Attention Mechanism-Enhanced Deep CNN Architecture for Precise Multi-class Leukemia Classification

Tahsen Islam Sajon, Barsha Roy, Md. Farukuzzaman Faruk, Azmain Yakin Srizon, Shakil Mahmud Shuvo, Md. Al Mamun, Abu Sayeed, and S. M. Mahedy Hasan

Abstract Leukemia is a life-threatening condition affecting people globally, making accurate diagnosis crucial for timely intervention. Consequently, researchers have been exploring automated methods to enable prompt action. The classification of leukemia into multiple subtypes according to WHO standards presents a unique challenge. Unlike binary classification, interclass features are highly similar, leading to misclassification. Ergo, we employ attention mechanisms to tackle this problem. Our proposed deep learning architecture combines transfer learning with attention mechanisms to classify subtypes of leukemia accurately. Using a publicly available dataset of blood cell images that adhered to WHO standards, we illustrate the potency of our approach. Our DenseNet201 with CBAM model achieves a remarkable 99.85% overall accuracy without resorting to data augmentation, surpassing previous methods on this dataset and attaining state-of-the-art results compared to other leukemia literature. To interpret the model's decision-making process and evaluate the efficacy of the attention mechanism in identifying discriminating features, we showcase GradCAM images and intermediate layer feature maps generated from our custom CNN. The proposed approach enhances leukemia classification accuracy and efficiency, providing clinical decision-making insights.

Keywords Leukemia classification · CNN · Transfer learning · Attention mechanism · CBAM · Feature map

1 Introduction

Leukemia is a malignant neoplasm of the hematopoietic system which manifests in the bone marrow and bloodstream. The uncontrolled proliferation of white blood cells disrupts the normal formation of essential blood components such as platelets,

T. I. Sajon (✉) · B. Roy · Md. F. Faruk · A. Y. Srizon · S. M. Shuvo · Md. Al Mamun · A. Sayeed · S. M. M. Hasan
Department of Computer Science and Engineering, Rajshahi University of Engineering and Technology, Rajshahi, Bangladesh
e-mail: sajon.tahsen@gmail.com

M. S. Arefin et al. (eds.), *Proceedings of the 2nd International Conference on Big Data, IoT and Machine Learning*, Lecture Notes in Networks and Systems 867,
https://doi.org/10.1007/978-981-99-8937-9_24

red blood cells, and other vital cells, leading to the development of leukemia [1]. Leukemic cells can spread throughout the body and harm other organs and tissues. The American Cancer Society estimates around 59,610 new leukemia cases to be diagnosed in the USA in 2023 [2]. Leukemia is a complex disease with different subtypes, each with distinct characteristics and treatments. Acute Lymphoblastic Leukemia (ALL) is one of the most prevalent childhood cancers, accounting for approximately 75% of leukemia cases in children and about 25% of all pediatric malignancies. In contrast, ALL is relatively rare in adults, representing only about 20% of all adult leukemia cases [3]. The onset of ALL is insidious, with non-specific symptoms such as fever, fatigue, and anemia, which may be mistaken for other common illnesses [1]. Rapid screening and therapy are critical for improving the chances of a favorable outcome. Traditionally, diagnosis involves a combination of clinical, laboratory, and morphological criteria, including the evaluation of bone marrow and blood samples. However, manual examination of these samples is subjective, time-consuming, and may lead to inaccuracies in diagnosis [4]. Consequently, accurate, efficient, and automated diagnostic tools are required to aid in the early diagnosis of ALL and enhance its management.

Although automated systems have shown promise in aiding leukemia diagnosis, several limitations and challenges persist, such as reliance on the French-American-British (FAB) categorization method instead of the expert-preferred World Health Organization (WHO) categorization, and underutilization of attention mechanisms. Addressing these limitations, we propose a novel three-tier architecture. In the initial tier, high-level features are extracted from blood smear images using a pretrained network. The second tier leverages a Convolutional Block Attention Module (CBAM) [5] to enhance model performance by capturing both spatial and channel information. Finally, the last tier consists of the classification module. We believe we are the first to employ CBAM for ALL classification. Moreover, to enhance the interpretability of the model, we present class activation maps and intermediate layer outputs to better understand the features learned by the model. Our research is motivated to explore the application of the WHO classification system and attention mechanisms to improve the accuracy and interpretability of ALL classification.

2 Literature Review

The classification of ALL has been a topic of active research. Using the ISBI-2019 challenge dataset, Zakir et al. designed a Convolutional Neural Network (CNN) architecture based on attention mechanism. They used VGG16 with Efficient Channel Attention (ECA) to amplify and enhance the semantic features of ALL cells. Their model achieved 91.1% accuracy on the test set [1]. In their paper, Krzysztof et al. utilized MobileNetV2 to extract features from images and applied Decision Tree (DT), Random Forest (RF), and XGBoost (XGB) algorithms to classify the images. Tested on the publicly available ALL-IBD dataset, their model obtained an average accuracy of 97.4% [6]. Mustafa et al. used ten different CNN architectures to extract features and classify 3256 PBS images from 89 suspected patients. Of the architectures tested, DenseNet201 achieved the highest accuracy of 99.85% [7]. Using images from the American Society of Haematology (ASH), Anilkumar et al. developed LeukNet, a 5-layer CNN for the automatic classification of ALL cells. Initially, they used AlexNet for classification and achieved an accuracy of 94.12%. They then applied all the preprocessing techniques used in AlexNet to LeukNet and achieved the same accuracy of 94.12% [8]. Adnan et al. proposed the use of Multi-Attention EfficientNet models to differentiate between leukemic and healthy cells. They utilized EfficientNetV2S and EfficientNetB3 transfer learning architectures, incorporating a multi-attention module and a weighted attention average module. Their models attained 99.73% and 99.25% accuracy on the C-NMC-2019 dataset [9]. Niranjana et al. introduced a specialized CNN architecture called ALLNET, which, trained on the C-NMC-2019 dataset, obtained an accuracy of 95.54% [10].

3 Materials and Methods

3.1 Dataset Collection and Description

A publicly available ALL dataset that was categorized per WHO standards served as the basis for our analysis. Images for the dataset were produced by the bone marrow lab at Taleqani Hospital in Tehran and were meticulously categorized by a qualified professional. This dataset comprised 3256 Peripheral Blood Smear (PBS) images obtained from 89 people with a presumptive diagnosis of ALL, including 25 individuals who were found to be healthy (benign hematogones) and 64 individuals who were diagnosed with ALL [7]. Table 1 provides an overview of the dataset's characteristics.

Table 1 Dataset characteristics

Type	Subtype	Samples count	Patients count
Benign	Hematogones	504	25
	Total	504	25
Malignant	Pro-B ALL	804	23
	Pre-B ALL	963	21
	Early Pre-B ALL	985	20
	Total	2752	64
Grand total		3256	89

605982_1_En_24_Chapter-online ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:26/12/2023 Pages: 12 Layout: T1-Standard

3.2 Data Preprocessing

CNNs are capable of recognizing essential features in raw images, eliminating the need for extensive preprocessing. Nonetheless, specific preprocessing measures were necessary to enhance the training and diagnosis processes. To this end, the photos were resized to a uniform dimension of $224 \times 224 \times 3$, and the pixel values were normalized between [0, 1] before they were fed into the neural network.

Data augmentation methods are commonly employed to increase the number of training samples and minimize the risk of overfitting. However, it is worth noting that overusing augmentation may potentially obscure critical image features. Furthermore, our findings revealed that the exclusive implementation of CBAM was sufficient in achieving exceptional accuracy, rendering data augmentation unnecessary. Therefore, data augmentation techniques were deliberately omitted from our approach.

3.3 Convolutional Neural Network

A Convolutional Neural Network (CNN) is a deep learning architecture optimized for image processing tasks. It employs a hierarchical approach to extract the features, utilizing convolution, pooling, normalization and fully connected layers, to extract higher-level features from the input data gradually. The final classification layer is utilized to assign probabilities to the output classes and identify the most probable class. The CNN architecture is a powerful tool for image processing research, enabling the development of accurate and sophisticated models.

Transfer Learning. Transfer learning is a type of machine learning technique involving a pretrained model to perform a related but different task. The pretrained model has learned valuable features from a larger dataset like ImageNet, and the model can be fine-tuned on a different smaller dataset for better performance on the new specific task.

In this study, we investigated four distinct transfer-learned architectures, namely DenseNet201, ResNet50, EfficientNetB6, and Xception, to extensively assess our strategy. DenseNet201 utilizes dense connections to alleviate the problem of vanishing gradients, allowing for better feature reuse and optimization [11]. EfficientNet is a family of scalable CNNs that use compound scaling to balance depth, width, and resolution dimensions, and achieve high accuracy while minimizing computational cost [12]. ResNet50 incorporates residual connections, enabling the reuse of earlier feature maps, enhancing training and generalization performance. It comprises residual blocks with identity and projection shortcuts for matching feature map dimensions [13]. Xception uses depth-wise separable convolutions, which factorize spatial and channel-wise dimensions of the convolution separately, reducing computation and increasing model capacity [14].

3.4 Attention Mechanism

Attention mechanisms have emerged as a promising approach to improving deep learning models. This neural network component selectively focuses on regions of input images, capturing fine-grained details and contextually relevant features. By dynamically weighing the importance of different image regions, attention mechanisms have shown potential to enhance accuracy, robustness, and interpretability in image classification models [5]. The concept of attention mechanisms initially gained popularity in the domain of Natural Language Processing (NLP) through the work of Vaswani et al. [15]. There, the attention mechanism was computed using three primary components: Query, Key, and Value. This idea was later adapted and extended to computer vision by Zhang et al. [16], who introduced the Self Attention Module.

Convolutional Block Attention Module. The Convolutional Block Attention Module (CBAM) is a type of attention mechanism that leverages both spatial and channel attention mechanisms to selectively focus on salient image features while filtering out noise and irrelevant information. CBAM comprises two modules: CAM, or the Channel Attention Module, and SAM, or the Spatial Attention Module, and they have distinct roles. The CAM module produces a 1D attention map by taking the max-pooled and average-pooled values from the input feature map and applying two dense layers to obtain channel-wise attention weights. In contrast, the SAM module generates a 2D spatial attention map by computing maximum and average values across the channel dimension, concatenating them, and passing the result through a convolutional layer.

$$B_{\mathrm{c}}(A) = \sigma\left(\mathrm{MLP}\left(\mathrm{AvgPool}(A)\right) + \mathrm{MLP}\left(\mathrm{MaxPool}(A)\right)\right) \tag{1}$$

$$B_{\mathrm{s}}(A) = \sigma\left(a^{7\times 7}\left(\left[\mathrm{AvgPool}(A)\,;\,\mathrm{MaxPool}(A)\right]\right)\right) \tag{2}$$

Equation (1) is for channel attention and (2) is for spatial attention. The attention maps are multiplied element-wise with the input feature map, resulting in adaptive refinement of features in both the spatial and channel dimensions. A convolutional layer processes the refined feature map to capture these features, and the resulting output is added to the original feature map, generating the final output of the CBAM module. The overall attention mechanism for CBAM can be summarized as:

$$\acute{A} = B_c(A) \otimes A \tag{3}$$

$$\acute{\acute{A}} = B_s(\acute{A}) \otimes \acute{A} \tag{4}$$

From (3) and (4), $\acute{\acute{A}}$ represents the final output of CBAM [5]. Figure 1 depicts each layer in our implementation of the CBAM module. Here, the two lambda layers in the SAM calculate the maximum and average values in order to compute the spatial attention map.

605982_1_En_24_Chapter-online ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:26/12/2023 Pages: 12 Layout: T1-Standard

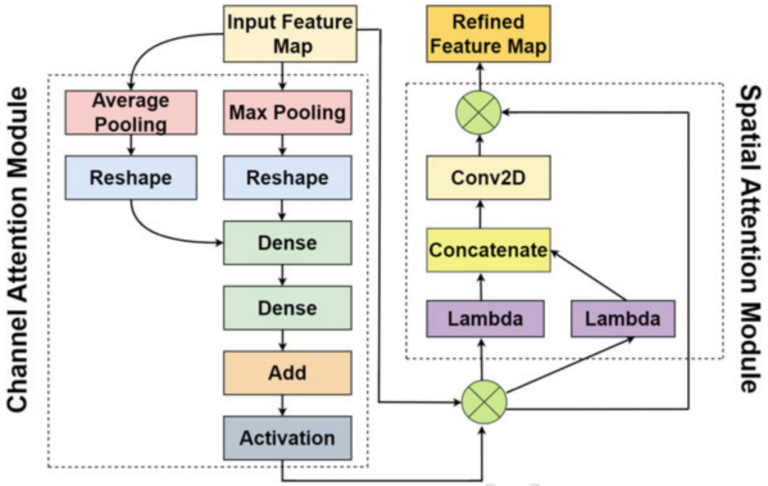

Fig. 1 Our implementation of the CBAM block as proposed in [5]

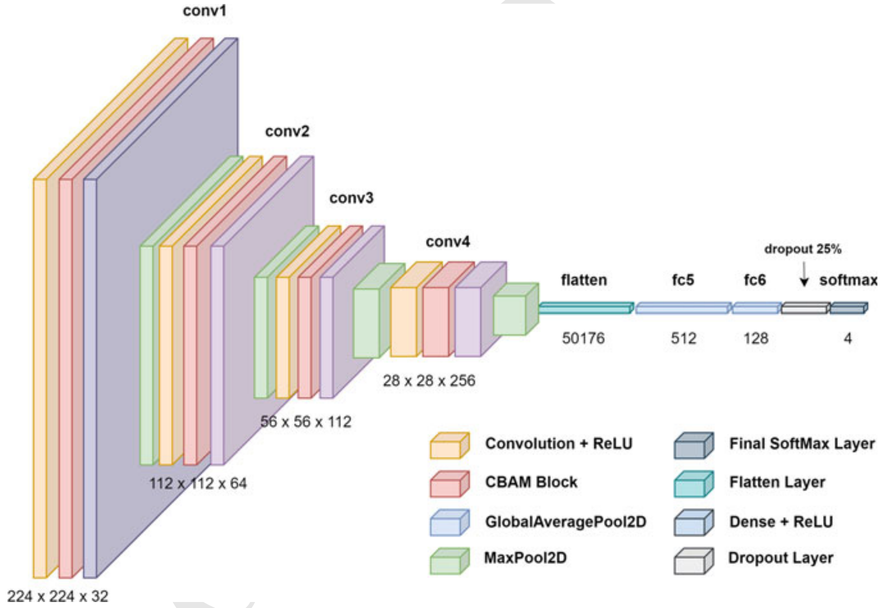

Fig. 2 Custom CNN with a CBAM layer following each convolution layer. Here, (con1-conv4) are the convolutional blocks, and (fc5, fc6) are the fully connected layers

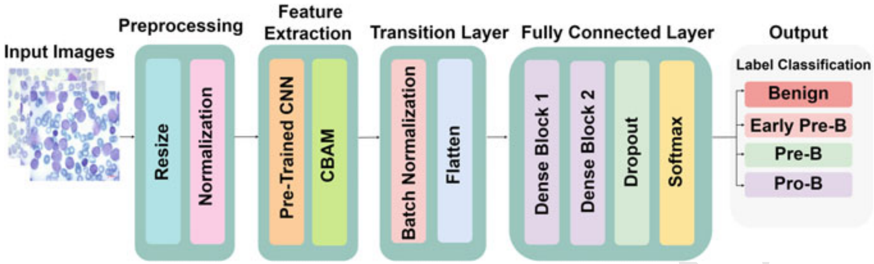

Fig. 3 Proposed transfer learning with CBAM attention architecture

3.5 Proposed Architecture

Our approach began with the implementation of a handcrafted CNN integrated with CBAM, as illustrated in Fig. 2. To enhance the intermediate feature maps obtained from the previous convolutional layer, we utilized CBAM as a layer in each convolutional block. This allowed us to refine the feature maps through the application of channel attention using CAM, followed by spatial attention using SAM. The resulting output feature maps were then used for subsequent processing.

To enhance the performance and robustness of our model, we adopted a transfer learning approach. Specifically, we utilized a pretrained CNN to extract features from the input data and obtained its outputs from the last convolutional layer. These outputs were then passed through a CBAM layer, which helped refine and highlight the most important features in the input. To further ameliorate the effectiveness of the network, we applied batch normalization to the output of the CBAM layer, which helps to improve the stability and speed of the training process. The normalized output was then flattened into a one-dimensional array. Next, we applied two dense layers with 512 and 128 nodes, respectively, with the activation function ReLu [17]. The purpose of these layers was to acquire high-level representations of the input features and further improve the discriminative capacity of the model. To prevent overfitting, we applied a 25% dropout after the second dense layer. Finally, the classification was performed through a softmax [18] layer, which allowed us to predict the class probabilities for the input image. This comprehensive pipeline of transfer learning, attention mechanisms, normalization, dense layers, and dropout helped to improve the model's performance and robustness. Figure 3 presents a overview of our proposed methodology.

Our research primarily aims to enhance the representation power of neural networks by incorporating attention mechanisms. The proposed methodology involves the utilization of two modules for attention-based feature refinement, namely channel and spatial. By integrating CBAM, we are able to efficiently modulate the flow of information inside the network by learning which features to prioritize and which to suppress. Our experimental findings demonstrate that our method offers significant performance gains while keeping computational overheads low.

4 Results and Performance Analysis

In our experimental setup, we trained the models for 75 epochs using a batch size of 24, until it was determined that the validation loss had essentially plateaued, with no further significant improvement in the remaining epochs. We employed Adam [19] as the optimizer, with a learning rate of 0.0001. We also made use of a callback to reduce the learning rate on a plateau. The categorical cross-entropy was selected as the loss function. Following preprocessing, the dataset was divided into train, validation, and test sets, with 60%, 20%, and 20% of the data, respectively, being assigned to each set.

Subsequently, to gauge the effectiveness of our proposed transfer learning strategy that integrates attention mechanisms, we evaluated a number of cutting-edge transfer learning architectures, including DenseNet201, EfficientNetB6, Xception, and ResNet50. All models showed improvements over the custom CNN, with our modified DenseNet201 achieving the highest accuracy of 99.85%. Figure 4 illustrates the training and validation accuracy and loss of the DenseNet201 model. The confusion matrices for both the custom CNN and the proposed DenseNet201 with CBAM model are depicted in Fig. 5. Table 2 provides a comprehensive summary of the performance of each model, including accuracy, precision, recall, F1-score, and support.

AQ2 Table 3 compares our proposed techniques with previous endeavors on the same dataset in terms of overall accuracy. Although [7] achieved commensurate outcomes with the DenseNet201 architecture, their other models exhibited significant variations in performance, with some achieving subpar results, indicating instability likely due to a lack of optimization. Our work addresses this issue, resulting in stable and consistent performance across our proposed models. Additionally, examination of the confusion matrix in Fig. 5 reveals that our DenseNet201 model made a single incorrect prediction, specifically for the Early Pre-B ALL class. In contrast, the previous highest-performing model, DenseNet201 from [7], had two erroneous predictions for the same class. Furthermore, we conducted a comparative evaluation of

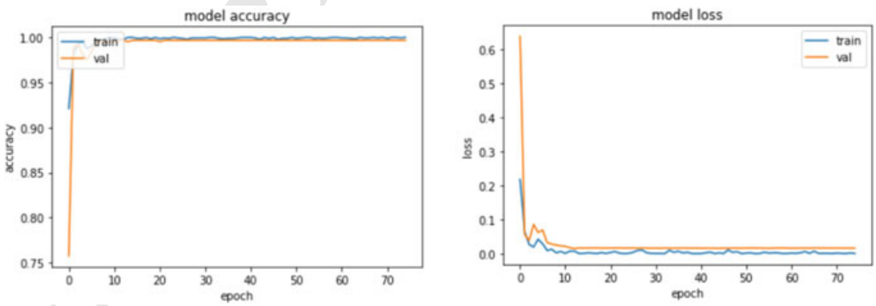

Fig. 4 Training and validation accuracy and loss of proposed DenseNet201 with CBAM model

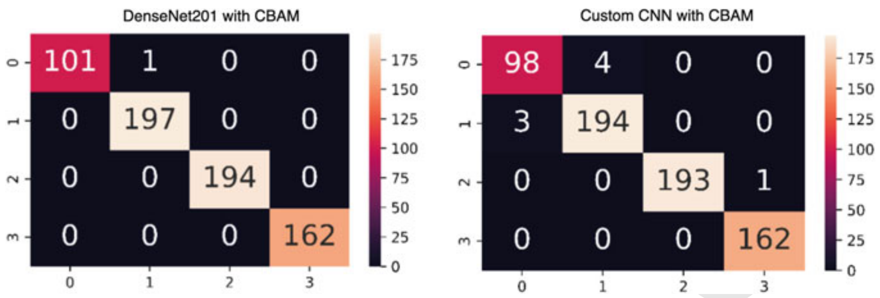

Fig. 5 Model performance assessment with confusion matrix

Table 2 Class-specific evaluation measures (accuracy, precision, recall, F1-score, and support) for each model using a scale ranging from 0.00 for 0% to 1.00 for 100%

Class	Accuracy	Precision	Recall	$F1$-Score	Support
Proposed DenseNet201 with CBAM					
Benign	0.990	1.00	0.99	1.00	102
Early	1.000	0.99	1.00	1.00	197
Pre-B	1.000	1.00	1.00	1.00	194
Pro-B	1.000	1.00	1.00	1.00	162
Proposed ResNet50 with CBAM					
Benign	0.990	1.00	0.99	1.00	102
Early	1.000	0.99	1.00	1.00	197
Pre-B	0.995	1.00	0.99	1.00	194
Pro-B	1.000	0.99	1.00	1.00	162
Proposed Xception with CBAM					
Benign	0.990	0.99	0.99	0.99	102
Early	1.000	1.00	1.00	1.00	197
Pre-B	0.990	0.99	0.99	0.99	194
Pro-B	0.994	0.99	0.99	0.99	162
Proposed EfficientNetB6 with CBAM					
Benign	0.990	0.97	0.99	0.98	102
Early	0.990	0.99	0.99	0.99	197
Pre-B	0.995	1.00	0.99	1.00	194
Pro-B	0.994	0.99	0.99	0.99	162
Custom CNN with CBAM					
Benign	0.961	0.97	0.96	0.97	102
Early	0.985	0.98	0.98	0.98	197
Pre-B	0.995	1.00	0.99	1.00	194
Pro-B	1.000	0.99	1.00	1.00	162

Table 3 Comparison between our proposal and notable previous works on the same dataset

Methods used	Data augmentation	Overall accuracy (%)
EfficientNet [7]	Yes	28.22
Xception [7]	Yes	96.70
ResNet50V2 [7]	Yes	97.85
NASNetLarge [7]	Yes	98.16
DenseNet201 [7]	Yes	99.85
Proposed EfficientNetB6 + CBAM	**No**	**99.24**
Proposed Xception + CBAM	**No**	**99.39**
Proposed ResNet50 + CBAM	**No**	**99.69**
Proposed DenseNet201 + CBAM	**No**	**99.85**

Table 4 Comparing proposed work with other literature

Dataset	Methods used	Overall accuracy (%)
ALL-IDB	MobileNetV2 + XGB,RF,DT [20]	97.40
ASH	LeukNet [8]	94.12
C-NMC-2019	VGG16 + ECA [1]	91.10
C-NMC-2019	EfficientNetV2s + Multi-Attention [9]	99.73
C-NMC-2019	ALLNET [10]	95.54
ALL dataset	Proposed DenseNet201 + CBAM	99.85

our proposed system against the relevant literature, and the findings are presented in Table 4. Notably, our suggested methods have outperformed previous efforts.

By evaluating multiple state-of-the-art transfer learning architectures, we were able to perform a comprehensive comparison and evaluation of our proposed transfer learning with attention architecture. This diverse selection of models allowed us to obtain a thorough understanding of the strengths and limitations of our strategy. As a result, we have reasonable grounds to draw informed conclusions about the efficacy of our methodology vis-á-vis existing techniques.

4.1 Model Interpretability: What Our CNN Sees

In this investigation, we offer two techniques to gain insight into model predictions: Class Activation Mapping and intermediate layer visualization.

The Gradient-weighted Class Activation Mapping (Grad-CAM) is a widely employed visualization technique utilized to comprehend the prominence of a specific class in a given image. Grad-CAM works by computing the gradients of the

605982_1_En_24_Chapter-online ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:26/12/2023 Pages: 12 Layout: T1-Standard

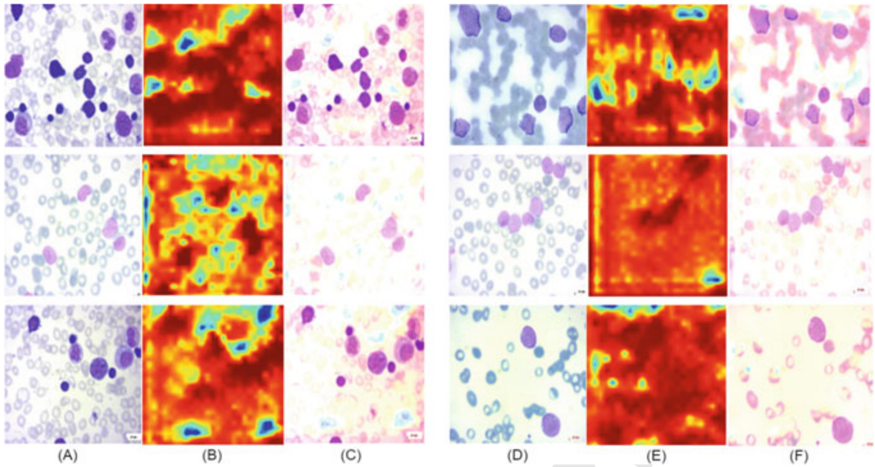

Fig. 6 Unpacking the complexity with Grad-CAM analysis: (**a**, **d**) are input samples, (**b**, **e**) are the respective class activation maps and in (**c**, **f**) activation maps superimposed on the samples provide a visual representation of the regions of interest

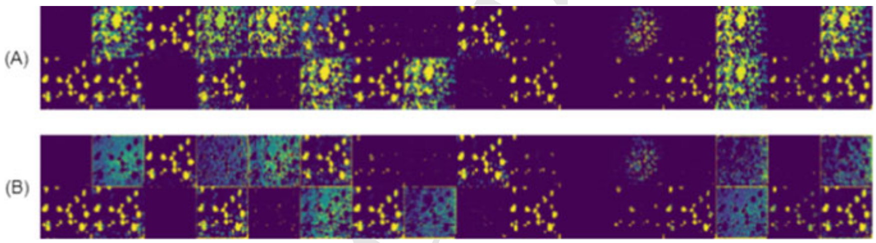

Fig. 7 Uncovering the impact of attention: **a** is the output from the first convolution layer, and **b** is the refined output achieved through subsequent CBAM attention layer

output class score relative to the feature maps of the final convolutional layer. The gradients are then globally averaged and weighted by the importance of each feature map. Lastly, the resulting weight map is multiplied with the feature map, which generates the Grad-CAM visualization [21]. We utilized the Grad-CAM technique on six randomly chosen input samples, displayed in Fig. 6, and the resulting heatmaps demonstrate the regions that the model used to make its predictions. Furthermore, as depicted in Fig. 7, we analyzed the 32 feature maps generated by the first convolutional layer and the CBAM layer immediately following it. Our observations reveal that CBAM effectively refines the feature maps and increases their discriminative power.

In sum, our proposed methods provide a deeper understanding of the model's decision-making process, and the insights gained can be leveraged to improve performance and interpretability.

605982_1_En_24_Chapter-online ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:26/12/2023 Pages: 12 Layout: T1-Standard

5 Conclusion

In this study, we introduce a new approach to automatic leukemia classification by incorporating transfer learning and attention mechanisms. Our research addresses two major gaps in the current literature: (1) the predominant use of the less-preferred FAB classification instead of the WHO system, and (2) the lack of attention mechanisms in prior studies.

Our approach consistently yielded promising results on a dataset of Acute Lymphoblastic Leukemia classified using the WHO methodology. Notably, our CNN architecture effectively refined the features of PBS images through the application of a CBAM module after the output of the pretrained network, thereby enhancing the discriminative ability of the model. Our optimization strategy involved techniques such as learning rate reduction, regularization, and dropout. Together, these approaches enabled us to achieve results that surpassed those of previous studies in the field, without using data augmentation.

Moreover, we provided insights into the interpretability of our proposed model. We presented Grad-CAM images and output feature maps to reveal the regions of interest in the input images and the effectiveness of the CBAM integration in refining the feature maps, respectively. Our model demonstrated not only superior classification performance but also high interpretability, which is crucial in medical diagnosis.

Moving forward, we plan to extend our study by exploring larger datasets, incorporating segmentation techniques, and investigating the potential of vision transformers in leukemia classification. We hope that our work will inspire further research in the field of automated medical diagnosis and contribute to the development of more precise and effective tools for diagnosing leukemia.

References

1. Zakir Ullah M, Zheng Y, Song J, Aslam S, Xu C, Kiazolu GD, Wang L (2021) An attention-based convolutional neural network for acute lymphoblastic leukemia classification. Appl Sci 11(22):10662
2. Society AC: American cancer society: cancer facts & statistics. https://cancerstatisticscenter.cancer.org/#!/cancer-site/Leukemia. Accessed on 7 Apr 2023
3. Das PK, Diya V, Meher S, Panda R, Abraham A (2022) A systematic review on recent advancements in deep and machine learning based detection and classification of acute lymphoblastic leukemia. IEEE Access
4. Sajon TI, Chowdhury M, Srizon AY, Faruk MF, Hasan SM, Sayeed A, Rahman AM (2023) Recognition of leukemia sub-types using transfer learning and extraction of distinguishable features using an effective machine learning approach. In: 2023 International conference on electrical, computer and communication engineering (ECCE). IEEE, pp 1–6
5. Woo S, Park J, Lee JY, Kweon IS (2018) Cbam: convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV), pp 3–19

6. Pałczyński K, Śmigiel S, Gackowska M, Ledziński D, Bujnowski S, Lutowski Z (2021) IoT application of transfer learning in hybrid artificial intelligence systems for acute lymphoblastic leukemia classification. Sensors 21(23):8025
7. Ghaderzadeh M, Aria M, Hosseini A, Asadi F, Bashash D, Abolghasemi H (2022) A fast and efficient CNN model for b-all diagnosis and its subtypes classification using peripheral blood smear images. Int J Intell Syst 37(8):5113–5133
8. Anilkumar K, Manoj V, Sagi T (2022) Automated detection of b cell and t cell acute lymphoblastic leukaemia using deep learning. Irbm 43(5):405–413
9. Saeed A, Shoukat S, Shehzad K, Ahmad I, Eshmawi A, Amin AH, Tag-Eldin E (2022) A deep learning-based approach for the diagnosis of acute lymphoblastic leukemia. Electronics 11(19):3168
10. Sampathila N, Chadaga K, Goswami N, Chadaga RP, Pandya M, Prabhu S, Bairy MG, Katta SS, Bhat D, Upadya SP (2022) Customized deep learning classifier for detection of acute lymphoblastic leukemia using blood smear images. In: Healthcare, vol 10. MDPI, p 1812
11. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4700–4708
12. Tan M, Le Q (2019) Efficientnet: rethinking model scaling for convolutional neural networks. In: International conference on machine learning. PMLR, pp 6105–6114
13. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
14. Chollet F (2017) Xception: deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1251–1258
15. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. Adv Neural Inf Process Syst 30
16. Zhang H, Goodfellow I, Metaxas D, Odena A (2019) Self-attention generative adversarial networks. In: International conference on machine learning. PMLR, pp 7354–7363
17. Nair V, Hinton GE (2010) Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th international conference on machine learning (ICML-10), pp 807–814
18. Ackley DH, Hinton GE, Sejnowski TJ (1985) A learning algorithm for Boltzmann machines. Cogn Sci 9(1):147–169
19. Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980
20. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC (2018) Mobilenetv2: inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4510–4520
21. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2017) Grad-cam: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision, pp 618–626

605982_1_En_24_Chapter-online ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:26/12/2023 Pages: 12 Layout: T1-Standard