Figure 1
Sample images from the Blood Cell Cancer (ALL) dataset



classification tasks. Such fine-grained morphological cues are often localized and thus difficult to capture using coarse or overly abstract feature representations. The stacked convolutional structure of VGG16 is particularly apt for blood cell cancer classification as it can progressively abstract these morphological elements in hierarchical layers.

Low-level edges and texture patterns, which are crucial for determining cytoplasmic borders and granularity, are detected by shallow convolutional layers. Intermediate layers capture structural details like chromatin condensation, cell shape deformities, and nuclear contour asymmetries, which are important morphological markers in differentiating leukemic subtypes. This hierarchical feature learning method is well suited for the morphological features observed in the images of ALL cells, where discriminative visual differences are small but clinically relevant.

In several medical imaging studies, it has shown a good generalization capacity, achieving a good accuracy in tasks such as bone anomaly recognition, retinal illness categorization, and brain tumor detection [27–30]. These results suggest that VGG16 is capable of learning complex biomedical features relevant for diagnostic purposes. In addition, its constant structure and the dense hierarchical feature extraction make it a stable backbone for further improvement with advanced feature integration strategies.

During training, the VGG16 backbone is optimized using categorical cross-entropy loss, which is defined as:

$$L(y, \hat{y}) = -\sum_{i=1}^{N}\sum_{j=1}^{C} y_{ij} \log(\hat{y}_{ij}) \quad (2)$$

where y_{ij} is the true label indicator, $\hat{y}_{ij}$ is the predicted probability, N is the number of samples, and C is the number of classes. The weight update rule is expressed as:

$$W_{t+1} = W_t - n\frac{\partial L(W_t)}{\partial W_t} \quad (3)$$

where W_t denotes the weights at iteration t, and η the learning rate. This optimization process enables VGG models to adapt effectively and improves their diagnostic accuracy in medical image classification tasks [31].

Furthermore, there are studies that have explored optimizing VGG16 for real-time and resource-efficient deployment. In field-programmable gate array (FPGA) implementations, accuracy was comparable, but inference time was greatly reduced by using quantization and compression, showing the suitability of VGG16 for edge computing environments [32]. The results demonstrate the trade-off of VGG16 between architectural simplicity and performance, making it a good basis for further improvements such as MFF.

2.3. Multiscale feature fusion

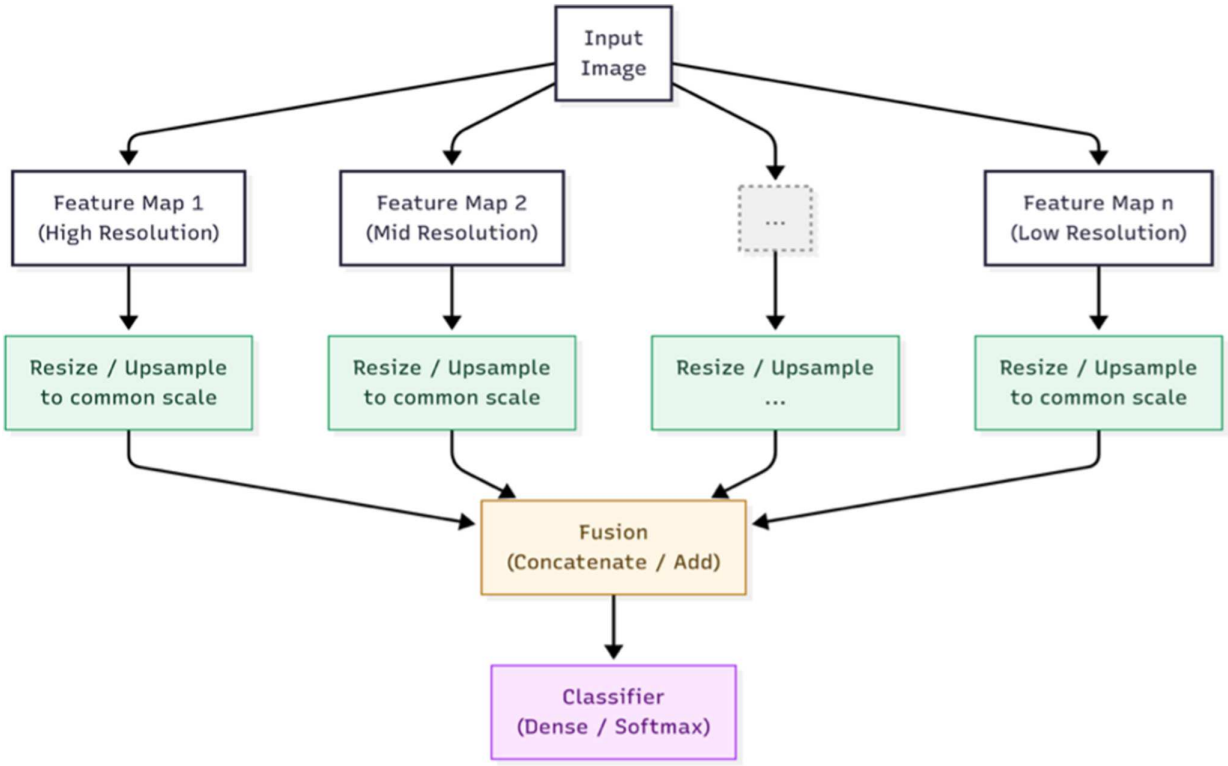
MFF is a popular technique in deep learning models to boost the performance of picture classification, in which the data extracted at multiple spatial resolutions are fused together. The MFF fuses the information from the different convolutional layers, which enables the model to simultaneously learn the global contextual patterns and the fine-grained local structures. In this way, it becomes easier to describe items that change their size, shape, or position from picture to picture. Liu et al. [33] claim that adding multiscale characteristics can greatly increase classification accuracy because various levels of abstraction offer complementing information for making decisions. Similarly, Pan et al. [22] demonstrated that fusing shallow and deep features allows networks to be more robust when handling variations in object scale and background complexity.

MFF has also been applied in lightweight architectures, proving effective even with limited data. For example, Ali et al. [34] proposed a compact fusion network that combined low-level edge features with high-level semantic features, improving recognition in small-sample image datasets. These results indicate that MFF is a practical and scalable method for improving the model accuracy in visual tasks with complex or heterogeneous image structures such as blood cell classification. The visualization of the MFF process is shown in Figure 2 for better clarification of the concept.

2.4. Proposed VGG16-based multiscale feature fusion framework

The proposed framework integrates a VGG16 backbone with an MFF module to enhance feature representation for blood cell cancer classification. The main goal is to get over the drawback of traditional CNNs' single-scale feature

Figure 2
General concept of multiscale feature fusion



extraction, which frequently misses the subtle and hierarchical morphological differences across seemingly similar ALL subtypes. By merging shallow, intermediate, and deep feature maps, the proposed design aims to preserve fine-grained structural details while simultaneously leveraging high-level semantic information.

Let the input blood cell image be denoted as $I \in \mathbb{R}^{H\times W\times 3}$, where H, W, and 3 represent the height, width, and RGB channels, respectively. The image is first processed through the convolutional blocks of the VGG16 backbone:

$$F_k = \phi_k(I), k \in \{1,2,3,4,5\} \tag{4}$$

where $\phi_k(\cdot)$ denotes the operation of the k-th convolutional block in VGG16, consisting of stacked 3×3 convolution layers with ReLU activation followed by a 2×2 max-pooling layer.

2.4.1. Feature-level extraction from VGG16 blocks

To enable multiscale representation, feature maps are extracted from three different depths of the VGG16 network:

$$F_3 = \phi_3(I), F_4 = \phi_4(I), F_5 = \phi_5(I) \tag{5}$$

corresponding to **block3_conv3**, **block4_conv3**, and **block5_conv3**, respectively.

These layers are deliberately selected to capture hierarchical morphological information:

1) Shallow features (F_3)

F_3 retains relatively high spatial resolution and captures low-level visual cues such as cytoplasmic granularity, cell boundary sharpness, and fine texture variations. These features are critical for identifying subtle differences in cytoplasmic appearance between benign cells and malignant lymphoblasts.

2) Intermediate features (F_4)

F_4 encodes mid-level structural information, including nucleus contour irregularity, chromatin condensation patterns, and nucleus-to-cytoplasm ratio. These morphological parameters are of great value to discriminate among ALL subtypes, particularly in differentiating Early Pre-B, Pre-B, and Pro-B lymphoblasts.

3) Deep features (F_5)

F_5 represents high-level semantic abstractions related to cell subtype identity and malignancy patterns. While these features are effective in capturing class-level characteristics, they often lack fine spatial detail when used alone.

By explicitly extracting features from these three levels, the proposed framework preserves both local morphological detail and global semantic context, which is essential for accurate blood cell cancer classification.

2.4.2. Spatial alignment and upsampling

Due to the pooling operations in VGG16, the spatial resolutions of F_3, F_4, and F_5 are different. To enable effective fusion, all feature maps are spatially aligned to a common resolution. Let the target resolution be that of F_3, that is, $H_3 \times W_3$. An upsampling operator $U_s(\cdot)$based on bilinear interpolation is applied:

$$\widehat{F}_4 = U_2(F_4), \widehat{F}_5 = U_4(F_5) \tag{6}$$

where the scaling factors 2 and 4 correspond to the relative downsampling ratios introduced by max-pooling layers. Bilinear interpolation is defined as:

$$\widehat{F}(x,y) = \sum_{i=0}^{1} \sum_{j=0}^{1} w_{ij} F(x_i, y_j) \tag{7}$$

where w_{ij} are the interpolation weights and (x_i, y_j) are the neighboring pixel coordinates. This operation ensures that all feature maps share identical spatial dimensions, enabling pixel-wise correspondence across feature levels.

2.4.3. Multiscale feature fusion strategy

Once spatially aligned, the feature maps are fused through channel-wise concatenation:

$$F_{fusion} = \text{Concat}(F_3, \widehat{F}_4, \widehat{F}_5) \tag{8}$$

resulting in a multiscale representation:

$$F_{fusion} \in \mathbb{R}^{H_3 \times W_3 \times (C_3 + C_4 + C_5)} \tag{9}$$

where C_3, C_4, and C_5 denote the number of channels in F_3, F_4, and F_5, respectively.

Concatenation is deliberately chosen instead of summation to preserve the complementary nature of features across different scales without enforcing feature-level alignment. This allows the model to learn optimal combinations of local texture, structural morphology, and semantic cues during subsequent training.

2.4.4. Global feature aggregation and classification

The fused feature map F_{fusion} is then passed through a Global Average Pooling (GAP) layer:

$$F_{gap} = \frac{1}{H_3 W_3} \sum_{x=1}^{H_3} \sum_{y=1}^{W_3} F_{fusion}(x,y) \tag{10}$$

This operation reduces spatial dimensionality, mitigates overfitting, and enforces spatial invariance while retaining discriminative information.

Finally, a fully connected layer followed by a softmax activation function is applied to obtain class probabilities:

$$\hat{y} = \text{Softmax}\left(WF_{gap} + b\right) \tag{11}$$

where W and b represent the learnable weights and bias parameters.

The output $\hat{y}$ corresponds to the probability distribution over the four classes: benign, malignant Early Pre-B, malignant Pre-B, and malignant Pro-B.

2.4.5. Design rationale and clinical interpretation

The design of the proposed VGG16-MFF framework is motivated by the hierarchical nature of morphological features in blood cell images. ALL subtypes often exhibit extremely subtle visual differences, which may manifest at different spatial scales. Shallow features capture cytoplasmic texture and edge patterns, intermediate features encode nucleus structure and chromatin organization, and deep features provide semantic abstraction of malignancy.

The proposed model integrates these three feature levels, allowing it to simultaneously attend to fine-grained cellular details and high-level diagnostic patterns. Such a multiscale representation is particularly useful in the context of ALL classification, where subtle morphological cues can be missed and lead to misclassification. Hence, the fusion strategy enhances the discriminative ability without too much architectural complexity.

The overall architecture of the proposed VGG16-MFF model is summarized in Figure 3.

2.5. Performance evaluation

We often use a confusion matrix to evaluate the classification models. A confusion matrix is just an easy way to see the results of a prediction. It provides a straightforward way to evaluate how well a model performs by breaking down the predictions into true positives, false positives, true negatives, and false negatives for each class. This matrix can be used to calculate a number of commonly used metrics including accuracy, precision, recall, and F1-score. Accuracy = correct predictions total predictions. Recall measures how well the model finds all relevant positive cases. Precision measures how precise the positive predictions are. In situations where the distribution of classes is uneven, or where it is important to minimize both false positives and false negatives, the F1-score offers a balance between precision and recall [35, 36].

These metrics are widely adopted in evaluating the performance of deep learning models across various image classification tasks. They offer complementary viewpoints on the accuracy, completeness, and balance of the model's performance. For example, a model may have high accuracy but low precision or recall, which may indicate poor class-level performance or bias. So, relying on one statistic may be misleading, especially when dealing with multiclass or imbalanced datasets. As pointed out by Kaushik and Khurana [37], a complete evaluation using all four metrics ensures that the model is not only accurate on average but also consistent across all prediction categories. This comprehensive evaluation framework enables a more reliable comparison of model performance and guides future model enhancements.

3. Results and Discussion

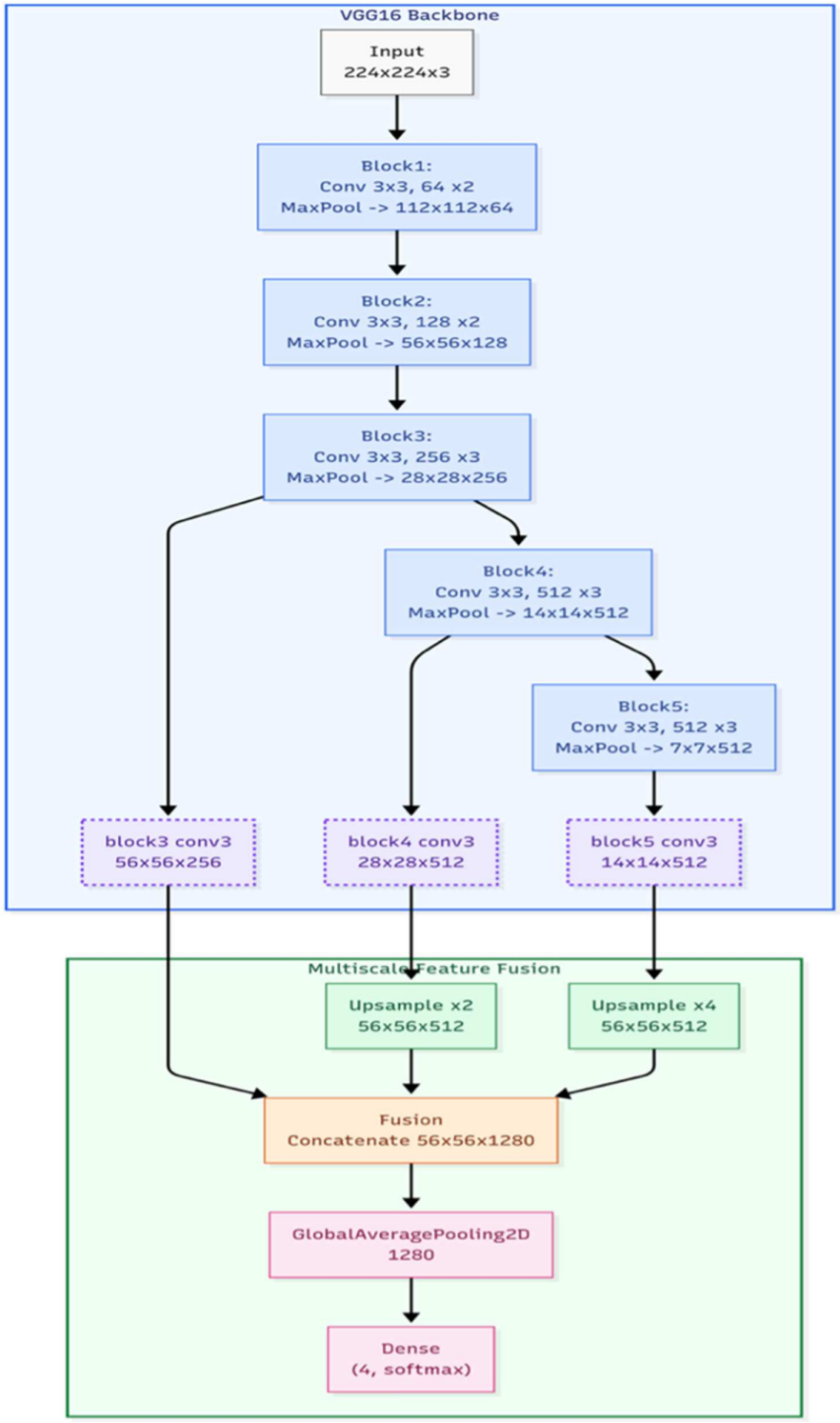
The experiments were conducted on Google Colab using Python 3 runtime with NVIDIA A100 GPU acceleration. The dataset was randomly split into training, validation, and testing sets at a stratified ratio of 80:10:10 to preserve the class distribution. All models are trained for 10 epochs with a batch size of 16 and an input resolution of 224 × 224 × 3. To make a fair comparison, the same preprocessing and training protocols were used for all architectures. In this section, we evaluate the effectiveness of the proposed multiscale features fusion design via baseline comparison and structured ablation analysis as shown in Table 1.

From Table 1, several important observations can be drawn. Deep residual learning is less successful in capturing fine-grained cellular morphology in blood cell cancer photos, as evidenced by ResNet50's lowest accuracy of 50.00%. Even though ResNet101 performs better (76.85%), it still performs worse than simple convolutional backbones, indicating that very deep residual structures might not be the best option for simulating minute cellular patterns.

However, VGG16 and VGG19 give a better performance with 88.27% and 87.34% accuracies, respectively. This indicates that uniform convolutional architectures are more suitable for microscopic medical images, where subtle texture and shape variations play a dominant role in the classification process.

The state of the art in modern design principles inspired by transformer architecture is ConvNeXt Tiny with an

Figure 3
Proposed architecture of VGG16-based multiscale feature fusion



accuracy of 84.26%, which is worse than VGG16. This implies that transformer-inspired convolutional designs do not necessarily guarantee better performance on small-scale medical datasets with relatively homogeneous visual patterns.

The ablation study further highlights the importance of deep and mid-level feature integration. Using only Deep features (***VGG16 + F5***) already provides a substantial performance gain (99.38%), demonstrating the strong semantic representation capability of high-level layers. When mid-level features are incorporated (***VGG16 + F4 + F5***), the accuracy further improves to 99.69%, indicating that intermediate texture information complements deep semantic cues. Finally, the proposed VGG16-MFF model integrating shallow, mid-level, and deep features (***F3 + F4 + F5***) achieves the best overall performance, confirming the

Table 1
Performance comparison of baseline models and proposed method

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Time (ms/step)
ResNet50	50.00	70.95	50.00	41.54	300
ResNet101	76.85	83.73	76.85	77.18	424
VGG16	88.27	90.35	88.27	88.72	175
VGG19	87.34	86.47	87.34	86.67	182
ConvNeXT Tiny	84.26	85.11	84.25	83.79	410
VGG16 + F_5	99.38	99.39	99.38	99.38	289
VGG16 + F_4 + F_5	99.69	99.69	99.69	99.69	276
Proposed model/VGG16-MFF (F_3 + F_4 + F_5)	99.69	99.69	99.69	99.69	191

effectiveness of MFF across hierarchical feature levels. An ablation study has therefore been conducted to evaluate the contribution of each feature level. It should be noted that the current ablation focuses on feature-level contribution rather than comparing different fusion strategies, which remains a direction for future work.

Crucially, this performance gain is attained at a relatively efficient computational cost of 191 ms/step, which is only slightly higher than the baseline VGG16 (175 ms/step). This efficiency is achieved because the proposed MFF module relies on channel-wise concatenation, which introduces minimal computational overhead compared to attention-based or multi-branch fusion mechanisms. In addition, the model does not significantly increase network depth. All experiments were conducted using an NVIDIA A100 GPU with a batch size of 16, contributing to stable and efficient training performance.

The training history is depicted in Figure 4, showing the learning curves for accuracy and loss.

The picture illustrates how accuracy converged to almost 100% while training and validation loss dropped quickly over the first few epochs and stabilized at values close to zero. The close alignment of training and validation curves indicates strong generalization and the absence of overfitting. This behavior contrasts with the baseline ResNet architectures, which exhibited slower convergence and a wider gap between training and validation performance. The proposed model thus demonstrates faster convergence and better stability during training.

The detailed classification performance per class is summarized in Table 2, while the corresponding confusion matrix is shown in Figure 5.

The proposed model is robust as shown in the confusion matrix and classification report. All malignant subtypes were classified with precision, recall, and F1-scores of 1.000. The only misclassification was in the benign class with a recall of 0.981, where one sample was classified as malignant. This type of error is safer clinically than missing malignant cases, but it is not ideal. The confusion matrix demonstrates the excellent sensitivity of the model in identifying malignant cells by verifying that nearly all samples were positioned appropriately.

In general, the results show that the integration of MFF in the VGG16 backbone can achieve significant improvement in classification accuracy while keeping computational economy. The combination of shallow and deep features provides complementary information that improves the model's discriminative capability, and hence, the model performs better than basic VGG and ResNet models. The results show the capability of the presented method to perform accurate medical image classification and its potential application in diagnostic support systems.

Figure 4
Training and validation accuracy and loss curves of the proposed VGG16-MFF model

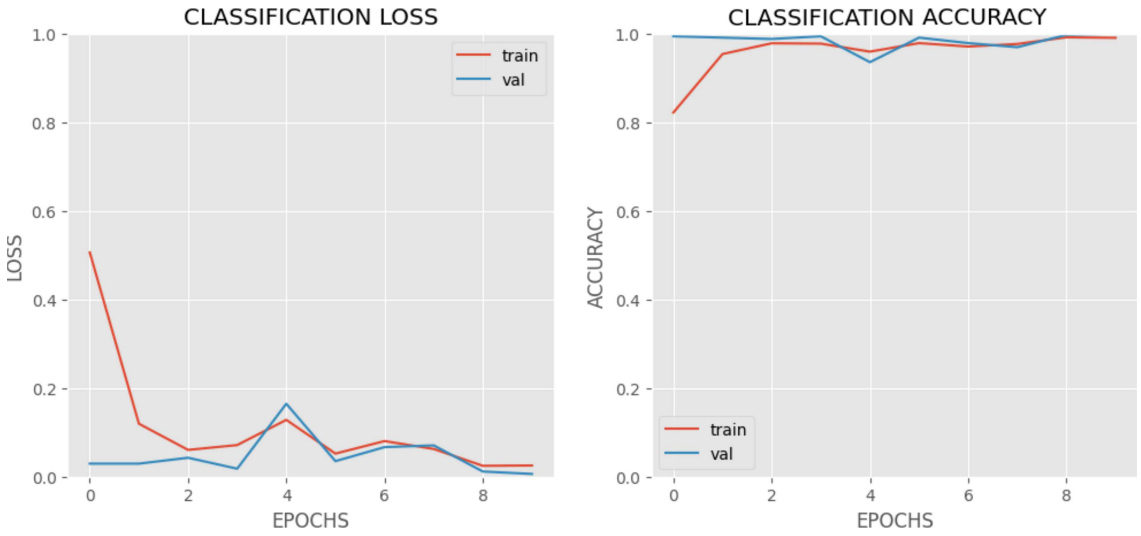


Table 2
Classification report of the proposed VGG16-MFF model

Class	Precision	Recall	F1-score	Support
Benign	1.000	0.981	0.990	52
[Malignant] Pre-B	1.000	1.000	1.000	95
[Malignant] Pro-B	1.000	1.000	1.000	79
[Malignant] early Pre-B	0.989	1.000	0.995	98
Overall accuracy	0.997			324

Figure 5
Confusion matrix of the proposed VGG16-MFF model

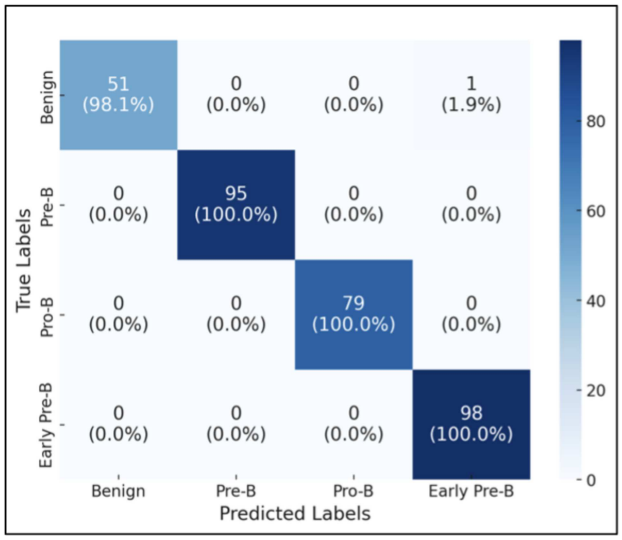


Table 3 is the summary of previous research for leukemia and blood cell cancer classification with CNN. As shown, most of the prior studies report lower accuracies than 95%, especially in multiclass settings, underlining the still ongoing challenge of distinguishing visually identical leukemic subtypes. These results suggest that conventional single-scale CNN representations are sometimes not able to accommodate small morphological variations. Ahad et al. [8] applied transfer learning with VGG19 and achieved only 72.29% accuracy, underscoring the limitations of conventional feature fusion. Similarly, Hidayat et al. [9] reported a baseline CNN accuracy of 85.32% that only slightly improved to 85.43% after incorporating XGBoost, demonstrating marginal gains despite hybridization. Other CNN-based works, such as those by Pooja et al. [10] and Rahman et al. [11], achieved accuracies between 85% and 88%, or below 90% in multiclass settings, highlighting the persistent difficulty of separating visually similar leukemia subtypes. Shourie et al. [12] added that the issue was made worse by class imbalance, which kept total performance below 90%.

Even more advanced strategies, including Imane et al. [13] for early-stage leukemia, Claro et al. [14] with ensemble CNNs, and Hamdi et al. [15] with hierarchical CNNs, produced improved but still imperfect outcomes, generally ranging from 90% to 94%. Claro et al. [14] reported accuracies of 94.73% and 94.59% with multilevel CNNs, but the approach heavily relied on data augmentation. Hamdi et al. [15] achieved around 90% using hierarchical CNNs, though performance degraded in real-world noisy datasets. Likewise, Al-Bashir et al. [16] found that several well-known CNN models, including AlexNet, ResNet, DenseNet, and VGG16, often produced validation accuracies near 94%, but test accuracies commonly dropped below 90% in multiclass scenarios.

Table 3
Summary of existing studies on blood cell cancer classification

Study	Dataset	Model	Classes	Accuracy (%)	Notes
Ahad et al. [8]	Leukemia dataset	VGG19 (TL)	Binary	72.29	Limited by transfer learning
Hidayat et al. [9]	Leukemia dataset	CNN + XGBoost	Binary	85.43	Marginal improvement after optimization
Pooja and Megha [10]	Bone marrow (ALL, MM)	CNN	Binary	85–88	Variation by subtype
Rahman et al. [11]	Multiclass leukemia	Deep CNN + FS	3–4	<90	Subtype separation remains difficult
Shourie et al. [12]	Leukemia subtypes	CNN	Multi	<90	Accuracy limited by imbalance
Imane et al. [13]	WBC images	CNN	Binary	<90	Early diagnosis challenge
Claro et al. [14]	4-class leukemia	Ensemble CNN	Multi	94.7	Relies on augmentation
Hamdi et al. [15]	Leukemia subtypes	Hierarchical CNN	Multi	~90	Drops on noisy real data
Al-Bashir et al. [16]	Mixed leukemia	AlexNet, ResNet, VGG, DenseNet	Multi	~90–94	Test acc <90 on multiclass

The proposed VGG16-MFF model demonstrated the potential of MFF for improving robustness in ALL classification on the Blood Cell Cancer (ALL) dataset. Unlike other methods that struggled with class imbalance, minor morphological changes, or dependence on augmentation, the network was able to capture both fine-grained local structures and more general contextual data by including MFF into the VGG16 backbone. This resulted in nearly perfect categorization between benign and malignant subtypes, demonstrating that the proposed design is a more reliable and long-lasting technique for blood cell cancer detection.

Although the performance is promising, this study has a limitation that a single dataset from one institution with relatively homogeneous conditions including consistent magnification and imaging protocols was used. These conditions may limit the generalizability of the model to real-world clinical settings, where variability in imaging devices, staining procedures, and patient demographics is common. Future work should therefore focus on validating the proposed approach on multi-institutional and more diverse datasets.

4. Conclusion

In this study, we proposed a VGG16-based MFF model, which greatly enhances the performance of blood cell cancer classification. The model consolidated shallow and deep features in a compact multiscale representation that effectively captured morphological details and semantic patterns that conventional CNNs often fail to distinguish. The experimental results show a significant enhancement. The baseline models: ResNet50, ResNet101, VGG16, and VGG19 attained 50.00%, 76.85%, 88.27%, and 87.34% accuracy, respectively, while the suggested VGG16-MFF achieved 99.69% in terms of accuracy, precision, recall, and F1-score. This near-perfect performance was achieved with only a small computational overhead (191 ms/step), demonstrating that MFF can greatly increase discriminative power without sacrificing efficiency. Importantly, the model was able to correctly identify all malignant subtypes and only misidentified one benign case, demonstrating its potential to be a very reliable diagnostic tool in a clinical setting. The results demonstrated that VGG16-MFF is an extraordinary methodological breakthrough and can be utilized as a reliable benchmark for blood cell cancer categorization studies.

Despite promising performance, this work is limited by the use of a single dataset collected at a single institution and under relatively homogeneous conditions. Such limitation may restrict the model's generalizability to real-world clinical settings. The model should be validated for this purpose by future studies on larger, more diverse, and multi-institutional medical imaging datasets. Additionally, future work will investigate lightweight model optimization techniques, such as pruning, quantization, and knowledge distillation, to enable deployment in resource-constrained environments, including point-of-care devices. Furthermore, future work will explore explainable AI techniques, such as Grad-CAM or LIME, to enhance model interpretability and provide visual insights into the decision-making process, which is essential for clinical adoption.

Ethical Statement

The authors declare that this study did not require formal ethical approval because it was based solely on secondary analysis of a publicly available dataset. The study did not involve direct interaction with human participants or animals, did not collect new biological samples, and did not use personally identifiable information. The dataset used in this study is openly accessible through Kaggle, and the analysis was conducted only on the publicly available image data provided by the dataset source. Therefore, formal informed consent was not required for this study.

Conflicts of Interest

The author declares that he has no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available in Kaggle at https://www.kaggle.com/datasets/mohammadamireshraghi/blood-cell-cancer-all-4class.

Author Contribution Statement

Simeon Yuda Prasetyo: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration.

References

[1] Yao, X., Wang, H., & Yang, L. (2026). Global, regional, and national burden of leukemia (1990–2021): A systematic analysis for the Global Burden of Disease Study 2021. *Acta Haematologica*, *149*(2), 137–154. https://doi.org/10.1159/000545724

[2] Yun, Z., Zhang, Y., Wang, J., Xue, C., Wu, J., Han, X., . . . , & Hou, L. (2024). Global, regional and national burden of leukemia, 1990-2021: Analysis of data from the Global Burden of Disease Study 2021. *Blood*, *144*(Supplement 1), 7907–7908. https://doi.org/10.1182/blood-2024-205478

[3] Chen, J., Pang, W., Deng, M., Zheng, R., Chen, Y., Zhang, Z., . . . , & Bai, Z. (2025). Global, regional, and national burden of leukemia, 1990–2021: A systematic analysis of the global burden of disease in 2021. *Frontiers in Medicine*, *12*, 1542317. https://doi.org/10.3389/fmed.2025.1542317

[4] Lan, X., Wu, J., Liao, Z., Wu, Y., & Hu, R. (2023). Association between maternal pesticide exposure and childhood acute lymphoblastic leukemia: A systematic review and meta-analysis. *BMC Cancer*, *23*, 1055. https://doi.org/10.1186/s12885-023-11581-z

[5] Sharma, R., & Jani, C. (2022). Mapping incidence and mortality of leukemia and its subtypes in 21 world regions in last three decades and projections to 2030. *Annals of Hematology*, *101*(7), 1523–1534. https://doi.org/10.1007/s00277-022-04843-6

[6] Daltveit, D. S., Morgan, E., Colombet, M., Steliarova-Foucher, E., Bendahhou, K., Marcos-Gragera, R., . . . , & Soerjomataram, I. (2025). Global patterns of leukemia by subtype, age, and sex in 185 countries in 2022. *Leukemia*, *39*(2), 412–419. https://doi.org/10.1038/s41375-024-02452-y

[7] Yang, F., Zhang, B., Lodder, P., & Guo, J. (2024). The burden of acute lymphoid leukemia among adolescents and young adults in the Western Pacific Region: Evidence from Global Burden Disease 2019. *Cancer Causes & Control*, *35*(5), 839–848. https://doi.org/10.1007/s10552-023-01843-3

[8] Ahad, M. T., Payel, I. J., Song, B., & Li, Y. (2024). DVS: Blood cancer detection using novel CNN-based ensemble approach. *arXiv Preprint: 2410.05272*.

[9] Hidayat, T., Hadinata, E., Damanik, I. S., Vikki, Z., & Irvanizam, I. (2023). An implementation of hybrid CNN-XGBoost method for leukemia detection problem. *Infolitika Journal of Data Science*, *1*(1), 15–21. https://doi.org/10.60084/ijds.v1i1.87

[10] Pooja, S., & Megha, G. S. (2021). Using CNN to detect cancerous cells in white blood cells from bone marrow through microscopic images. *Journal of Advancement in Parallel Computing*, *4*(2), 1–7.

[11] Rahman, W., Faruque, M. G. G., Roksana, K., Sadi, A. H. M. S., Rahman, M. M., & Azad, M. M. (2023). Multiclass blood cancer classification using deep CNN with optimized features. *Array*, *18*, 100292. https://doi.org/10.1016/j.array.2023.100292

[12] Shourie, P., Anand, V., Chauhan, R., Pokhariya, H. S., & Gupta, S. (2023). Exploring CNNs for blood subtype recognition: Binary and multi-class image classification. In *2023 3rd International Conference on Smart Generation Computing, Communication and Networking*, 1–6. https://doi.org/10.1109/SMARTGENCON60755.2023.10442491

[13] Imane, K., Abbas, M., Miloudi, A., & Meftah, M. C. E. (2022). A CNN model for early leukemia diagnosis. *International Journal of Organizational and Collective Intelligence*, *12*(1), 1–20. https://doi.org/10.4018/IJOCI.304889

[14] Claro, M. L., Veras, R. D. M. S., Santana, A. M., Vogado, L. H. S., Braz Junior, G., Medeiros, F. N. S. D., & Tavares, J. M. R. S. (2022). Assessing the impact of data augmentation and a combination of CNNs on leukemia classification. *Information Sciences*, *609*, 1010–1029. https://doi.org/10.1016/j.ins.2022.07.059

[15] Hamdi, I., El-Gendy, H., Sharshar, A., Saeed, M., Ridzuan, M., Hashmi, S. K., . . . , & Yaqub, M. (2024). Breaking down the hierarchy: A new approach to leukemia classification. In *Applications of Medical Artificial Intelligence: Second International Workshop* (pp. 104–113). https://doi.org/10.1007/978-3-031-47076-9_11

[16] Al-Bashir, A. K., Khnouf, R. E., & Bany Issa, L. R. (2024). Leukemia classification using different CNN-based algorithms-comparative study. *Neural Computing and Applications*, *36*(16), 9313–9328. https://doi.org/10.1007/s00521-024-09554-9

[17] Gao, H., Zhang, Y., Chen, Z., & Li, C. (2021). A multiscale dual-branch feature fusion and attention network for hyperspectral images classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 8180–8192. https://doi.org/10.1109/JSTARS.2021.3103176

[18] Yang, J., Wu, C., Du, B., & Zhang, L. (2021). Enhanced multiscale feature fusion network for HSI classification. *IEEE Transactions on Geoscience and Remote Sensing*, *59*(12), 10328–10347. https://doi.org/10.1109/TGRS.2020.3046757

[19] Qing, Y., Huang, Q., Feng, L., Qi, Y., & Liu, W. (2022). Multiscale feature fusion network incorporating 3D self-attention for hyperspectral image classification. *Remote Sensing*, *14*(3), 742. https://doi.org/10.3390/rs14030742

[20] Gao, J., Huang, Q., Gao, Z., & Chen, S. (2022). Image segmentation of retinal blood vessels based on dual-attention multiscale feature fusion. *Computational and Mathematical Methods in Medicine*, *2022*(*1*), 8111883. https://doi.org/10.1155/2022/8111883

[21] Wang, G., Yan, P., Tang, Q., Yang, L., & Chen, J. (2023). Multiscale feature fusion for skin lesion classification. *BioMed Research International*, *2023*(*1*), 5146543. https://doi.org/10.1155/2023/5146543

[22] Pan, H., Zhu, Y., Ge, H., Liu, M., & Shi, C. (2023). Multiscale cross-fusion network for hyperspectral image classification. *The Egyptian Journal of Remote Sensing and Space Sciences*, *26*(3), 839–850. https://doi.org/10.1016/j.ejrs.2023.09.002

[23] Liu, Y., Wang, L., Cheng, J., & Chen, X. (2021). Multiscale feature interactive network for multifocus image fusion. *IEEE Transactions on Instrumentation and Measurement*, *70*, 5019316. https://doi.org/10.1109/TIM.2021.3124058

[24] Hosseini, A., Eshraghi, M. A., Taami, T., Sadeghsalehi, H., Hoseinzadeh, Z., Ghaderzadeh, M., & Rafiee, M. (2023). A mobile application based on efficient lightweight CNN model for classification of B-ALL cancer from non-cancerous cells: A design and implementation study. *Informatics in Medicine Unlocked*, *39*, 101244. https://doi.org/10.1016/j.imu.2023.101244

[25] Awasthi, S., Kapoor, E., Srivastava, A. P., & Sanyal, G. (2020). A new Alzheimer's Disease classification technique from brain MRI images. In *2020 International Conference on Computation, Automation and Knowledge Management*, 515–520. https://doi.org/10.1109/ICCAKM46823.2020.9051485

[26] Muhammed, S., Upadhya, J., Poudel, S., Hasan, M., Donthula, K., Vargas, J., . . . , & Poudel, K. (2023). Improved classification of Alzheimer's disease with convolutional neural networks. In *2023 IEEE Signal Processing in Medicine and Biology Symposium* (pp. 1–7). https://doi.org/10.1109/SPMB59478.2023.10372725

[27] Ankita, & Mittal, S. (2024). Image classification of satellite using VGG16 model. In *2024 2nd International Conference on Disruptive Technologies*, 401–404. https://doi.org/10.1109/ICDT61202.2024.10489685

[28] Aksoy, S. (2025). Brain tumor detection and classification via VGG16-based deep learning on MRI imaging. *International Journal for Multidisciplinary Research*, *7*(2), 1–15.

[29] Navale, V. S. (2023). Bone abnormalities detection and classification using deep learning-VGG16 algorithm. *International Journal for Research in Applied Science and Engineering Technology*, *11*(7), 122–129. https://doi.org/10.22214/ijraset.2023.54582

[30] Ratnakar, A. (2024). Deep learning for early detection of diabetic retinopathy: A Colour Fundus image-based approach with VGG16 and VGG19. *International Journal of Scientific Research in Engineering and Management*, *8*(7), 1–10. https://doi.org/10.55041/IJSREM36355

[31] Prasetyo, S. Y., Alam, I. N., Rachmawati, I. D. A., Purwanto, E. S., & Ngarianto, H. (2023). VGG-powered convolutional neural networks in diabetic retinopathy classification: A comparative investigation. In *2023 3rd International Conference on Smart Cities, Automation & Intelligent Computing Systems*, 26–31. https://doi.org/10.1109/ICON-SONICS59898.2023.10435080

[32] Mouri Zadeh Khaki, A., & Choi, A. (2025). Optimizing deep learning acceleration on FPGA for real-time and resource-efficient image classification. *Applied Sciences*, *15*(1), 422. https://doi.org/10.3390/app15010422

[33] Liu, S., Zhang, Q., & Huang, L. (2023). Graphic image classification method based on an attention mechanism and fusion

of multilevel and multiscale deep features. *Computer Communications*, *209*, 230–238. https://doi.org/10.1016/j.comcom.2023.07.001

[34] Ali, A., Mu, C., Wang, Y., & Liu, Y. (2024). L-MFFN: Lightweight multiscale feature fusion network with limited samples for HSI classification. In *2024 9th International Conference on Signal and Image Processing*, 569–574. https://doi.org/10.1109/ICSIP61881.2024.10671478

[35] Park, W., Huh, J.-K., & Lee, J.-H. (2023). Automated deep learning for classification of dental implant radiographs using a large multi-center dataset. *Scientific Reports*, *13*(1), 4862. https://doi.org/10.1038/s41598-023-32118-1

[36] Sharmily, R. R., Karthik, B., & Vijayan, T. (2024). Improved MRI based automatic brain tumor categorization employing deep learning techniques. In *2024 4th Asian Conference on Innovation in Technology*, 1–5. https://doi.org/10.1109/ASIANCON62057.2024.10837709

[37] Kaushik, P., & Khurana, S. (2025). Bone age classification using deep learning: A ResNet50 model approach. In *2025 Fourth International Conference on Power, Control and Computing Technologies*, 1–5. https://doi.org/10.1109/ICPC2T63847.2025.10958638

How to Cite: Prasetyo, S. Y. (2026). VGG16-Based Multiscale Feature Fusion for Robust Blood Cell Cancer Classification. *Artificial Intelligence and Applications*. https://doi.org/10.47852/bonviewAIA62027495