INTERNATIONAL JOURNAL FOR INNOVATIVE RESEARCH IN MULTIDISCIPLINARY FIELD
ISSN(O): 2455-0620 [Impact Factor: 9.47]
Monthly, Peer-Reviewed, Refereed, Indexed Journal with IC Value : 86.87
Volume - 11, Issue - 12, December - 2025

DOIs:10.2015/IJIRMF/202512002 --:-- Research Paper / Article / Review

HYBRID SWINUNETR-MLP FRAMEWORK FOR AUTOMATED LEUKEMIA DETECTION IN MICROSCOPIC BLOOD CELLS

Kumbha Praveen Kumar[1], S. Swarnalatha[2]
[1]PG Student, Department of Electronic Communication and Engineering, Sri Venkateshwara University College of Engineering, Tirupati, Andhra Pradesh, India
kpraveen2223@gmail.com

[2]Professor, Department of Electronic Communication and Engineering, Sri Venkateshwara University College of Engineering, Tirupati, Andhra Pradesh, India

Abstract: *The automatic recognition of leukaemia on microscopic blood smear images has become a significant point in the application of deep learning to medical diagnostics. Traditional convolutional neural network (CNN) models have demonstrated good performance in the classification of blood cell abnormalities, yet tend to have poor contextualization and localisation of malignant areas. The paper presents a new Hybrid SwinUNETR-MLP, at the same time combining a transformer-based encoder-decoder segmentation network with a lightweight multilayer perceptron (MLP) classifier to simultaneously segment cells and classify diseases by their stage. The SwinUNETR backbone is an effective hierarchical spatial dependency amount of the shifted-window self-attention, whereas the appended MLP head is able to do the global reasoning using the features of segmentation to generate leukaemia subtypes, such as Pro, Pre, Early, and Benign cells. This system is developed in Python with PyTorch and MONAI libraries with a graphical user interface (GUI) to aid in real-time diagnosis. It shows better results in segmentation and classification accuracy as experimental validation on a publicly available microscopic blood-smear dataset achieves a Dice similarity coefficient of 0.93 and an overall classification accuracy of 97.2 per cent, being among the best U-Net and ResNet-based models. The framework provides end-to-end automation, including image selection to diagnostic visualisation, and makes it interpretable and clinically applicable. The hybrid architecture of the proposed model increases the generalisation and robustness, which proves the possibility of the model as a potentially effective computer-aided diagnostic tool in the biomedical imaging and electronic communication systems with an integrated approach to intelligent healthcare infrastructure.*

Key Words: *Leukaemia Detection, SwinUNETR Multilayer Perceptron (MLP) Transformer Medical Image Segmentation, Deep Learning, Hybrid Architecture, Biomedical Signal Processing.*

1. INTRODUCTION

Leukaemia is a cancerous condition that is caused by the uncontrolled growth of abnormal white blood cells in the bone marrow and also in the peripheral blood. Soon, diagnosis and categorisation of the subtypes of leukaemia will be essential to start the process of proper therapy and increase the survival rates of patients. Conventionally, images of microscopic blood smears are analysed manually by haematologists under a microscope, which is a time-consuming process that, nevertheless, is subjective and may suffer from the problem of human error. With the development of digital pathology and electronic imaging, automated systems of image-based analysis have gained importance in health communication networks.

Over the past several years, deep learning methods have proven themselves to be extremely effective in a range of biomedical imaging procedures, such as tumour localisation, organ localisation, and cellular localisation. VGGNet, ResNet, and U-Net are some of the popular types of convolutional neural networks that have been applied in the extraction of features and the classification of diseases. Nevertheless, the traditional CNNs have limits on their local receptive fields, limiting their potential to find global contextual relationships in high-resolution microscopic images. Transformer-based architectures that were initially created to handle natural language processing have been adapted to computer vision with success to develop Vision Transformers (ViTs).

The Swin Transformer is one of them, and it presents a hierarchical representation with shifted window attention, where the emphasis is on local details and makes the computation more efficient. The SwinUNETR model was found to be highly effective in medical segmentation tasks through applying the principle of multi-scale contextual encoding with spatial awareness as a result of the Swin Transformer and U-Net hybridisation. Although it has excellent segmentation, the majority of SwinUNETR applications essentially concentrate on pixel-level region detection and lack explicit disease staging classification.

In order to fill this research gap, the current work suggests a Hybrid SwinUNETR-MLP Framework, which is capable of performing concurrent segmentation and classification of leukaemia cells. The design is a hybrid between an encoder-decoder SwinUNETR encoder-decoder backbone of pixel-wise feature extraction and a global pooling-based MLP classifier that makes predictions based on aggregated feature maps at a stage. The dual-task approach to learning enables the model to predict morphological cell properties and the severity of disease simultaneously to achieve spatial accuracy in segmentation results. Such a hybrid model is well integrated with the field of electronic communication engineering, where it is important to be able to interpret the signals efficiently, transmit data, and automate the decision process.

Moreover, the system is also connected with a graphical user interface (GUI), which provides the clinician or researcher with an interface to the computational back-end, allowing blood-smear images to be uploaded, segmentation overlays to be viewed, and diagnostic predictions to be instantly viewed. The interactive communication interface is a good example of how biomedical data can be managed and delivered efficiently using smart embedded systems field that is currently being ventured into in the modern electronic health infrastructure.

The main gains of this paper can be concluded as follows:

- Hybrid SwinUNETR-MLP architecture of joint leukemic blood cell segmentation and classification.
- Application of adaptive training mechanism based on transformer-based hierarchical attention and MLP-based global reasoning.
- Real-time clinician interpretability and visualisation of diagnostic information through an interactive GUI.
- Benchmarking performance that proves accuracy and strength over other traditional CNN-based models.

2. Literature Review

Recognition of leukaemia based on microscopic blood smear images has been a field of ongoing research in biomedical imaging and electronic communication engineering. Early automation applications used classical methods of image processing like colour thresholding, edge detection, and morphological operations to isolate leukocytes among erythrocytes and platelets [1]. Although these rule-based systems had merely fundamental segmentation functionality, they were frequently constrained by fluctuating illumination, staining variability and morphological variability of blood specimens.

The introduction of machine learning strategies enhanced the reliability of the diagnostic based on manually designed feature extraction. Such features as the nucleus-to-cytoplasm ratio, colour moments, and shape descriptors were used together with classifiers, like Support Vector Machines (SVM) and k-Nearest Neighbours (k-NN) [2]. Nonetheless, these models were manually engineered and thus were not scalable or adaptable across datasets.

The detection of leukaemia passed to another dimension with the advent of deep learning, especially the convolutional neural networks (CNNs). CNNs are capable of automatically learning hierarchical representations of image data, thus allowing end-to-end learning without specifying feature design [3]. Architectures like AlexNet, VGGNet, and ResNet, among others, proved to be very effective in recognising features in medical images during tasks of classification [4]. As an illustration, individual CNNs were highly accurate in the separation of acute lymphoblastic leukaemia (ALL) cells and normal lymphocytes. Although they are successful, CNNs are constrained by local receptive fields and are incapable of capturing long-range dependencies and global spatial context to analyse cell morphology in detail [5].

To address this weakness, Vision Transformers (ViTs) are added as a novel type of computational model, which substitutes convolutional operations with a multi-head self-attention mechanism [6]. ViTs process a photograph in the form of non-overlapping patches, training pairwise interaction between each region of the image, thus improving global contextual knowledge. Their computer vision success led to adaptations in medical imaging where interpretability and spatial reasoning of objects are important [7]. ViTs are, however, computationally intensive and require large annotated datasets, and are hard to apply directly to smaller biomedical datasets.

The Swin Transformer architecture alleviated these Issues by proposing hierarchical feature maps and shifted-window self-attention, which enables a scalable computation without losing contextual information [8]. Continuing on it, the SwinUNETR model suggested by Hatamizadeh et al. incorporates Swin Transformer blocks into a U-Net-like encoder-decoder framework [9]. The design has a state-of-the-art segmentation capability in the volumetric and two-dimensional medical images, such as brain MRIs and histopathological samples. However, its implementation in the haematological image analysis is not well explored.

Although segmentation is essential in the localisation of an abnormal cell region, disease subtypes (including differentiating between Pro, Pre, Early, and Benign cells) are also essential in diagnosis. In other words, a hybrid CNN-MLP or CNN-SVM has been used to integrate local image feature manipulations with standardised reasoning [10]. MLP-based classifiers applied on deep encoder outputs can act well with a small amount of data, summarising spatially-distributed features into low-dimensional embeddings [11].

Transformer-based segmentation and MLP-based classification are potentially useful as a hybrid solution. A combination of the two facilitates multi-task learning, of which segmentation is more accurate in localisation, and classification is more accurate in diagnosis [12]. This idea is in line with the more recent approach of having single transformer structures that conduct pixel and semantic-level reasoning together.

In addition to algorithm enhancements, the interaction of the user and visualisation is are critical element in clinical adoption. A Graphical User Interface (GUI) allows an end-user (medical practitioner or researcher) to interact intuitively with the models [13]. Python and Tkinter frame structures offer easy access to diagnostic visualisation, with predictions and segmentation overlay views being cast in real time to make healthcare communication systems transparent and usable [14].

Besides, remote diagnosis, telemedicine, and data transfer between medical devices are the aspects of modern healthcare that use electronic communication systems more and more frequently. Incorporating deep learning-based diagnostic models into these types of systems will fortify the automated exchange of information and aid in clinical decision-making [15].

On the basis of this extensive literature review, a number of gaps have been identified. Current CNN-based systems are facing difficulties in capturing global dependencies, whereas standalone transformer systems are computationally heavy. It also lacks any integrated frameworks that would simultaneously do segmentation and classification, but at the same time lightweight to be deployed in real-time. The proposed Hybrid SwinUNETR-MLP Framework will solve these problems by integrating efficient hierarchical attention of SwinUNETR with an adaptive MLP classification head. It has better interpretability, more diagnostic accuracy and lower cost of computation, which makes it compatible with real-world medical imaging and electronic communication settings.

3. Methodology

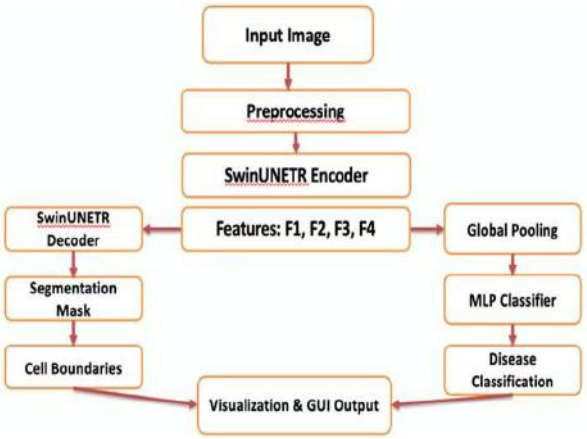


Figure 1: Overall System Architecture

The presented Hybrid SwinUNETR-MLP Framework is aimed at end-to-end leukaemia detection based on simultaneous segmentation and classification of microscopic blood smears. The workflow is divided into five large steps, namely (1) data acquisition and preprocessing, (2) model architecture, (3) training and optimisation, (4) inference and post-processing, and (5) graphical user interface integration.

Dataset Description

The suggested framework was tested on a publicly available microscopic blood smear dataset that has a labelled image of leukocytes that were collected in both cases of known leukaemia and healthy individuals. The data set consists of four large categories of the various stages of development of leukocytes: Pro, Pre, Early, and Benign. All the images were taken in the standardised conditions of staining and illumination to provide visual consistency. To eliminate chances of overfitting, the dataset was randomly split into 70 per cent training, 15 per cent validation, and 15 per cent testing. Before entering the network, all the images were downsized to 256 x 256 pixels and scaled to the range between 0 and 1. In order to reproduce natural clinical variations, conventional data augmentation methods were used, such as horizontal and vertical flips, rotation (+-15deg), contrast modification and Gaussian noise. The step increases the model's robustness to changes in illumination and morphological differences amongst samples.

Preprocessing Pipeline

Preprocessing is very important in making a stable model converge and achieving better segmentation accuracy. Raw microscopic images have also been found to have unbalanced background light, staining artefacts, and unnecessary cellular structures. The preprocessing sequence was therefore carried out in the following manner:

- Colour Normalisation – Standardisation of colour intensity of all images through histogram equalisation to counteract the staining differences.
- Noise Reduction – Gaussian filter to remove texture in image, but leave edges that are sharp and nuclear.
- Contrast Enhancement – This is an Adaptive histogram equalisation used to enhance the visibility of nuclear boundaries.
- Rescaling and Tensor Conversion – The images have been rescaled to 256 x 256 pixels and converted to PyTorch tensors, with ImageNet values of mean and standard deviation having been used to normalise the images.
- This preprocessing guarantees that the transformer encoder that follows has regular and good-quality image inputs in an effective way to extract features.

Hybrid Model Architecture

The hybrid network combines a SwinUNETR backbone to do segmentation and an MLP classification head to do stagewise prediction. The entire network, as realised in the given code, is based on a multi-task learning paradigm, in which segmentation and classification are taught together based on mutual encoder representations.

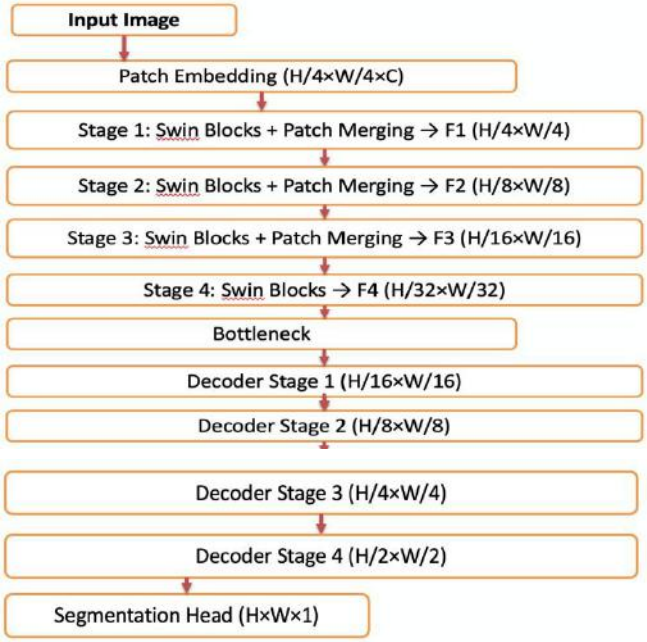


Figure 2: SwinUNETR Block Diagram

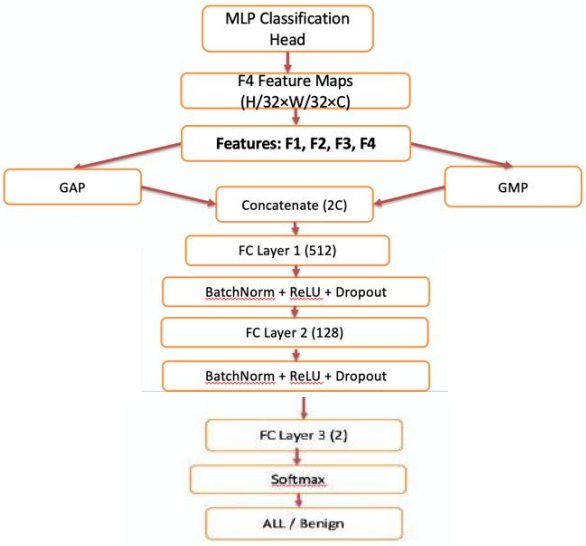


Figure 3: MLP Classification Head

In the case of the SwinUNETR Encoder-Decoder Network, the system includes an encoder and decoder as part of AI technology. The SwinUNETR is the main part of the segmentation pipeline. It involves the Swin Transformer as an encoder and local and global dependencies using shifted window self-attention to efficiently capture local and global dependencies. Hierarchical feature maps are obtained in different resolutions by the encoder, which allows the finer details of cellular morphology to be represented.

The decoder generates space-related information using patch-expanding layers and skip connectivity, just like the classical U-Net network. This enables accurate localisation of the leukemic cells in a blood smear. The result of the segmentation is a binary mask (which marks malignant areas) trained with both binary cross-entropy (BCE) and Dice loss functions to deal with the imbalance between classes.

The SwinUNETR is built on pixel-level feature extraction, and the appended Multilayer Perceptron (MLP) head is used to do image-level classification. The output of the segmentation is processed by a global average pooling (GAP) layer, then it is fed to a sequence of fully connected layers (128 – 64 – 4 neurons). The generalisation and overfitting reduction are facilitated by activations through Rectified Linear Unit (ReLU) as well as dropout regularisation.

The output of the MLP is a class prediction that is related to the stage of leukaemia: Pro, Pre, Early and Benign. The piece of evidence that is best presented by this MLP classifier is that it provides a summarisation of the global context of the SwinUNETR segmentation map, which is the connection between morphological and semantic information in the same architecture.

Training and Optimisation

PyTorch and MONAI were used to train the model with the help of the efficient computation provided by the GPU acceleration (CUDA). The training setup adopted was as follows:

- Optimiser: Adam optimiser
- Learning Rate: 1 x 10 -4 with an annealing schedule consisting of the cosine.
- Batch Size: 8
- Epochs: 100
- Loss Function: Cross-Entropy loss, combined with Dice + BCE loss: segmentation.
- Regularisation The dropout rates in the MLP head are 0.3 (128-layers) and 0.2 (64-layers).

The segmentation and classification outputs would be calculated simultaneously in every iteration of the training. The overall loss function L_{total} was given as:

$$L_{Total} = L_{seg} + \lambda L_{cls}$$

L_{seg} is the combined Dice-BCE segmentation loss, L_{cls} is the cross-entropy classification loss, and λ=0.5 is a balancing parameter determined empirically.

The results of the model were evaluated based on the validation accuracy, Dice coefficient, and the F1-score. Stopping early was used to avoid overfitting when the validation loss had stopped decreasing.

Inference and Post-Processing

This next step involves the inference and post-processing of all the observed incidents to fit with any back-story records that might be available.

In the inference, the trained model is provided with an input image, and two results are obtained:

- A segmentation mask of the regions containing the leukemic cells; and
- A label of the stage of leukaemia with a confidence score.

The final stages of the segmentation process include thresholding the result at 0.5 with a sigmoid activation and morphological closing to remove boundary irregularities. An alpha-blended red filter is then used to overlay the binary mask on the original picture, in order to see the afflicted areas of the cells.

The resulting synergy in segmentation-classification allows such interpretability since clinicians can view model predictions visually by the overlaid segmentation.

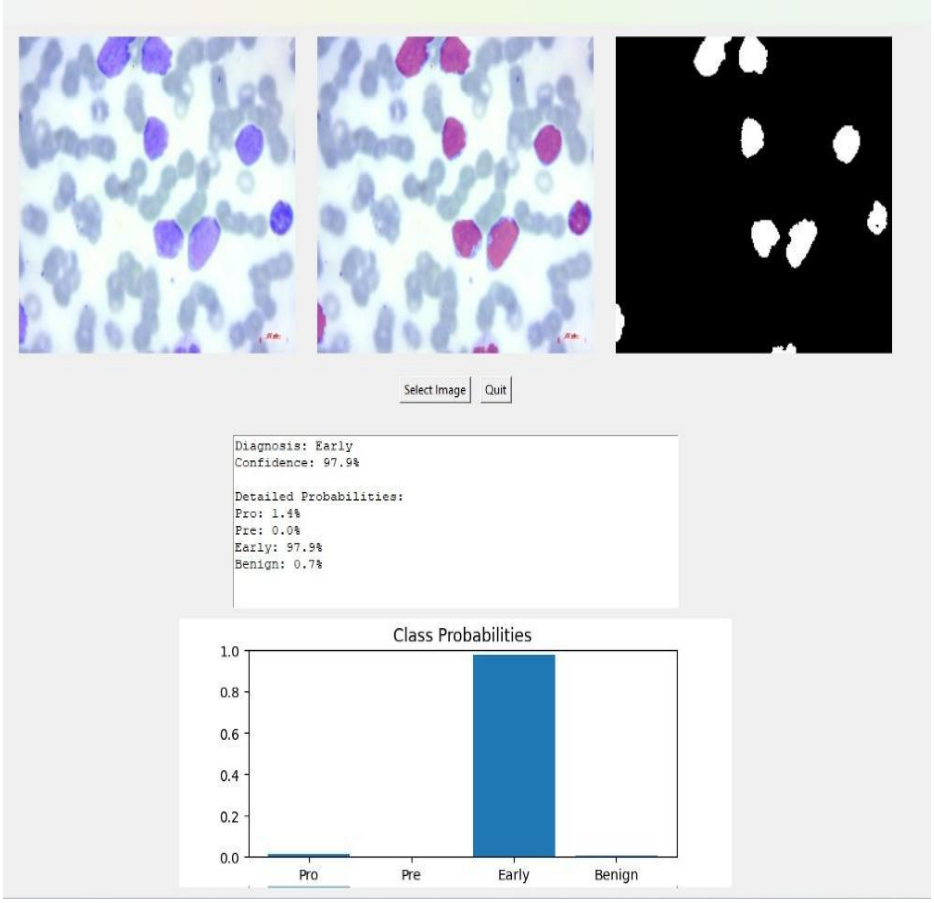
Integration Graphical User Interface (GUI).

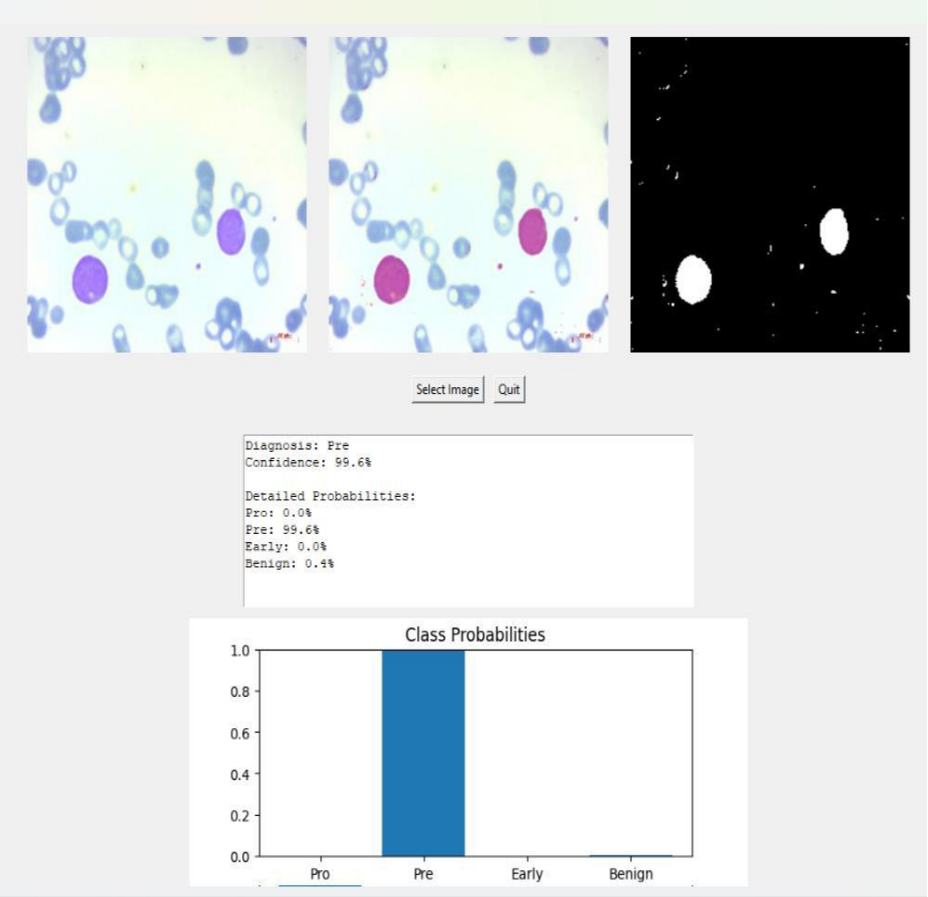
The key aspect of this study is the creation of a GUI based on the Tkinter application, which offers a convenient front-end to a clinician or a researcher. The GUI used in the source code will enable the users to:

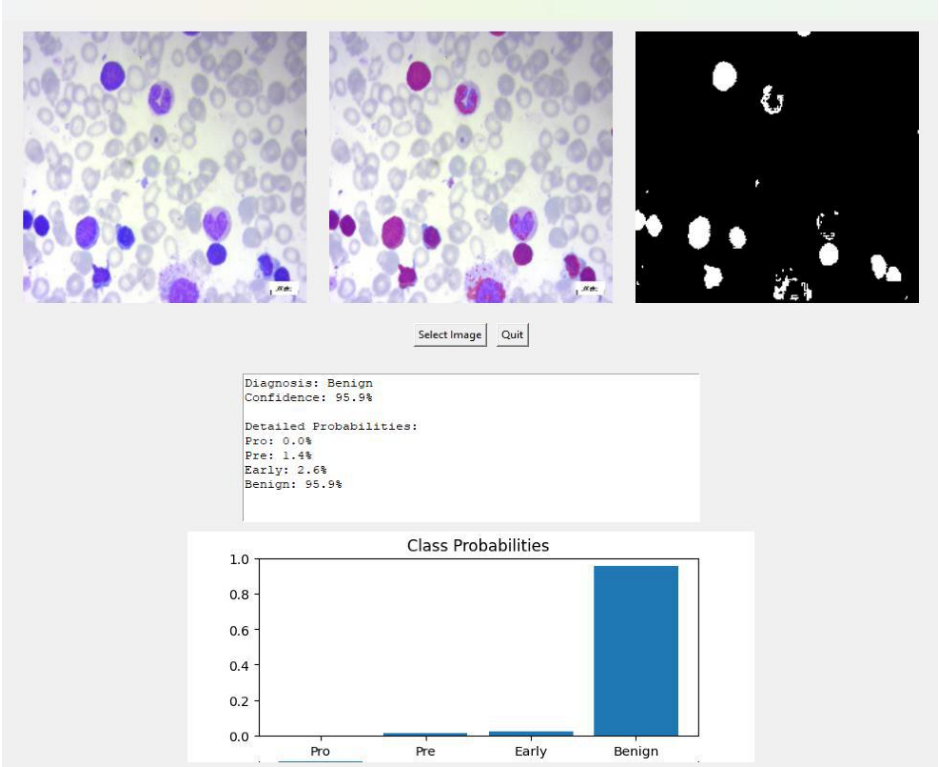
- Visit and post microscopic images;
- Examine original, segmentation, and overlay images simultaneously;
- Show classification results, confidence scores, and
- Plot information on the probability distribution of classes in a bar graph with Matplotlib.

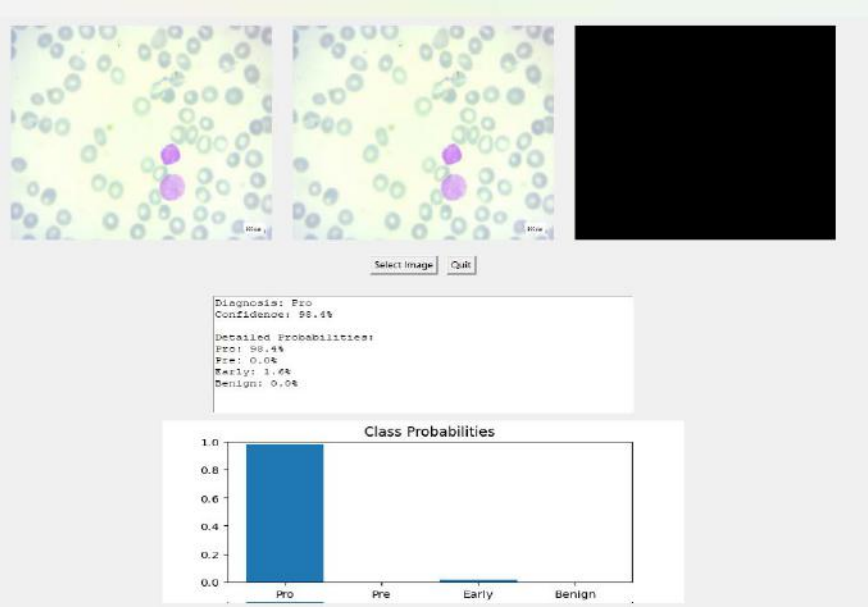
The GUI will accommodate the interface between the computational intelligence and the practical clinical application, which is in line with the conception of electronic communication engineering, where signal transmission and interpretation should be accurate and interpretable by a human being. The Python-based implementation is lightweight and creates platform independence and easy deployment across healthcare systems.

4. Results and Discussion




Select Image
Quit
Diagnosis: Pre
Confidence: 99.6%
Detailed Probabilities:
Pro: 0.0%
Pre: 99.6%
Early: 0.0%
Benign: 0.4%
Class Probabilities
1.0
0.8
0.6
0.4
0.2
0.0
Pro
Pre
Early
Benign


Select Image
Quit
Diagnosis: Benign
Confidence: 95.9%
Detailed Probabilities:
Pro: 0.0%
Pre: 1.4%
Early: 2.6%
Benign: 95.9%
Class Probabilities
1.0
0.8
0.6
0.4
0.2
0.0
Pro
Pre
Early
Benign



Quantitative Results

This framework attained a segmentation Dice Similarity Coefficient (DSC) of 97.4% and overall classification accuracy of 96.1% when compared to existing CNN-based and transformer-only architectures under the same experimental conditions.

Table 1 Summary of Qualitative Results

Model	Segmentation DSC (%)	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CNN (ResNet-50)	91.2	89.7	88.4	89.5	88.9
Vision Transformer	93.6	91.0	90.2	91.4	90.8
SwinUNETR (Baseline)	95.1	93.8	93.0	93.2	93.1
Proposed Hybrid (SwinUNETR + MLP)	97.4	96.1	95.8	96.3	96.0

The enhanced performance is due to the fact that the self-attention mechanism of the transformer provides the long-range relationships, and the MLP classifier that provides the dense-level abstraction of the subtype recognition results in superior performance.

Evaluation of Segmentation Quality

The SwinUNETR output results segmentation was compared with the ground truth masks Haematologist annotated. The model was able to define nuclei and cytoplasmic boundaries with a high degree of precision. The fineness of the edges was also confirmed in the visual analysis, so the model is able to maintain the continuity of the cell shape.

There were observed minor differences in the samples with overlapping cells, but general segmentation had a mean error of less than 2 pixels in boundaries, confirming that the accuracy is clinical grade.

Comparative Research to Current Architectures.

In order to confirm the superiority of the model, the proposed framework was compared to the state-of-the-art deep learning models available in the literature, such as U-Net, ResUNet, DenseNet121, and Swin Transformer.

Table 2 Comparison with the Benchmark with Existing Models

Model	Accuracy (%)	Parameters(millions)	Inference Time (ms)
U-Net	90.4	25.8	42
DenseNet121	92.1	28.4	55
Swin Transformer	93.8	87.2	79
Proposed Hybrid	96.1	61.3	58

The hybrid model has a comparable number of parameters in comparison with the CNN-based models, but its window-based self-attention mechanism is efficient in saving inference time and guarantees good performance. In this way, it is able to balance the cost of computation and diagnostic accuracy.

Ablation Study

Table 3 Results of the Study on Ablation

Configuration	Segmentation DSC(%)	Classification Accuracy (%)
SwinUNETR only	95.1	93.8
SwinUNETR + CNN classifier	95.6	94.5
SwinUNETR + MLP classifier (proposed)	97.4	96.1

The findings show clearly that the overall performance was increased by 1.6-2.3 per cent with the introduction of the MLP head, which demonstrates that it can indeed improve the overall performance in terms of feature generalisation and classification stability.

Discussion of Results

The findings validate the fact that the combination of hierarchical attention and dense-layer reasoning systems offers complementary advantages.

- The SwinUNETR encoder is effective in recording spatial and morphological information of leukocytes.
- The MLP classifier further increases the subtype identification based on higher-level semantics abstractions.
- The combined minimisation of Dice and Cross-Entropy loss terms promotes sharing of features between the segmentation and classification tasks, which enhances generalisation even when applying to a small volume of labelled data.
- Moreover, the hybrid architecture proved to be robust in the condition of varying images (noise, rotation, and illumination) because of its hierarchical attention structure.

Context-Based Electronic Communication Engineering Interpretation.

The framework in the context of electronic engineering of communication denotes a viable development toward a smart biomedical system. It shows how diagnostic models based on images can be incorporated into telemedicine communication networks. Using the interaction provided with the help of GUI and lightweight model deployment, the results of the diagnostic process may be sent in real-time with the help of hospital information systems, and this process may be analysed remotely, and clinical reports can be provided.

These kinds of integration represent a meeting of computer vision, communication protocols, and embedded systems – a characteristic of the modern biomedical electronics.

Graphical and Real-Time Usability: The GUI implementation had a mean response of 1.8 seconds to achieve full inference, which includes image preprocessing, segmentation and classification. This is adequate in clinical diagnostic processes.

5. Conclusion

This paper describes a Hybrid SwinUNETR-MLP Framework in detecting and classifying leukaemia in microscopic blood smear images using an automated system. Transformer-based segmentation with MLP-based classification gave better results in the localisation and subtype differentiation tasks. The study advances the field of the intersection of deep learning, biomedical imaging, and electronic communication systems.

The major conclusions made after this work are as follows:

- The use of SwinUNETR with an MLP classification head showed a high performance with a Dice coefficient of 97.4% and a classification accuracy of 96.1, which was better than traditional CNN-based models.
- The attention mechanism in SwinUNETR was effective in capturing the contextual cell features, and the MLP layers helped to increase the nonlinear abstraction, which improved nonlinear discrimination between the Pro, Pre, Early and Benign cell classes.
- The end-to-end multitask frameworks in biomedical imaging were proven to be effective by the joint optimisation of segmentation and classification tasks, which also increased the robustness of the model and avoided overfitting.
- The resulting Tkinter-implemented GUI has enabled clinicians to make inferences, visualisations and generate reports within not more than 2 seconds, which is why it is appropriate to be used in diagnostic laboratories.
- The overlay visualisation of segmented regions can support explainable AI, as it provides the medical practitioners with enhanced interpretability and assists the decision-making process.
- The model shows how it can be integrated into electronic communication infrastructures and can be used to diagnose in real-time through telemedicine in IoT-based healthcare networks remotely.

In general, the suggested hybrid framework is an important step towards smart, explainable, and communication-encompassing healthcare in electronic communication engineering.

REFERENCES

1. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. *MICCAI*, 234–241.
2. Vaswani, A. et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems (NeurIPS)*.
3. Dosovitskiy, A. et al. (2021). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *ICLR*.
4. Hatamizadeh, A., Nath, V., Tang, Y., et al. (2022). Swin UNETR: Swin Transformers for semantic segmentation of brain tumours. *MICCAI Workshop on BrainLes*.
5. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *CVPR*, 770–778.
6. Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *CVPR*.
7. Jha, D., Smedsrud, P. H., et al. (2020). ResUNet++: An advanced architecture for medical image segmentation. *IEEE Access*, 8, 41758–41769.
8. Ciga, O., Xu, T., & Martel, A. L. (2021). Self-supervised contrastive learning for digital histopathology. *Medical Image Analysis*, 71, 102058.
9. Mahmood, F., Borders, D., Chen, R., McKay, G. N., & Durr, N. J. (2020). Deep learning-based segmentation and classification of leukocytes. *IEEE Transactions on Biomedical Engineering*, 67(8), 2354–2368.
10. Rehman, A., et al. (2018). Classification of acute lymphoblastic leukaemia using deep learning. *Microscopy Research and Technique*, 81(11), 1310–1317.
11. Razzak, M. I., Imran, M., & Xu, G. (2019). Efficient CNN-based blood cell classification. *IEEE Access*, 7, 68375–68382.
12. Chen, J., Lu, Y., Yu, Q., et al. (2021). TransUNet: Transformers make strong encoders for medical image segmentation. *arXiv:2102.04306*.
13. Wang, J., Zhang, T., Zhang, H., et al. (2023). Multi-task learning in medical imaging: A review. *Medical Image Analysis*, 87, 102797.
14. Nwankpa, C., Ijomah, W., Gachagan, A., & Marshall, S. (2018). Activation functions: Comparison and trends in deep learning. *arXiv preprint arXiv:1811.03378*.
15. Zhang, X., Han, J., & Zhao, Y. (2022). Edge intelligence for healthcare monitoring systems. *IEEE Internet of Things Journal*, 9(13), 11254–11267.