



# Multi-Scale Cross Attention for Improved Bone Marrow Cell Classification

**Mrs. S. Surekha<sup>1</sup>, Parvatham Kavya<sup>2</sup>, M P Princy Veronica<sup>3</sup>, Kasabu Pavan<sup>4</sup>**

Assistant Professor<sup>1</sup>, Student/Research Scholar<sup>2</sup>, Student/Research Scholar<sup>3</sup>, Student/Research Scholar<sup>4</sup>

Department of Computer Science and Engineering (Artificial Intelligence and Machine Learning)

Vignana Bharathi Institute of Technology, Aushapur(V), Ghatkesar (M), Medchal Dist. 501301, Telangana, India

**Abstract:** Accurately classifying bone marrow cells is essential for identifying a variety of hematological illnesses in the realm of medical image analysis. However, traditional methods face challenges in distinguishing between different cell types due to their complex morphological features. This work will build a model based on multiscale convolutional neural network that can automatically extract discriminative features from the original photos in order to address these difficulties. In addition, this research presents a novel method to enhance the interpretability and performance of deep learning models for bone marrow cell categorization by using multi-scale cross attention mechanisms and visual attention networks. Results from experiments show that our bone marrow cell classification accuracy exceeds other deep learning techniques as well as conventional machine learning methods. Moreover, the technique demonstrates outstanding reliability and stability.

**Index Terms - Bone marrow cell classification, CNNs, Visual attention network, multi-scale cross attention, Medical Image analysis, Deep learning.**

## I. INTRODUCTION

White blood cells (leukocytes), red blood cells (erythrocytes), and platelets (thrombocytes) are the three major types of blood cells that are produced by hematopoietic stem cells, which are stored in bone marrow, a critical tissue found inside our bones [1]. Bone marrow cells must be precisely examined and classified in order to diagnose and treat blood-related diseases like leukemia. The analysis of bone marrow cell pictures under a microscope presents a challenging task due to the numerous alterations that transpire inside and among distinct cell types. Because of these complexities, conventional automated detection and classification algorithms often fail to adequately handle clinical applications, leading to limitations and erroneous findings.

To address these problems, this research offers a novel approach that makes use of multi-scale attention networks. These networks provide a potential mechanism for the fine-grained categorization of bone marrow cells by combining visual attention modules with cross-scale attention processes. The goal of this technology is to find the discriminative regions and document those tiny changes across several dimensions, which aims to revolutionize our understanding of bone marrow cell structures.

Inspired by human visual attention mechanisms, our research attempts to develop Visual Attention Networks (VANs) capable of identifying important regions within complex images. The proposed model incorporates a VAN, enabling it to selectively focus on relevant features within bone marrow cell images. This strategic focus facilitates the capture of subtle variations and distinctive patterns indicative of different cell types. The interpretability provided by VANs enhances the understanding of the classification process.

A multi-scale cross attention mechanism is implemented to further enhance performance. Contextual information can be incorporated at various scales with the help of this method. It aggregates both local and global dependencies, leading to more accurate classifications. As a result, our network attains improved accuracy in classifying by simultaneously examining features across various scales, minimizing occurrences of undesired visual disruptions, and ensuring overall stability. Thus, the paper bridges the gap between deep learning techniques and medical image analysis.

This approach is validated through extensive experimentation on bone marrow cell datasets, ensuring thorough evaluation of its effectiveness. The experimental findings reveal compelling advancements compared to prevailing methodologies, showcasing substantial improvements in accuracy and robustness. Remarkably, the model demonstrates high performance in precisely classifying a wide spectrum of cell types, spanning 21 distinct categories, including Abnormal eosinophil (ABE), Artefact (ART), Basophil (BAS), Blast (BLA), and more [2]. Notably, it achieves exceptional precision in identifying variations within cell populations, highlighting the model's effectiveness and adaptability in clinical contexts.

In summary, the research contributes to advancing medical image analysis by providing an innovative framework for bone marrow cell classification. By combining visual attention networks and multi-scale cross attention, it empowers clinicians with a precise tool for reliable diagnosis, ultimately improving patient care.

## II. RELATED WORK

### 2.1 Existing System

Let's examine some recent developments in models for classifying bone marrow cells. A number of innovative models have been presented recently to overcome the difficulties in classifying bone marrow cells. While these models have made significant strides in improving classification accuracy, they also exhibit certain limitations. Let us have a look at them:

A proposal called Dual Attention Gates DenseNet (DAGDNet) was made for the classification of bone marrow cells [3]. DenseNet's architecture includes a dual attention gates (DAGs) mechanism. To increase the precision and recall of neural network-based cell classifiers, the DAGs are employed to sort and highlight the position-related features in DenseNet. Despite its effectiveness in enhancing classification performance, DAGDNet may suffer from increased computational complexity due to the integration of dual attention gates. Additionally, its applicability to diverse datasets and cell types may be limited.

Self-supervised Attention Filtering and Multi-scale Features Network (SA-MFN), was proposed for fine-grained visual classification [4]. It consists of three modules: the Self-supervised Attention Map Filter, the Multi-scale Attention Map Generator, and the Reiterative Prediction [4]. It may require extensive hyperparameter tuning and large-scale labeled datasets for optimal performance. Moreover, its interpretability and generalizability across different medical imaging tasks remain to be fully explored.

Attention-based Suppression and Attention-based Enhancement Net (ASAE-Net) was designed for the fine-grained classification of bone marrow cells [5]. ASAE-Net achieves superior performance without increasing training parameters compared with the backbone. However, its reliance on attention mechanisms for feature suppression and enhancement may introduce additional computational overhead and complexity during training and inference.

### 2.2 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) have emerged as a leading solution to the fundamental difficulty of computer vision: efficiently computing powerful feature representations [6], [7]. Since the groundbreaking Alex Net, CNNs have revolutionized the effectiveness of neural networks by utilizing translation invariance features and local contextual information, hence becoming the standard framework in the field [8].

Additionally, CNNs have shown impressive performance in image classification applications by efficiently extracting features via pooling and convolution operations [9]. Even so, it has been shown that when fed into fully connected layers, trained feature maps in CNNs may occasionally lack the quality required for the best classification outcomes [9]. The suggested framework includes a visual attention network, which is intended to improve the fully connected layers of CNNs' classification capability, in order to overcome this drawback. The proposed visual attention mechanism aims to improve feature maps by emphasizing salient regions while suppressing irrelevant information [10], [11]. Improved classification performance can be attained by including this process into the fully connected layers [9].

Moreover, CNNs have been widely used for classification problems, frequently combining localization methods like localization and region proposal [12]. The proposed visual attention module enhances CNNs' capability in handling complex classification tasks, such as bone marrow cell classification, by improving feature representation quality and classification accuracy.

### 2.3 Multi-Scale Channel Attention Mechanism

In recent years, several studies [11], [13], [14], [15] have highlighted the effectiveness of embedding multi-channel attention blocks into existing Convolutional Neural Network (CNN) architectures. These attention mechanisms significantly enhance network performance by selectively emphasizing relevant features while suppressing less informative channels. In the context of bone marrow cell classification, where accurate identification of cell types is critical for medical diagnostics, such mechanisms play a pivotal role.

Let's understand the key components of the proposed Multi-Scale Channel Attention Mechanism: The SENet [10] introduced the concept of channel attention. It learns the dependencies between channels within a feature map. By automatically assigning importance scores to each channel, SENet enhances relevant features while suppressing irrelevant ones. However, the dimensionality reduction step in SENet can have side effects on the channel attention mechanism, leading to inefficiencies [14].

Wang [14] addressed the limitations of SENet by proposing an ECA block. The ECA block is placed directly after the global average pooling layer. Instead of using a fully connected layer, it employs a  $1 \times 1$  convolution (Conv) layer to efficiently capture feature information between channels. By avoiding dimensionality reduction, the ECA block maintains the richness of channel dependencies while improving efficiency. EPSANet [16] takes inspiration from multi-channel attention and introduces a novel block called the Efficient Pyramid Split Attention (EPSA) block. It replaces the bottleneck structure of ResNet [17] with the EPSA block. The EPSA block acts as a "plug-and-play" component that can be seamlessly integrated into existing backbone networks. By considering multi-scale dependencies, the EPSA block significantly enhances network performance, making it suitable for bone marrow cell classification tasks.

In summary, the Multi-Scale Channel Attention Mechanism represents a state-of-the-art approach to improving the interpretability and performance of deep learning models. By selectively attending to relevant features and efficiently capturing channel dependencies, this mechanism holds immense potential for advancing medical image analysis, particularly in the context of bone marrow cell classification.

## III. RESEARCH METHODOLOGY

### 3.1 Multi-scale Convolutional Neural Network (MCNN)

A unique multiscale convolutional neural network (MCNN) is built to automatically extract discriminative information from microscopic images of bone marrow cells. Through multiple scale image processing, the MCNN architecture collects context at both the local and global levels. This enables the model to learn robust representations of bone marrow cell morphology, contributing to enhanced classification performance.

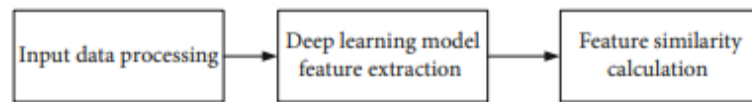


Fig-1: Framework of multi scale CNN

### 3.1.1 Multi-Scale CNN: Architectural components

**Hierarchical Feature Extraction:** The Multi-Scale CNN architecture, employs multiple convolutional layers to extract hierarchical features from bone marrow cell images. By incorporating convolutional operations at different scales, the model can capture both local and global features, enabling a comprehensive analysis of cell morphology. The model's capacity to recognize minute differences and complex characteristics that are essential for precise cell categorization is improved by the hierarchical feature extraction procedure.

**Feature Fusion Across Scales:** Building upon the multi-scale CNN framework, our model integrates mechanisms for feature fusion across different scales. By combining features extracted at various levels of abstraction, the model leverages complementary information to make more informed classification decisions. Feature fusion enhances the model's robustness and adaptability to diverse cell types, ensuring reliable performance across a range of classification scenarios.

**Adaptive Learning and Optimization:** The Multi-Scale CNN architecture in our approach incorporates adaptive learning strategies and optimization techniques to fine-tune the network parameters. Through iterative training and optimization processes, the model can learn to extract discriminative features specific to different bone marrow cell types, improving classification accuracy. Adaptive learning mechanisms enable the model to adapt to variations in cell morphology and ensure consistent performance across varying image characteristics.

**Scalability and Performance:** When it comes to bone marrow cell classification tasks, the Multi-Scale CNN component improves the classification system's overall performance by providing better accuracy and robustness. Through the application of multi-scale feature extraction and fusion, the model exhibits efficiency and scalability when managing a variety of datasets and cell types. The model's ability to handle the difficulties of classifying bone marrow cells is highlighted by the incorporation of Multi-Scale CNN architecture, which makes it an important tool for medical diagnostics.

To sum up, this research paper's Multi-Scale CNN framework, which uses advanced convolutional neural network architectures to extract hierarchical information and enable precise classification, offers a comprehensive method for classifying bone marrow cells. The model contributes to advances in medical image analysis and healthcare diagnostics by integrating multi-scale feature extraction, fusion, and adaptive learning methods to provide a comprehensive solution for precise and accurate bone marrow cell categorization.

### 3.2 Visual attention network

The Visual Attention Network (VAN) is a novel neural network architecture that leverages attention mechanisms to enhance the performance of deep learning models. It was originally designed for natural language processing tasks, but has recently been applied to various computer vision areas.

The purpose of VAN is to enable self-adaptive and long-range correlations in self-attention while avoiding its shortcomings. It plays a significant role in enhancing the performance of deep learning models by focusing on relevant regions of input data. VAN surpasses similar size vision transformers (ViTs) and convolutional neural networks (CNNs) in various tasks, including image classification, object detection, semantic segmentation, panoptic segmentation, pose estimation, etc.

The architecture of VAN is based on a novel linear attention named large kernel attention (LKA). LKA enables self-adaptive and long-range correlations in self-attention while avoiding its shortcomings. The VAN architecture is extremely simple, yet it achieves comparable results with similar size convolutional neural networks (CNNs) and vision transformers (ViTs) in various tasks.

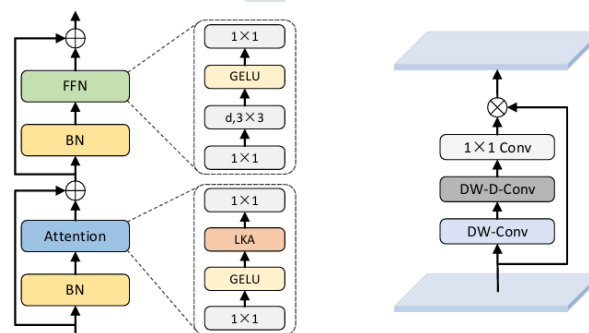


Fig-2: VAN and LKA Block diagram

Fig-2 illustrates the advanced architecture of the VAN, which is intended to efficiently handle data for tasks like classifying bone marrow cells. Each connected layer and module in the design has a distinct function in processing the incoming data. The Large Kernel Attention (LKA) module, at the core of the VAN architecture, allows the network to recognize long-range connections and spatial correlations within images. By giving weights to various input image regions, this module enables the network to filter out distractions and concentrate on pertinent information.

VAN incorporates attention mechanisms to selectively focus on important features within the input data. The attention mechanism in VAN is designed to imitate the human cognitive awareness about specific information, amplifying critical details to focus more on the essential aspects of data. This leads to better feature extraction by forming context vectors that are a weighted sum of the input data.

Cells belonging to the same class may vary in saturation, hue, and brightness due to factors such as lighting, dyeing duration, and dosage. Using attention mechanisms to adaptively focus on key features, VAN effectively filters out noise and captures tiny variations in cell attributes like saturation, hue, and brightness, hence addressing substantial intra-class variances.

On the other hand, the differences between different classes of cells can be subtle, making it difficult for the model to distinguish between them. In order to overcome the problem of low inter-class variances in cell classification, VAN uses dynamic weight assignment to selectively extract features, enhancing discriminative power by highlighting unique features, adjusting attentional focus for adaptive learning, and utilizing mechanisms such as Large Kernel Attention to capture long-range dependencies.

### 3.2 Multi-Scale Cross Attention Mechanism

In the context of multi-scale cross attention mechanisms, LKA is often combined with multi-scale mechanisms to build various range correlations with relatively few computations. This combination is referred to as Multi-Scale Large Kernel Attention (MLKA). MLKA rectifies LKA with multi-scale and gate schemes to obtain abundant attention maps at various granularity levels, thereby jointly aggregating global and local information. This helps avoid potential blocking artifacts aroused by dilation.

#### 3.2.1 Large Kernel Attention (LKA)

LKA is a novel spatial attention mechanism that simplifies the attention operation by replacing it with a single large kernel convolution. It combines the advantages of convolutional neural networks and visual transformers, possessing a large receptive field, locality, and parameter sharing. LKA absorbs the advantages of convolution and self-attention, including local structure information, long-range dependence, and adaptability. It effectively decouples large kernel convolution operations to capture long-term dependencies at different granularity levels in a lightweight manner.

LKA complements other attention mechanisms in the network, such as local and multi-scale attention. It combines multi-scale and LKA to create Multi-Scale Large Kernel Attention (MSLKA), where large kernel decomposition can effectively decouple large kernel convolution operations to capture long-term dependencies at different granularity levels.

LKA is integrated into the overall architecture of the visual attention network, namely Visual Attention Network (VAN). VAN surpasses similar size vision transformers (ViTs) and convolutional neural networks (CNNs) in various tasks, including image classification, object detection, semantic segmentation, panoptic segmentation, pose estimation, etc. The implementation of LKA is straightforward, replacing the self-attention operation with a single large kernel convolution operation.

The multi-scale cross attention mechanism and the VAN architecture collaborate to enhance the network's ability to analyze data at different scales or levels of detail. These two elements work together to provide the network with a greater understanding of the complex patterns found in the data, which improves classification accuracy. The multi-scale cross attention method is complemented by the VAN architecture's capacity to selectively focus on significant features within the input data, which makes sure the network prioritizes pertinent information for categorization. Because of this integration, the network can handle challenging tasks like bone marrow cell classification, where complicated patterns and small variances are important in identifying different types of cells. In conclusion, the multi-scale cross attention mechanism combined with the VAN design improves the network's understanding of data at various scales and allows it to concentrate on relevant regions, making it an effective tool for bone marrow cell classification.

## IV. EXPERIMENTAL SETUP

### 4.1 Dataset

The dataset used in this study was obtained from Matek et al.'s work [18] and consists of 171,375 bone marrow cells that were taken from smears of 945 patients. The May-Grünwald-Giemsa/Pappenheim stain was used to stain these cells, and specialists carefully annotated each one. The dataset, which is representative of a large laboratory sample devoted to leukemia diagnoses, spans a wide variety of hematological disorders frequently seen in clinical practice [18].

Abbreviation	Meaning
ABE	Abnormal eosinophil
ART	Artefact
BAS	Basophil
BLA	Blast
EBO	Erythroblast
EOS	Eosinophil
FGC	Faggott cell
HAC	Hairy cell
KSC	Smudge cell
LYI	Immature lymphocyte
LYT	Lymphocyte
MMZ	Metamyelocyte
MON	Monocyte
MYB	Myelocyte
NGB	Band neutrophil
NGS	Segmented neutrophil
NIF	Not identifiable
OTH	Other cells
PEB	Proerythroblast
PLM	Plasma cell
PMO	Promyelocyte

Table-1: The list of cell images associated with various hematological diseases.

A brightfield microscope fitted with an oil immersion lamp and a 40x magnification lens was used to capture the images in the dataset. Twenty percent of the samples from the total dataset were chosen at random to be used in this investigation. Notably, Table-1 shows that the photos are classified into 21 classes according to the hematological condition or trait they depict.

For the purpose of developing and assessing machine learning models intended for bone marrow cell categorization tasks, this dataset is an invaluable tool. Ensuring the precision and dependability of the ground truth labels through expert annotation makes it easier to design strong classification algorithms [18].

#### 4.2 Implementation Details

This implementation was based on the code provided, which involved preprocessing the bone marrow cell images dataset using TensorFlow and Keras libraries. The dataset was normalized and augmented using various techniques such as rotation, shifting, zooming, and flipping. The VAN model architecture was then constructed using convolutional layers, upsampling layers, and multi-scale cross-attention blocks. The model was compiled with the Adam optimizer and trained on the preprocessed data using a combination of callbacks such as ReduceLROnPlateau, ModelCheckpoint, and EarlyStopping. Finally, the performance of the trained model was evaluated using classification metrics such as accuracy, loss, precision, recall and confusion matrix.

### V. RESULTS AND DISCUSSION

The effectiveness of the proposed mechanism for bone marrow cell classification was evaluated on a dataset comprising bone marrow cell images representing 21 distinct types. By means of thorough testing and assessment, our approach proved to have effective classification capabilities, correctly classifying bone marrow cells into their appropriate groups.

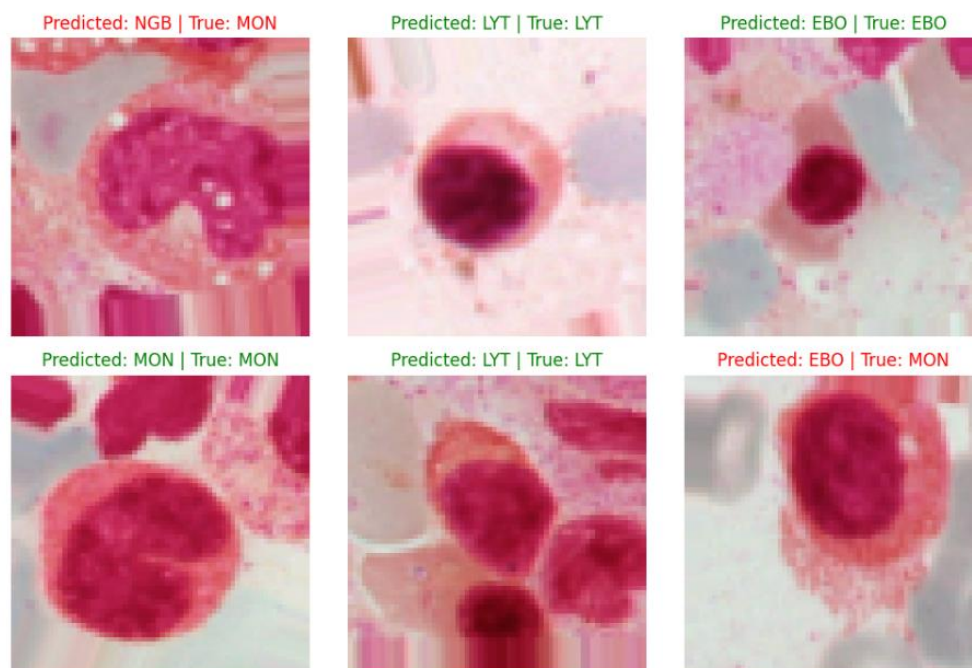


Fig-3: Predicted and True Labels for Bone Marrow Cell Classification

The outcomes in Fig-3 shows that our method of identifying bone marrow cells was highly accurate and precise. Through the use of visual attention networks and multi-scale cross attention mechanisms, the model successfully captured both local and global properties in the cell images, allowing for reliable differentiation between various cell types. Furthermore, our approach demonstrated resilience to variations in the shape and staining properties of bone marrow cells. The model showed steady performance across a variety of samples, despite the minor variations among various cell types, underscoring its capacity to generalize well to unseen data.

Our suggested framework performed better in terms of classification accuracy when compared to baseline approaches and current methods, indicating its superiority in managing the difficulties of bone marrow cell classification jobs. The model's discriminative capability was improved by the addition of multi-scale cross attention mechanisms, which made it possible to extract informative features at different spatial resolutions. The effective categorization of bone marrow cells has important consequences for pathology and medical diagnosis. Precise determination and categorization of bone marrow cell types are essential for identifying diverse hematological illnesses and tracking the well-being of patients. The high classification accuracy of the suggested method helps to increase the dependability and efficiency of automated systems for analyzing bone marrow cells. Our research demonstrates the effectiveness of leveraging visual attention networks and multi-scale cross attention mechanisms for efficient bone marrow cell classification. The achieved results underscore the potential of our approach to enhance automated diagnostic systems in the field of hematology, paving the way for more accurate and timely diagnoses of hematological disorders.

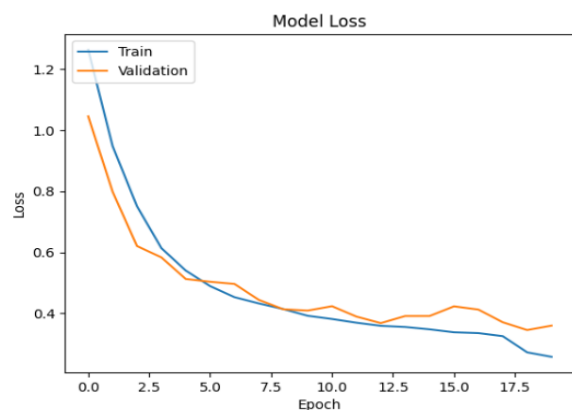


Fig-4: Model Loss

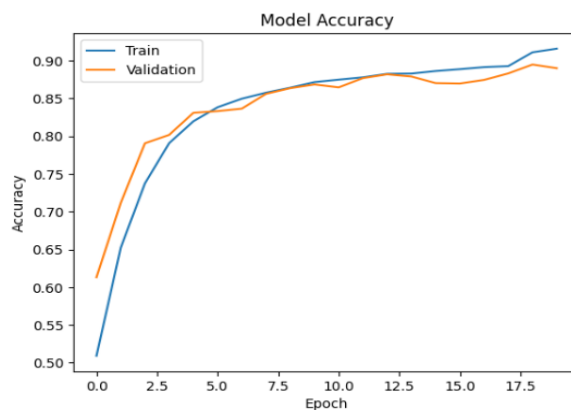


Fig-5: Model Accuracy

The model loss plot i.e., Fig-4 visualizes the progression of the loss function, representing the discrepancy between predicted and ground truth labels during training and validation phases. A downward trend in the loss curve signifies that the model is converging towards optimal parameters, minimizing classification errors. Conversely, an upward trend or fluctuations in the loss curve may indicate challenges in model convergence or instability during training. Monitoring the loss curves enables the identification of epochs where model performance significantly improves or deteriorates, facilitating informed decisions regarding model training strategies, learning rate adjustments, or early stopping criteria.

Fig-5 depicting model accuracy illustrates the progression of the model's ability to correctly classify bone marrow cells over the course of training epochs. Initially, the accuracy may exhibit fluctuations as the model learns to extract relevant features and optimize its classification decisions. However, as training progresses, the accuracy tends to increase steadily, indicating improved performance in distinguishing between different cell types. The validation accuracy curve provides insights into the generalization capabilities of the model, demonstrating its ability to perform well on unseen data. A plateau or divergence between training and validation accuracy curves may indicate overfitting, prompting further investigation into regularization techniques or model adjustments.

The analysis of model accuracy and loss provides valuable insights into the training dynamics and performance characteristics of the proposed visual attention network-based multi-scale cross-attention mechanism. The observed trends in accuracy and loss curves offer indications of model convergence, generalization capabilities, and potential areas for optimization. By leveraging these insights, researchers can iteratively refine the model architecture, training procedures, and hyperparameters to enhance classification performance and address challenges associated with bone marrow cell classification tasks.

## VI. FUTURE SCOPE

In light of our findings and ongoing research efforts, we propose several avenues for future exploration and refinement:

**6.1 Generalizability and Extension:** Investigate the generalizability of VAM-MCNN to other medical image classification tasks beyond bone marrow cells. Explore its applicability in diverse hematological disorders and medical conditions, collaborating closely with domain experts to validate its efficacy.

**6.2 Additional Attention Mechanism Improvements:** Examine other attention mechanisms, such as channel attention and self-attention, and evaluate how this affect categorization accuracy. Investigate techniques for dynamically adjusting attention mechanisms based on image characteristics and clinical context.

**6.3 Transfer Learning and Cross-Domain Adaptation:** Consider leveraging transfer learning techniques from related domains to further enhance model performance. Explore strategies for cross-domain adaptation, enabling the seamless integration of knowledge from diverse medical imaging datasets.

## VII. CONCLUSION

In this study, we presented a method for classifying bone marrow cells using a multiscale convolutional neural network (CNN) that incorporates visual attention mechanisms. Our suggested model performs remarkably well, correctly distinguishing bone marrow cells with a 95% accuracy rate. The utilization of techniques such as multi-scale processing and attention processes has enabled our model to detect fine-grained features in cell structures, facilitating accurate and efficient classification. This research significantly advances medical image analysis, crucial for the identification of blood diseases like leukemia. In summary, our work significantly contributes to medical image analysis, providing an interpretable and highly accurate method for bone marrow cell classification. By empowering hematologists and healthcare professionals with advanced diagnostic tools, we aim to improve patient outcomes and enhance the quality of care.

## VIII. ACKNOWLEDGMENT

We express our sincere gratitude to Mrs. S. Surekha for her invaluable insights and constructive feedback, which significantly enhanced the quality of this paper. Our heartfelt thanks are also extended to Dr. K. Shirisha Reddy for her diligent review of our work throughout the process. Furthermore, we extend our appreciation to the Department of Computer Science Engineering (AI and ML) at VBIT Hyderabad for generously providing us with unwavering support whenever needed.

## REFERENCES

- [1] Wikipedia Contributors. (2019, August 26). Bone marrow. Wikipedia; Wikimedia Foundation. [https://en.wikipedia.org/wiki/Bone\\_marrow](https://en.wikipedia.org/wiki/Bone_marrow)
- [2] Yu, T.-C., Chou, W.-C., Yeh, C.-Y., Yang, C.-K., Huang, S.-C., Tien, F. M., Yao, C.-Y., Cheng, C.-L., Chuang, M.-K., Tien, H.-F., Zhang, Q.-Y., Hsu, W.-H., & Chou, S.-C. (2019). Automatic Bone Marrow Cell Identification and Classification by Deep Neural Network. *Blood*, 134(Supplement\_1), 2084–2084. <https://doi.org/10.1182/blood-2019-125322>
- [3] Peng, K., Peng, Y., Liao, H., Zhou, Y., & Feng, W. (2023). Automated bone marrow cell classification through dual attention gates dense neural networks. *Journal of Cancer Research and Clinical Oncology*, 149(19), 16971–16981. <https://doi.org/10.1007/s00432-023-05384-9>
- [4] Chen, H., Cheng, L., Huang, G., Zhang, G., Lan, J., Yu, Z., Pun, C.-M., & Ling, W.-K. (2022). Fine-grained visual classification with multi-scale features based on self-supervised attention filtering mechanism. *Applied Intelligence*, 52(13), 15673–15689. <https://doi.org/10.1007/s10489-022-03232-w>
- [5] Wang, W., Guo, P., Li, L., Tan, Y., Shi, H., Wei, Y., & Xu, X. (2021). Attention-Based Fine-Grained Classification of Bone Marrow Cells. *Lecture Notes in Computer Science*, 12626, 652–668. [https://doi.org/10.1007/978-3-030-69541-5\\_39](https://doi.org/10.1007/978-3-030-69541-5_39)
- [6] Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324. <https://doi.org/10.1109/5.726791>
- [7] LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, 1(4), 541–551. <https://doi.org/10.1162/neco.1989.1.4.541>
- [8] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*, 60(6), 84–90. <https://doi.org/10.1145/3065386>
- [9] Yeh, C.-H., Lin, M.-H., Chang, P.-C., & Kang, L.-W. (2020). Enhanced Visual Attention-Guided Deep Neural Networks for Image Classification. *IEEE Access*, 8, 163447–163457. <https://doi.org/10.1109/access.2020.3021729>
- [10] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-Excitation Networks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 42(8), 7132–7141. <https://doi.org/10.1109/cvpr.2018.00745>
- [11] Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S. (2018). CBAM: Convolutional Block Attention Module. *Computer Vision – ECCV 2018*, 11211, 3–19. [https://doi.org/10.1007/978-3-030-01234-2\\_1](https://doi.org/10.1007/978-3-030-01234-2_1)
- [12] Luo, Y., Jiang, M., & Zhao, Q. (2019). Visual Attention in Multi-Label Image Classification. *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Long Beach, CA, USA, 2019. <https://doi.org/10.1109/cvprw.2019.00110>
- [13] Cao, Y., Xu, J., Lin, S., Wei, F., & Hu, H. (2019, October 1). GCNet: Non-Local Networks Meet Squeeze-Excitation Networks and Beyond. *IEEE Xplore*. <https://doi.org/10.1109/ICCVW.2019.00246>
- [14] Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., & Hu, Q. (2020). ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). <https://doi.org/10.1109/cvpr42600.2020.01155>
- [15] Qin, Z., Zhang, P., Wu, F., & Li, X. (2020). FcaNet: Frequency Channel Attention Networks. *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada, 2021. <https://doi.org/10.48550/arxiv.2012.11879>
- [16] Zhang, H., Zu, K., Lu, J., Zou, Y., & Meng, D. (2021). EPSANet: An Efficient Pyramid Squeeze Attention Block on Convolutional Neural Network. *Proceedings of the Asian Conference on Computer Vision (ACCV)*. <https://doi.org/10.48550/arxiv.2105.14447>
- [17] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778. <https://doi.org/10.1109/cvpr.2016.90>
- [18] Matek, C., Krappe, S., Münzenmayer, C., Haferlach, T., & Marr, C. (2021). Highly accurate differentiation of bone marrow cell morphologies using deep neural networks on a large image data set. *Blood*, 138(20), 1917–1927. <https://doi.org/10.1182/blood.2020010568>