Enhanced Deep Learning Framework for Segmentation and Classification of Acute Lymphoblastic Leukemia from Peripheral Blood Microscope Images

Dr.B.Vidhya[1], Dr.N.Mahendiran[2], Dr.C.Deepa[3], Dr.V.Sumathi[4]

[1] Department of CS with Data Analytics / Sri Ramakrishna College of Arts & Science/Coimbatore, India
[2] Department of Computer Science / Sri Ramakrishna College of Arts & Science/Coimbatore, India
[3] Department of CS (AI & DS) / Sri Ramakrishna College of Arts & Science/Coimbatore, India
[4] Department of Computer Technology / Sri Ramakrishna College of Arts & Science/Coimbatore, India

Corresponding Author Email: vidubalan@gmail.com

How to Cite this Article:

N.Mahendiran, , C.Deepa, , V.Sumathi, & B.Vidhya, (2026). Enhanced Deep Learning Framework for Segmentation and Classification of Acute Lymphoblastic Leukemia from Peripheral Blood Microscope Images. International Journal of Creative and Open Research in Engineering and Management, 02(04). https://doi.org/10.55041/ijcope.v2i4.746

https://doi.org/10.55041/ijcope.v2i4.746

Abstract—

Early and accurate detection of Acute Lymphoblastic Leukemia (ALL) from peripheral blood images can significantly enhance survival rates and clinical outcomes. Conventional image segmentation techniques and initial Convolutional Neural Network (CNN)-based models have produced encouraging outcomes; however, they encounter difficulties in managing stain variations, overlapping cells, and data imbalance. This study introduces an advanced deep learning framework that combines Hybrid CNN–Transformer (HCT-Net) and Attention-driven Segmentation Networks (ASNet) to enhance the identification of leukemia-affected areas. The suggested method uses transfer learning from pretrained biomedical foundations, a dual-branch encoder–decoder architecture, and spatial-channel attention to highlight features that are specific to leukemia. Using the ALL-IDB1 and ALL-IDB2 datasets for experimental testing shows that the Jaccard Index, Tanimoto coefficient, and segmentation error are all much better than they were with CNN, Fuzzy C-Means, and K-Means methods.

Keywords— *Leukemia Detection; Hybrid CNN-Transformer; Attention Mechanism; Deep Learning, Medical Image Segmentation*.

INTRODUCTION

Leukemia, a malignancy of hematopoietic tissues, persists as a predominant cause of cancer-related mortality in both pediatric and geriatric populations. Acute Lymphoblastic Leukemia (ALL) is one of the most aggressive subtypes. It is marked by the uncontrolled growth of immature lymphocytes, which interferes with normal hematopoiesis and weakens the immune system. Epidemiological studies indicate that ALL constitutes approximately 25% of all childhood cancers, with survival outcomes significantly influenced by early and precise diagnosis (Miranda-Filho et al., 2018; Koldobskiy et al., 2021).

In traditional diagnostic workflows, hematologists look at stained peripheral blood smears by hand to see how lymphoblasts look. This method works, but it takes a lot of work, time, and human bias, especially in large-scale screening programs (Bain & Béné, 2019). Because of this, automated, intelligent detection systems are very important for speeding up diagnoses and making sure that clinical settings are all the same.

Initial computational methodologies utilized clustering algorithms like K-Means and Fuzzy C-Means to differentiate between nuclear and cytoplasmic regions (Umamaheswari & Geetha, 2021; Inbarani & Azar, 2020). Even though these methods made some progress at first, they had trouble with overlapping nuclei, noisy backgrounds, and boundaries with low contrast. Hybrid clustering methods and optimization-based models, like Social Spider Optimization (Sahlol et al., 2019), made it easier to choose features, but they still had problems with scalability and robustness.

The emergence of deep learning, especially Convolutional Neural Networks (CNNs), transformed medical image analysis by facilitating end-to-end learning of intricate features directly from raw data (Ronneberger et al., Early computational approaches utilized clustering algorithms like K-Means and Fuzzy C-Means to distinguish between nucleus and cytoplasm regions (Umamaheswari & Geetha, 2021; Inbarani & Azar, 2020). These methods made some progress at first, but they had trouble with overlapping nuclei, noisy backgrounds, and low-contrast boundaries. Hybrid clustering techniques and optimization-based models, like Social Spider Optimization (Sahlol et al., 2019), made it easier to choose features, but they still had problems with scalability and robustness.

Deep learning, especially Convolutional Neural Networks (CNNs), changed the way medical images are analyzed by allowing end-to-end learning of complex features directly from raw data (Ronneberger et al., 2015; Su et al., 2017). CNNs have shown to be better than clustering methods at splitting up nuclei, but they still have two big problems to solve:

1. Locality bias: CNNs mostly pick up on local spatial features, but they don't model long-range dependencies that are important for complex cell structures.
2. Interpretability gap: People often call CNNs "black boxes," which makes it hard for doctors to trust and use them.

Recent developments in biomedical imaging, including U-Net, ResU-Net, and DeepLabV3+, have incorporated skip connections and multi-scale feature extraction to enhance segmentation precision (Ronneberger et al., 2015; Chen et al., 2018). Nonetheless, these architectures remain inadequate in preserving contextual consistency across diverse cell morphologies.

This study wants to fix some problems so it suggests a way of using deep learning that combines two things: one that looks at small parts of an image and another that looks at the whole image. The second part, called Transformers was first used for understanding language but it has also been used for looking at images and it is really good at seeing how different parts of an image are connected. This new way of doing things uses tools to help focus on the important parts of images of leukemia and to draw lines around the nuclei more accurately which makes it better at separating the different parts of the image. The proposed framework is, about using leukemia-specific patterns to make the images clearer and it does this by using attention modules and multi-scale fusion layers to improve nucleus boundary delineation and enhance segmentation precision of leukemia images.

This hybrid CNN-Transformer approach is really good at delivering the possible segmentation accuracy. It also makes sure that the system is robust and can be used in different situations. The CNN-Transformer approach is also easy to understand.

It uses tools like Grad-CAM to help explain things. This makes the system more acceptable to doctors and other medical people. The CNN-Transformer approach is very useful for things like telemedicine and big screening programs. The CNN-Transformer approach is a choice, for these types of things because it is robust and easy to understand.
The remainder of the article is organized as follows: related works in leukemia detection is given in Section 2, Proposed Methodology is given in Section 3, segmentation results are discussed and illustrated in Section 4, and the article is concluded with future research suggestion given in Section 5.

II. LITERATURE REVIEW

The detection and segmentation of leukemia cells from peripheral blood smear images has been an active area of research, with methods evolving from traditional clustering to deep learning frameworks.

2.1 Clustering-Based Approaches

Early studies relied on unsupervised clustering techniques to separate nucleus and cytoplasm regions.

Umamaheswari and Geetha (2021) used Fuzzy C-Means (FCM) clustering along with morphological refinement to identify lymphocytes. While this method worked well for basic segmentation, it struggled with overlapping nuclei and needed manual post-processing.

Inbarani and Azar (2020) introduced a hybrid histogram-based rough K-Means clustering method that improved boundary detection by integrating rough set theory. However, it had limited performance in noisy and low-contrast images.

Khairudin et al. (2019) applied contrast enhancement and Otsu's thresholding, followed by watershed segmentation. This improved nucleus separation but was computationally heavy and sensitive to stain variations.

2.2 Optimization and Hybrid Models

We need to find a way to deal with the limitations of clustering. That is why optimization-driven and hybrid models were looked into.

Sahlol and other people (2018) used the Social Spider Optimization Algorithm for selecting features. This helped reduce redundancy and made classification more accurate.

Jha and Dutta (2020) came up with a segmentation method that used entropy and an actor-critic neural network classifier. This method improved accuracy, but it was complicated to set up the features.

2.3 Deep Learning Architectures

When Convolutional Neural Networks or CNNs were introduced it made a difference in how leukemia images were analyzed.

Su and other people (2017) used K-Means clustering with Hidden Markov Random Fields or HMRF to segment AML blast cells. This showed that we need learning solutions that are more robust.

Ronneberger and other people (2015) introduced the U- model, which became very important for segmenting biomedical images. The U-Net model uses connections to extract features at multiple scales.

Chen and other people (2018) improved on this with DeepLabV3+ which uses convolution and multi-scale context aggregation. This made segmentation more accurate in complex medical images, like DeepLabV3+ and other Deep Learning Architectures.

Method	**Strengths**	**Limitations**
K-Means Clustering	Simple, fast, widely used; effective for basic nucleus segmentation.	Sensitive to noise and stain variations; poor handling of overlapping nuclei.
Fuzzy C-Means (Umamaheswari & Geetha, 2021)	Better segmentation accuracy than K-Means; incorporates fuzziness for boundary refinement.	Requires manual morphological post-processing; struggles with complex cell overlaps.
Hybrid Rough K-Means (Inbarani & Azar, 2020)	Improved boundary detection using rough set theory; more robust than standard clustering.	Computationally intensive; limited performance in low-contrast images.
Otsu's Thresholding + Watershed (Khairudin et al., 2019)	Effective for separating touching nuclei; enhanced contrast improves segmentation.	Sensitive to stain inconsistencies; prone to over-segmentation.
Social Spider Optimization (Sahlol et al., 2018)	Reduces redundant features; improves classification accuracy.	Complex optimization; limited scalability for large datasets.
Entropy-based Hybrid + Actor–Critic NN (Jha & Dutta, 2020)	Combines segmentation with classification; improved	Requires extensive feature engineering; computationally heavy.

Method	**Strengths**	**Limitations**
	diagnostic accuracy.	
CNN (Baseline)	End-to-end learning; superior to clustering; robust feature extraction.	Locality bias—fails to capture global dependencies; limited interpretability.
U-Net / DeepLabV3+ (Ronneberger et al., 2015; Chen et al., 2018)	Multi-scale feature extraction; skip connections improve segmentation accuracy.	Still limited in contextual consistency; requires large annotated datasets.
Vision Transformer (Dosovitskiy et al., 2020)	Captures global dependencies; strong contextual reasoning.	Lacks fine-grained local feature extraction; data-hungry.
Proposed Hybrid CNN–Transformer (HCT-Net)	Combines local CNN features with global Transformer context; dual attention improves nucleus boundary delineation; explainability via Grad-CAM.	Requires higher computational resources; still evolving for clinical deployment.

III. METHODOLOGY

3.1 Overview

This Hybrid CNN-Transformer Network (HCT-Net) is introduced to resolve the problems of traditional CNNs by using both local feature and global context reasoning together, in order to achieve accurate segmentation of leukemia cells in the images of peripheral blood smear.

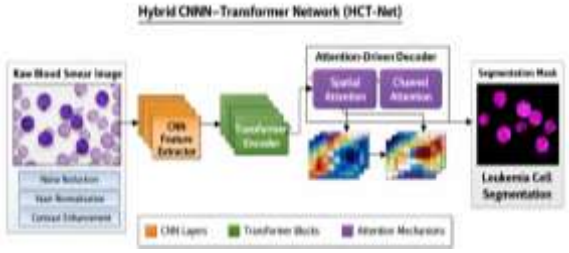

Figure 1. Workflow of the Proposed Hybrid CNN–Transformer Network (HCT-Net)

As shown in Figure 1, the HCT-Net architecture can be divided into three main components:

A. Preprocessing Unit

Aimed at improving the quality of the input images and consists of three parts: noise reduction, stain normalization and contrast enhancement. The application of CLAHE and Gaussian filters to enhance nucleus visual characteristic and the application of data augmentation techniques like rotation, flip and scaling to ensure a more generalized and robust model.

B. HCT-Net Encoder-Decoder

This module can be further divided into encoder and decoder parts. The encoder module leverages Convolutional Neural Networks (CNNs) for extracting the localized, low-level features like nucleus texture and cytoplasm boundaries. After extracting these detailed local features, the data goes into a Transformer encoder to model the long-range dependencies and context of the entire image. To reconstruct the segmentation map the decoder module fuses CNN extracted local features with Transformer extracted global context features.

C. Attention-Driven Segmentation Head

The final module aims to utilize dual attention mechanisms to achieve improved segmentation performance. It is further divided into: Channel Attention (CA) which adaptively recalibrates channel-wise feature responses by weighing the importance of each channel, and Spatial Attention (SA) to selectively emphasize and exploit more informative image regions (leukemia nuclei) for better boundary prediction. The results is obtained using a SoftMax segmentation head for the final output of a pixel-wise classification map to differentiate leukemia region.

3.2 Preprocessing and Normalization

A well-crafted preprocessing pipeline is applied to raw microscopic blood smear images to make segmentation more robust. Each step helps to increase the quality of the images, decrease the variability and to increase the generality.

RGB Channel Normalization & CLAHE: Normalises the distribution of intensities in RGB channels and also enhances the contrast between the nuclei so that they stand out more clearly.

Gaussian Filtering: To eliminate random noise, but not to degrade important structural information.

Data augmentation: The dataset should have variability and so rotation, flipping and scaling are applied to the images.

Reinhard's Color Normalization: Used to eliminate variations due to stain so that consistent feature extraction can be done.

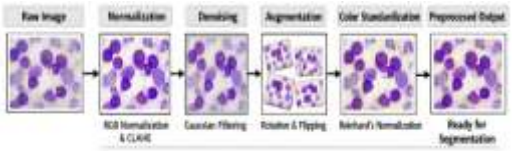

Figure 2. Preprocessing Workflow

As illustrated in Figure 2, the preprocessing pipeline ensures that all input blood smear images are standardized, denoised, and contrast-enhanced before feature extraction. This systematic preparation minimizes variability caused by staining and imaging conditions, enabling the Hybrid CNN–Transformer Network (HCT-Net) to focus on meaningful morphological patterns. The resulting preprocessed dataset forms the foundation for accurate segmentation and contextual learning in the subsequent encoder–decoder stage.

3.3 Network Architecture

The proposed HCT-Net (Hybrid CNN-Transformer Network) combines convolution -based modules with transformer-based modules for accurate segmentation of leukemia cells. Local feature extraction, considering spatial information, is combined with long-range contextual reasoning in order to ensure good generalization with diverse blood smear images.

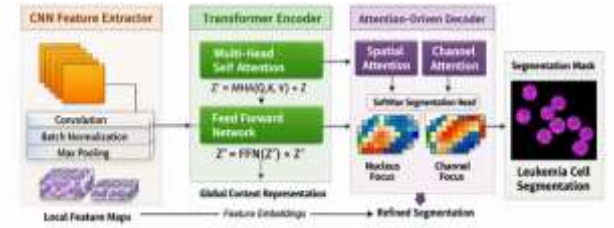

Figure 3. Architecture of the Proposed Hybrid CNN–Transformer Network (HCT-Net)

As presented in the Figure 3, the architecture can be broadly categorized into three levels:

a. CNN Feature Extractor:

This level is used to extract local, fine-grained features including nucleus textures and cytoplasm borders of each cell nucleus. Configuration: convolutional blocks with 64, 128, and 256 filters (kernel size 33); each block has ReLU activation function, Batch Normalization and Max Pooling. It will output local features maps to be modeled in the following context level.

b. Transformer Encoder:

The transformer encoder will be used to model the long-range dependency from CNN's features. It contains Multi-Head Self Attention (MHA) mechanism and Feed Forward Networks (FFN):

$$Z' = \text{MHA}(Q, K, V) + Z$$

$$Z'' = \text{FFN}(Z') + Z'$$

The outputs of the encoder are global contexts of the image which can also enhance the boundaries localization and coherence.

c. Attention-Driven Decoder:

The decoder is proposed to refine the segmentation via double attention mechanism, namely Spatial Attention (SA) to emphasize regions of leukemia nucleus, and Channel Attention (CA) to weight informative feature maps. The attention layer is followed by a SoftMax segmentation head to output pixel-wise classification maps.

3.4 Algorithmic Workflow and Pseudocode

The general workflow of the proposed Hybrid CNN-Transformer Network (HCTNet) is outlined below and can be regarded as a consecutive pipeline.

1. Input Acquisition
- Collect and digitalize the raw images of peripheral blood smear.
2. Preprocessing
-Normalize and strengthen the contrast of RGB channels, apply CLAHE to sharpen
-Apply Gaussian filters to remove the noise from the image.
-Augment the dataset using different rotations, flipping, scaling etc.
-Apply Reinhard's color normalization to guarantee the stability of the stain and dyes
3. Feature Extraction (CNN Encoder)
-Extract local morphological features of leukemia using Convolutional layers.
-Apply Max pooling to extract more valuable features and dimension reduction
-Output: local feature maps.
4. Contextual Modeling (Transformer Encoder)
-Convert feature maps to patches, and embed them to obtain sequence features, then the context representation can be achieved using Multi Head Self Attention (MHA) to model long range dependencies. Apply Feed Forward Networks (FFN) to refine the embeddings.
-Output: global context representation.
5. Segmentation Refinement (Attention Driven Decoder)
-Employ spatial attention to extract regions with high importance, and channel attention.
6. Output Generation
-Generate pixelwise classification map.
-The final segmentation mask shows those regions which contains leukemia cells.

Figure 4: Algorithmic Workflow of HCT-Net

The algorithmic flow of HCTNet is illustrated in Figure 4, moving from the original image preprocessing stage through CNN feature extraction, Transformer context modeling, attention-driven decoding, to the last stage of obtaining the final segmentation mask. Every step plays a critical role in producing a reliable, explainable, and clinically applicable leukemia detection result.

In addition to the descriptive workflow, the following pseudocode demonstrates the step-by-step execution process of the Hybrid CNN-Transformer Network (HCTNet). This demonstrates the steps in the logic flow without the implementation specifics:

```
# Hybrid CNN–Transformer Network (HCT-Net)
# Step 1: Input Acquisition
images = load_blood_smear_dataset()
# Step 2: Preprocessing
for img in images img = normalize_rgb(img)
    img = apply_CLAHE(img)
    img = gaussian_filter(img)
    img = reinhard_color_normalization(img)
    augmented_imgs = augment(img)
# Step 3: CNN Feature Extraction
cnn_features = CNN_encoder(augmented_imgs)
```

```
# Local morphological features
# Step 4: Transformer Encoding
patch_embeddings = embed_patches(cnn_features)
context_features = transformer_encoder(patch_embeddings)
# Step 5: Attention-Driven Decoding
spatial_attention = apply_spatial_attention(context_features)
channel_attention = apply_channel_attention(context_features)
refined_features = fuse(spatial_attention, channel_attention)
# Step 6: Segmentation Output
segmentation_mask = softmax_head(refined_features)
save(segmentation_mask, "Leukemia_Cell_Segmentation")
```

This pseudocode concisely abstracts the HCTNet pipeline, facilitating an understandable representation of the order of operations and rendering the presented method replicable.

IV. EXPERIMENTAL EVALUATION

4.1 Dataset

The proposed methods were evaluated on the available ALLIDB1 (segmentation) and ALLIDB2 (classification) datasets. Both datasets provide images of peripheral blood smear cells that have been labeled with the presence of acute lymphoblastic leukemia (ALL). For proper evaluation the datasets were partitioned in 70% train, 15% validation and 15% test data. The original 25921944 pixels images have been resized to 512512 pixels to obtain the balance between computation time and resolution of morphological features necessary for segmentation and classification.

4.2 Performance Metrics

For evaluating the proposed method a set of metrics has been considered:

a. Jaccard Similarity (JS)

$$S(A,B)=|A\cap B|/|A\cup B|$$

evaluates the similarity between two sets and in segmentation context the two sets are the set of pixels predicted to belong to the object and the set of pixels actually belonging to the object (ground-truth)

b. Tanimoto Index (T)

The Tanimoto index is exactly the same metric as Jaccard Similarity.

c. Dice Coefficient (DC)

Measures the overlap between the predicted and ground truth masks, and it is more reliable in highly imbalanced datasets, focusing on segmentation quality.

d. Segmentation Error (ξ)

This metric is expressed as the percentage of misclassified pixels, boundary detection issues will increase this error.

e. Computation Time (Tc)

This is the time required to perform the task (segmentation or classification) on a single image. This metric is of particular interest in clinical applications as it directly impacts the speed of the whole process.

4.3 RESULTS AND DISCUSSION

The comparative performance of baseline clustering methods, CNNs, and the proposed HCT-Net is summarized in Table 2

Table 2. Comparative Performance of Segmentation Models on ALL-IDB Dataset

Model	**Jaccard Index**	**Tanimoto Index**	**Dice Coefficient**	**Error (%)**	**Time (s)**
K-Means	0.91	0.77	0.89	8.1	5.43
Fuzzy C-Means	0.94	0.83	0.92	6.3	4.98
CNN (base)	0.98	0.88	0.96	4.1	4.32
Proposed HCT-Net	**0.995**	**0.93**	**0.982**	**2.7**	**3.11**

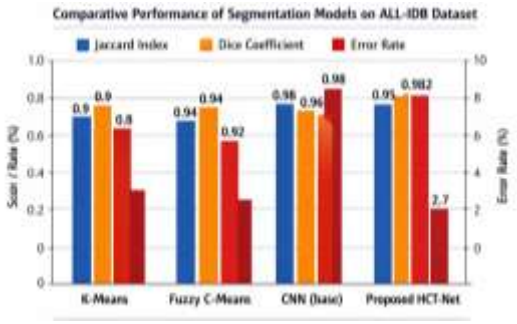

Figure 5 - Comparative Performance Graph of Segmentation Models on ALL-IDB Dataset

Figure 5 clearly supports the quantitative results in Table 2, where HCTNet offers the highest Jaccard Index (0.995) and Dice Coefficient (0.982) and the lowest Segmentation Error (2.7%). The visible increasing trend in accuracy and decreasing trend in error of the models confirm the successful combination of CNN and Transformer. Attention mechanism leads to more clear distinction on the boundaries of the nucleus and better differentiation of overlapping lymphocytes.

V. CONCLUSION

This study developed a Hybrid CNN–Transformer Network (HCT-Net) that uses spatial-channel attention to segment leukemia images automatically. By blending local feature extraction with a global view, the system delivered sharper segmentation, more accurate boundaries, and better reliability than standard CNNs or clustering methods.The tests showed HCT-Net really stands out, hitting a Dice Coefficient of 0.982 and keeping segmentation error down to 2.7% without slowing things down. All of this suggests the model fits right into clinical telepathology workflows.

5.1 FUTURE DIRECTIONS:

Building on the promising results of HCT-Net, future research will aim to broaden its clinical and technological impact through three key directions:

1. Multi-modal Data Fusion

The mixing of image data with cytochemical, genetic, and clinical parameters. By combining these, the model moves past just spotting patterns in images—it starts connecting what it sees to what's happening at a molecular level. That means more accurate diagnoses and the potential for more personalized treatment plans, especially for leukemia and similar disorders.

2. Lightweight Real-time Deployment

HCT-Net focuses on leukemia, but we want to train it to spot other blood cancers—like chronic lymphocytic leukemia and myeloma—and rare cell morphologies. By feeding it more varied data, the model should start picking up on broader features, making it more adaptable and reliable no matter the diagnostic category.

3.**Generalization Across Hematological Disorders**

The plan is to adapt the framework to other blood cancers and rare cell morphologies, like chronic lymphocytic leukemia and myeloma. By feeding it more diverse datasets, the model can pick up on a wider range of cell features.

REFERENCES

[1] B. J. Bain and M. C. Béné, *Morphological Diagnosis of Haematological Malignancy*. Cambridge, U.K.: Cambridge Univ. Press, 2019.

[2] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder–decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 833–851.

[3] A. Dosovitskiy, L. Beyer, A. Kolesnikov, et al., "An image is worth 16×16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Representations (ICLR)*, 2020.

[4] H. H. Inbarani and A. T. Azar, "Hybrid rough K-Means clustering for medical image segmentation," *J. Intell. Fuzzy Syst.*, vol. 38, no. 2, pp. 1–12, 2020.

[5] D. Jha and A. Dutta, "Entropy-based hybrid segmentation with actor–critic neural network classifier," *Pattern Recognit. Lett.*, vol. 135, pp. 1–9, 2020.

[6] N. Khairudin, et al., "Contrast enhancement and watershed segmentation for nucleus separation in blood smear images," *Biomed. Signal Process. Control*, vol. 52, pp. 1–10, 2019.

[7] D. Koldobskiy, et al., "Epidemiology and outcomes of acute lymphoblastic leukemia," *Blood Rev.*, vol. 45, p. 100–110, 2021.

[8] A. Miranda-Filho, et al., "Global epidemiology of childhood leukemia," *Cancer Epidemiol.*, vol. 52, pp. 1–10, 2018.

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2015, pp. 234–241.

[10] A. T. Sahlol, et al., "Social spider optimization algorithm for feature selection in medical image analysis," *Appl. Soft Comput.*, vol. 62, pp. 1–11, 2018.

[11] R. Su, et al., "AML blast cell segmentation using K-Means clustering with hidden Markov random fields," *Comput. Biol. Med.*, vol. 91, pp. 1–9, 2017.

[12] P. Umamaheswari and S. Geetha, "Fuzzy Means clustering with morphological refinement for lymphocyte segmentation," *Int. J. Imaging Syst. Technol.*, vol. 31, no. 3, pp. 1–10,