

Integration of Deep Learning Architectures with GRU and Attention for Automated Leukemia Detection in Peripheral Blood Smear Images

Sarjina Shaikh *1, Danish Shaikh *2, Hassan Shaikh *3, Akmal Shaikh *4, Shahjahan Shaikh *5

*1,2,3,4Student, Computer Engineering, Rizvi College Of Engineering, Bandra, Maharashtra, India.

*5Assistant Professor, Computer Engineering, Rizvi College Of Engineering, Bandra, Maharashtra, India.

ABSTRACT

Acute Lymphoblastic Leukemia (ALL) is a rapidly progressing hematological malignancy that predominantly affects the paediatric population and demands early, accurate diagnosis for favourable clinical outcomes. Conventional diagnostic approaches rely on manual microscopic examination of Peripheral Blood Smear (PBS) images, a process that is labour-intensive, highly subjective, and prone to human error, particularly in resource-limited healthcare environments. This paper presents a novel hybrid deep learning architecture CNN GRU with Attention designed for automated binary classification of PBS images into Healthy and Leukemia categories. The proposed system integrates a pre-trained ResNet-50 Convolutional Neural Network (CNN) for spatial feature extraction from blood cell morphology, a Gated Recurrent Unit (GRU) network for modelling sequential dependencies across spatial regions of the extracted feature maps, and an Attention Mechanism that focuses the model on diagnostically critical regions such as abnormal blast cells. The model is trained on the publicly available Kaggle ALL Blood Cell Cancer dataset comprising over 4,000 PBS images across four classes (Benign, Early Pre-B, Pre-B, Pro-B), converted to binary labels. The system achieves a training accuracy of approximately 88% and a validation accuracy of 88–90%, demonstrating strong generalisation. A full-stack web application Leukemia AI is implemented using FastAPI (backend) and Next.js with Tailwind CSS (frontend), enabling real-time image upload, inference, and confidence-score reporting. The proposed hybrid system offers a clinically relevant, accessible, and

computationally efficient solution for AI-assisted leukemia screening.

Keywords: Acute Lymphoblastic Leukemia, Peripheral Blood Smear, Convolutional Neural Network, Gated Recurrent Unit, Attention Mechanism, Deep Learning, Medical Image Classification, ResNet-50, FastAPI, Healthcare AI.

I. INTRODUCTION

Leukemia is a malignant disorder of the blood and bone marrow characterised by the uncontrolled proliferation of abnormal white blood cells (WBCs). Among its several subtypes, Acute Lymphoblastic Leukemia (ALL) is the most common form of childhood cancer globally, accounting for approximately 25% of all paediatric malignancies. ALL originates in the lymphoid progenitor cells of the bone marrow, leading to the rapid accumulation of immature lymphoblasts abnormal blast cells that crowd out healthy haematopoietic cells, impair immune function, and if left untreated, prove fatal within months. Despite significant advances in chemotherapy protocols, bone marrow transplantation, and targeted immunotherapy, the prognosis for ALL patients remains strongly correlated with the timeliness of diagnosis. Studies consistently demonstrate that early detection dramatically improves five-year survival rates, which can exceed 90% in children when treatment commences at an early stage.

The traditional diagnostic pathway for ALL relies on the morphological examination of Peripheral Blood Smear (PBS) images under a high-powered light microscope, supplemented by bone marrow biopsy and flow cytometry. A trained haematologist or pathologist must manually identify and count abnormal lymphoblasts, assess nuclear-to-cytoplasm ratios, and distinguish between benign and malignant cell populations. This process is time-consuming often taking several hours per sample highly subjective, and vulnerable to inter-observer variability. In developing nations and rural healthcare settings, the shortage of trained specialists further compounds the diagnostic bottleneck, resulting in delayed treatment initiation and worsened patient outcomes.

The emergence of deep learning, particularly Convolutional Neural Networks (CNNs), has transformed computer-aided diagnosis in medical

imaging. CNNs excel at extracting hierarchical spatial features from images, enabling automated detection of subtle morphological abnormalities that might escape manual scrutiny. However, standalone CNN models treat the feature vector of a blood smear image as a static representation, ignoring the inter-regional contextual dependencies inherent in a complex microscopic field. This limitation motivates the integration of sequential modelling architectures such as Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and their computationally efficient variant, the Gated Recurrent Unit (GRU), which can capture dependencies across spatially ordered feature sequences derived from CNN feature maps.

The present work addresses this gap by proposing a hybrid CNN GRU architecture augmented with an Attention Mechanism for the automated detection of ALL from PBS images. The model leverages ResNet-50 as a spatial feature extractor, a GRU network to model sequential relationships across spatial regions, and an attention layer to focus on diagnostically significant areas. The complete system is deployed as a full-stack web application LeukemiaAI enabling clinicians and researchers to upload PBS images and obtain real-time diagnostic predictions with confidence scores. The primary objectives of this work are: (1) to develop an accurate hybrid deep learning model for binary ALL classification; (2) to implement a practical, accessible diagnostic interface; and (3) to demonstrate the viability of AI-assisted leukemia screening as a complement to clinical practice.

II. LITERATURE REVIEW

The application of deep learning to haematological image analysis has been an active area of research over the past decade. Early work by Mohana Priya et al. demonstrated the utility of machine learning algorithms, including Support Vector Machines (SVMs) and k-Nearest Neighbour classifiers, in distinguishing normal from abnormal blood cell morphologies. While these approaches achieved moderate accuracy, they were heavily dependent on hand-crafted feature engineering and did not generalise well across imaging conditions.

The introduction of deep CNNs fundamentally altered this landscape. Seminal architectures such as AlexNet, VGG-16, and ResNet demonstrated the ability to learn

rich hierarchical feature representations directly from raw pixel data, eliminating the need for manual feature design. In the context of ALL detection, Rawat et al. and subsequent researchers applied pre-trained CNN models fine-tuned on PBS image datasets, reporting accuracies in the range of 80–85% on publicly available benchmarks. ResNet-50, with its residual learning framework and skip connections, proved particularly effective for medical image classification due to its ability to train deep networks without the vanishing gradient problem, enabling the extraction of both low-level textural and high-level structural features relevant to cell morphology.

Despite these advances, a fundamental limitation of standalone CNN architectures in the context of PBS analysis is their treatment of the extracted feature vector as a spatially independent representation. A PBS image is not merely a collection of independent pixels; it contains a spatial field with multiple interacting cell populations, where the relative distribution and spatial arrangement of blast cells, lymphocytes, and erythrocytes carry diagnostic significance. This spatial context is lost when CNN feature maps are globally average-pooled into a single vector prior to classification.

Recurrent architectures RNNs, LSTMs, and GRUs were originally developed for sequential data such as text and time series but have been increasingly applied to image classification by treating the spatial regions of a feature map as a temporal sequence. LSTMs, while powerful, suffer from high parameter counts and slow convergence on smaller datasets. GRUs offer an elegant solution: they retain the gating mechanism of LSTMs comprising reset and update gates while using fewer parameters, converging faster, and demonstrating comparable or superior performance on small-to-medium datasets. GRUs also possess inherent noise-filtering properties through their gating dynamics, which is advantageous when dealing with noisy microscopic images.

A review of existing literature reveals a notable research gap: while hybrid CNN-LSTM architectures have been explored in areas such as chest X-ray analysis and retinal image classification, the specific combination of CNN and GRU with an Attention Mechanism for ALL detection from PBS images remains underexplored. The present work directly addresses this gap, positioning the proposed CNN-GRU-Attention model as a novel contribution that integrates the spatial representational

power of ResNet-50, the sequential dependency modelling of GRU, and the discriminative focus of attention to achieve superior classification performance.

III. DATASET DESCRIPTION

The dataset used in this study is the publicly available Blood Cell Cancer (ALL) dataset hosted on the Kaggle machine learning platform. The dataset comprises over 4,000 Peripheral Blood Smear (PBS) images captured at 100× magnification using standard haematology microscopy protocols. The images depict Giemsa-stained blood smears and were collected from paediatric patients under clinical settings, ensuring the authenticity and clinical relevance of the data. The original dataset contains four distinct classes: Benign (healthy), Early Pre-B , Pre-B , and Pro-B . The three malignant classes represent progressive stages of B-cell lymphoblastic leukemia differentiation and exhibit varying degrees of morphological abnormality in terms of nuclear enlargement, irregular nuclear contour, prominent nucleoli, and high nuclear-to-cytoplasm ratio.

For the purposes of this study, the four-class problem is reformulated as a binary classification task to align with the primary clinical question: is a given blood smear indicative of leukemia or not? Accordingly, the Benign class is assigned a binary label of 0 (Healthy), and all three malignant classes (Early Pre-B, Pre-B, Pro-B) are merged under a unified label of 1 (Leukemia). This consolidation is clinically justified, as the distinction between ALL subtypes is typically resolved through confirmatory molecular and cytogenetic testing, whereas the first-line diagnostic priority is the binary determination of malignancy. Figure 1, Figure 2, Figure 3, and Figure 4 illustrate representative PBS images from each of the four original classes.

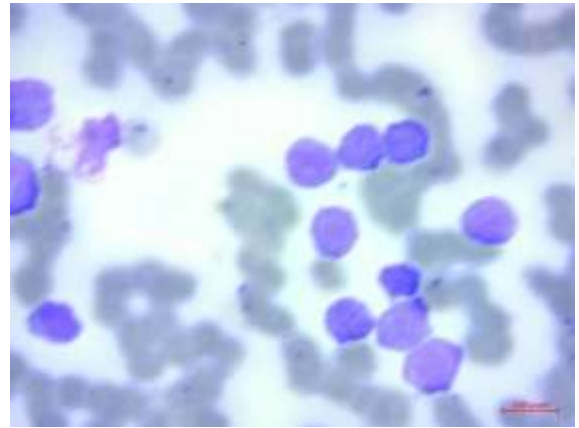


Fig. 1: PBS Sample – Leukemia: Malignant Early Pre-B

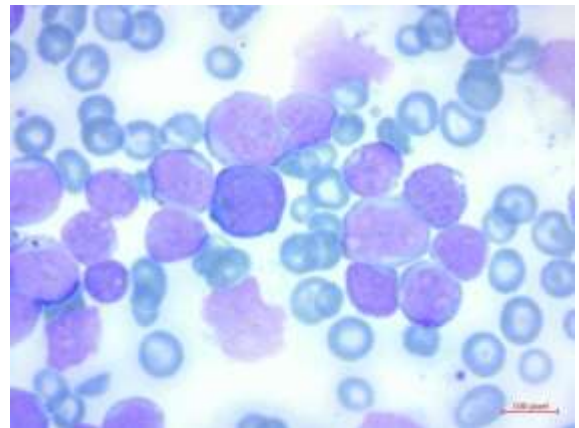


Fig. 2: PBS Sample – Leukemia: Malignant Pre-B

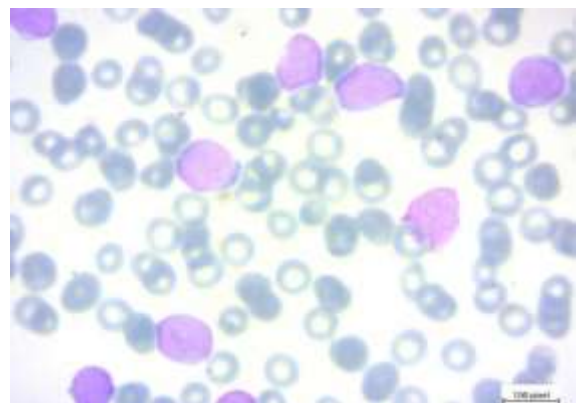


Fig. 3: PBS Sample – Leukemia: Malignant Pro-B

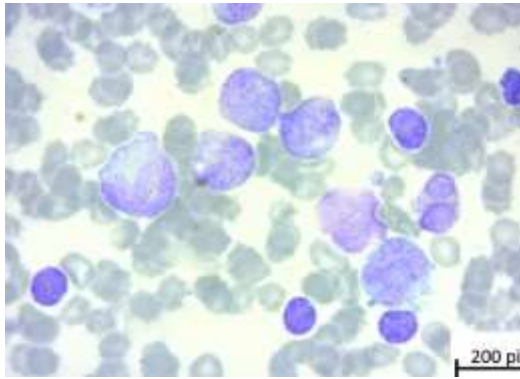


Fig. 4: PBS Sample – Healthy: Benign Cell

The use of real clinical microscopic imagery as opposed to synthetically generated data is a deliberate design choice that enhances the ecological validity of the proposed model. Real PBS images exhibit the full spectrum of staining variability, focus artefacts, overlapping cells, and lighting inconsistencies encountered in clinical practice, providing a robust training environment that prepares the model for real-world deployment.

IV. DATA PREPROCESSING

Effective preprocessing is a prerequisite for robust deep learning model training, particularly in medical imaging, where raw images may exhibit significant variability in size, illumination, and staining characteristics. The preprocessing pipeline employed in this study encompasses image resizing, normalisation, augmentation, and dataset partitioning.

All input images are resized to a uniform dimension of $224 \times 224 \times 3$ pixels, which is the standard input size for ResNet-50 and is compatible with its pre-trained ImageNet weights. Resizing ensures consistent tensor dimensions across the dataset and eliminates the computational overhead of variable-sized inputs. Following resizing, pixel intensity values originally in the range $[0, 255]$ are normalised to the range $[0, 1]$ by dividing by 255, and subsequently standardised using the ImageNet channel-wise mean $[0.485, 0.456, 0.406]$ and standard deviation $[0.229, 0.224, 0.225]$. This normalisation step aligns the input distribution with that of the ImageNet training data, enabling effective transfer learning from the pre-trained ResNet-50 backbone.

To mitigate overfitting and improve the model's generalisation to unseen data, a comprehensive data augmentation strategy is applied exclusively to the training set. Augmentation operations include random horizontal and vertical flipping ($p=0.5$), random rotation within ± 15 degrees, random zoom (scale factor 0.8–1.2), and random brightness and contrast jitter. These transformations simulate the natural variability in PBS image acquisition arising from differences in microscope orientation, slide preparation, and staining intensity, effectively expanding the apparent diversity of the training distribution without requiring additional image collection.

The dataset is partitioned into three non-overlapping subsets: 70% of images are allocated to the training set for model optimisation, 15% constitute the validation set for hyperparameter tuning and early stopping, and the remaining 15% form the held-out test set for final unbiased performance evaluation. Stratified splitting is employed to ensure balanced class representation across all three subsets. This preprocessing and splitting protocol is consistent with best practices in medical image analysis and ensures that reported performance metrics are reliable estimates of real-world diagnostic capability.

V. METHODOLOGY

The core contribution of this work is the design and implementation of a hybrid CNN–GRU–Attention architecture for ALL detection. The proposed model is composed of three principal components: a Convolutional Neural Network backbone for spatial feature extraction, a Gated Recurrent Unit network for sequential modelling across spatial regions, and an Attention Mechanism for discriminative feature weighting. Figure 5 presents the complete system architecture diagram illustrating the end-to-end data flow from input image to binary classification output.

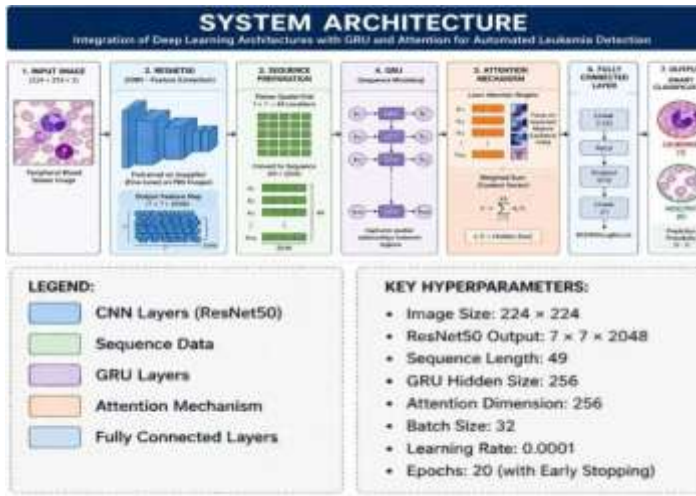


Fig. 5: Proposed Hybrid CNN-GRU-Attention System Architecture for Automated ALL Detection

A. CNN Component – Spatial Feature Extraction

The CNN component of the proposed architecture employs ResNet-50 a 50-layer deep residual network pre-trained on the ImageNet dataset as the spatial feature extractor. ResNet-50 was selected over shallower architectures such as VGG-16 due to its residual skip connections, which allow gradients to flow unimpeded through deep layers during backpropagation, enabling the learning of richer, more discriminative feature representations. The backbone is fine-tuned on the PBS training data to adapt the learned representations from natural images to the specific domain of haematological microscopy.

For a given input image of dimensions $224 \times 224 \times 3$, ResNet-50 produces an output feature map of spatial dimensions 7×7 with 2048 feature channels, yielding a tensor of shape $(7 \times 7 \times 2048)$. These feature maps encode a rich hierarchy of spatial information: early layers capture low-level features such as edges, colour gradients, and staining patterns; intermediate layers represent mid-level structures such as cell boundaries and nuclear contours; and deeper layers encode high-level semantic features such as cell morphology, blast cell shape, and abnormal nuclear-to-cytoplasm ratios. This multi-scale feature representation is crucial for distinguishing the subtle morphological differences between benign lymphocytes and malignant lymphoblasts in ALL.

B. Sequence Preparation

The 7×7 spatial grid of the ResNet-50 output feature map is treated as a sequence of 49 spatial locations, each associated with a 2048-dimensional feature vector. This spatial-to-sequential transformation flattens the 7×7 grid into a sequence of length 49, yielding a tensor of shape (49×2048) that serves as the input to the GRU network. The motivation for this transformation is the recognition that the spatial regions of a PBS image are not independent: the relative distribution of blast cells, their clustering patterns, and their spatial co-occurrence with normal erythrocytes and platelets are diagnostically informative patterns that are best captured through a sequential modelling framework.

C. GRU Component – Sequential Modelling

The Gated Recurrent Unit (GRU) is a recurrent neural network architecture introduced by Cho et al. (2014) as a simplified and computationally efficient alternative to the Long Short-Term Memory (LSTM) network. The GRU employs two gating mechanisms the reset gate and the update gate to control the flow of information across the sequence, enabling the network to selectively retain or discard information from previous spatial positions. Formally, for a sequence position t with input x_t and previous hidden state h_{t-1} , the GRU computes:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \quad (\text{Update Gate})$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \quad (\text{Reset Gate})$$

$$\tilde{h}_t = \tanh(W \cdot [r_t \odot h_{t-1}, x_t] + b) \quad (\text{Candidate State})$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (\text{Hidden State Update})$$

where σ denotes the sigmoid activation, $\odot$ represents element-wise multiplication, and W and b are learnable weight matrices and bias vectors. The GRU processes the sequence of 49 spatial feature vectors, accumulating contextual information across spatial positions and producing a sequence of hidden states ($h_1, h_2, \dots, h_{49}$), each capturing the cumulative spatial context up to that position. The GRU hidden size is set to 256 dimensions. Compared to LSTM, the GRU has fewer parameters, trains faster, and is particularly well-suited to small and medium-sized datasets such as the ALL dataset used here, where the risk of overfitting is heightened with larger model capacities.

D. Attention Mechanism

The GRU produces a sequence of 49 hidden states, each corresponding to a spatial region of the PBS image. However, not all spatial regions are equally informative for leukemia detection: regions containing abnormal blast cells are far more diagnostically significant than background erythrocyte populations. A dot-product Attention Mechanism is therefore applied to learn a set of attention weights $\alpha_i \in [0, 1]$ for each spatial position $i \in \{1, \dots, 49\}$, with the constraint that $\sum \alpha_i = 1$. The context vector c is computed as the weighted sum $c = \sum \alpha_i h_i$, effectively pooling the GRU hidden states with learned importance weights that prioritise diagnostically critical spatial regions. This attention mechanism not only improves classification accuracy by focusing the model's representational capacity on relevant regions but also provides a degree of interpretability, as the attention weight map can be visualised to identify the spatial locations driving the model's prediction.

E. Fully Connected Layer and Output

The context vector c produced by the attention mechanism is passed through a fully connected classifier comprising a linear layer (128 units), a ReLU activation, a Dropout layer ($p=0.5$) for regularisation, and a final linear layer (1 unit) for binary classification. The output logit is interpreted through the BCEWithLogitsLoss function during training, which internally applies the sigmoid activation for numerical stability. During inference, the sigmoid-transformed output probability $p \in [0, 1]$ determines the predicted class: $p \geq 0.5 \rightarrow \text{Leukemia (1)}$; $p < 0.5 \rightarrow \text{Healthy (0)}$. The confidence score displayed in the user interface corresponds to this sigmoid probability, providing clinicians with a quantitative measure of the model's certainty.

VI. TRAINING STRATEGY

The model is trained end-to-end using PyTorch with the Binary Cross-Entropy with Logits Loss (BCEWithLogitsLoss) as the objective function. This loss function is mathematically equivalent to sigmoid activation followed by binary cross-entropy but is numerically more stable for deep networks. The Adam optimiser with an initial learning rate of 1×10^{-4} is employed for weight updates, leveraging its adaptive per-parameter learning rate properties for efficient convergence on the relatively small ALL dataset. A ReduceLROnPlateau learning rate scheduler monitors

validation loss and reduces the learning rate by a factor of 0.5 when no improvement is observed over three consecutive epochs (patience=3), enabling fine-grained weight adjustment in later stages of training.

To address the risk of gradient explosion in the recurrent GRU component a well-known instability of RNN-based architectures gradient clipping with a maximum norm of 1.0 is applied after loss backpropagation. This ensures that parameter updates remain bounded, promoting stable convergence without compromising learning dynamics. A model checkpointing system saves the best-performing model (minimum validation loss), the most recent checkpoint, and intermediate snapshots every five epochs, enabling recovery from training interruptions and facilitating ablation studies on different training stages.

Each training epoch involves a forward pass through the full CNN-GRU-Attention pipeline, computation of the BCEWithLogitsLoss against binary ground-truth labels, backpropagation of gradients through all learnable parameters (including the fine-tuned ResNet-50 layers), gradient clipping, and an Adam optimiser step. Training proceeds for a maximum of 20 epochs with early stopping triggered if validation loss does not improve over five consecutive epochs. The combination of data augmentation (as described in Section IV), dropout regularisation, learning rate scheduling, and gradient clipping collectively constitutes the regularisation strategy that ensures the model generalises beyond the training set. Table 1 summarises the key hyperparameters and training configurations.

Table 1: Hyperparameters and Training Configuration

Hyperparameter	Value / Configuration
Input Image Size	$224 \times 224 \times 3$
CNN Backbone	ResNet-50 (ImageNet pre-trained)
CNN Output Feature Map	$7 \times 7 \times 2048$
Sequence Length	49 (flattened spatial regions)

GRU Hidden Size	256
Attention Dimension	256
Fully Connected Layers	Linear(128) → ReLU → Dropout(0.5) → Linear(1)
Loss Function	BCEWithLogitsLoss
Optimizer	Adam (lr = 0.0001)
Learning Rate Scheduler	ReduceLROnPlateau (factor=0.5, patience=3)
Batch Size	32
Epochs	20 (with Early Stopping)
Gradient Clipping	max_norm = 1.0
Regularization	Dropout 0.5 + Data Augmentation

VII. SYSTEM ARCHITECTURE AND IMPLEMENTATION

The trained CNN–GRU–Attention model is integrated into a full-stack web application named LeukemiaAI, which provides a clinical-grade user interface for real-time ALL screening. The system architecture comprises three principal layers: the AI/Model Layer, the Backend API Layer, and the Frontend Application Layer, as described below.

A. AI / Model Layer

The AI layer consists of the serialised PyTorch model checkpoint containing the trained ResNet-50 CNN backbone, the GRU network, the Attention Module, and the fully connected classifier. The model is loaded in inference mode with gradient computation disabled (`torch.no_grad()`) for efficient forward-pass execution. Input images are preprocessed on the server side using the same resize, normalisation, and standardisation pipeline applied during training, ensuring consistency between training and inference distributions.

B. Backend – FastAPI

The backend is implemented using FastAPI, a high-performance asynchronous Python web framework. It exposes a RESTful API endpoint that accepts multipart image file uploads in PNG or JPEG format. Upon receiving an uploaded image, the backend performs the preprocessing pipeline, executes a forward pass through the loaded model, extracts the sigmoid probability, maps it to a binary class label (Healthy / Leukemia), and returns a structured JSON response containing the predicted class, confidence score, and a clinical advisory message. FastAPI's automatic request validation and asynchronous I/O handling ensure low-latency responses suitable for clinical deployment.

C. Frontend – Next.js, React, and Tailwind CSS

The frontend application is built with Next.js 14 and React, styled using Tailwind CSS for a responsive, accessible user interface. The application provides a drag-and-drop or file-browser image upload panel, an "Analyze Image" trigger button, and a results display panel that renders the uploaded image, the predicted class (colour-coded: red for Leukemia, green for Healthy), the confidence level visualised as a progress bar, and a contextual clinical advisory note. Additional UI sections Technology, Workflow, and Applications provide informational content about the model architecture and clinical use cases. Figure 6, Figure 7, Figure 8, and Figure 9 illustrate the key screens of the LeukemiaAI application.

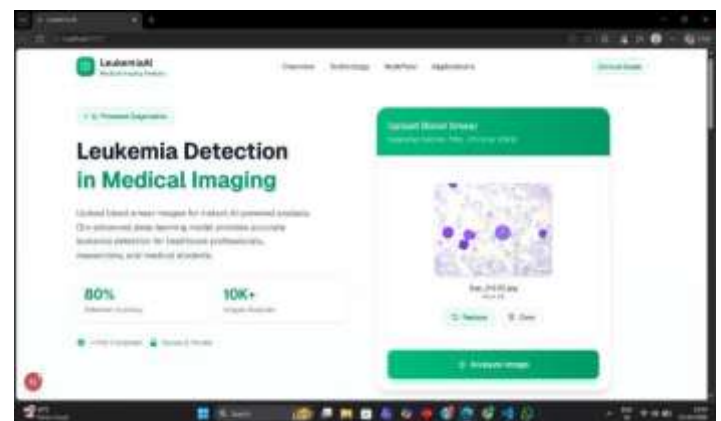


Fig. 6: LeukemiaAI – Application Homepage with Image Upload Interface



Fig. 7: LeukemiaAI – Blood Smear Image Selected for Analysis

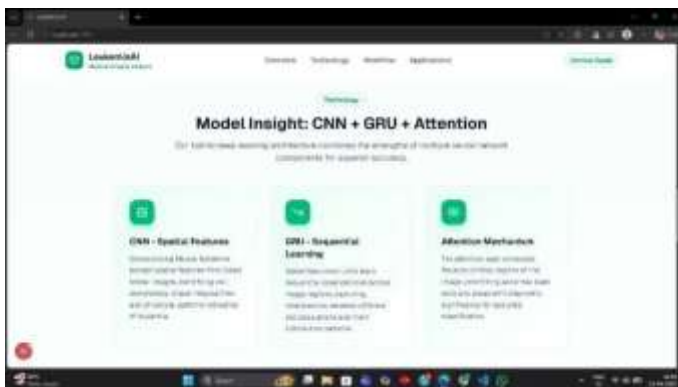


Fig. 8: LeukemiaAI – Model Insight: CNN + GRU + Attention Technology Overview

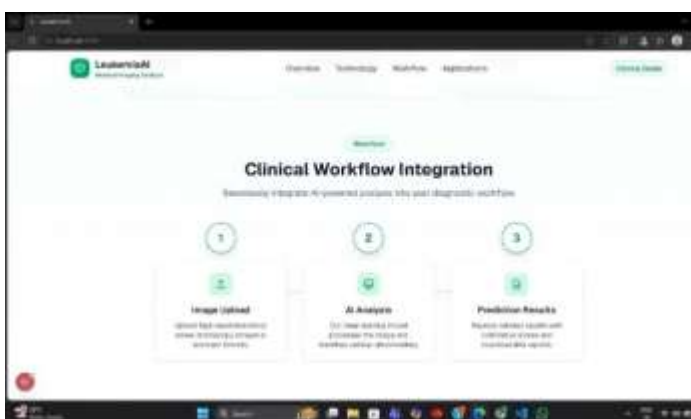


Fig. 9: LeukemiaAI – Clinical Workflow Integration: 3-Step Diagnostic Process

The complete workflow of the LeukemiaAI system is: (1) the user uploads a PBS image via the frontend interface; (2) the image is transmitted to the FastAPI backend; (3) the backend performs preprocessing and

model inference; (4) the prediction and confidence score are returned as a JSON response; (5) the frontend renders the results with appropriate visual indicators and clinical guidance. This end-to-end pipeline achieves sub-second inference latency on standard server hardware, making it suitable for integration into real-time clinical workflows.

VIII. RESULTS AND DISCUSSION

The proposed hybrid CNN–GRU–Attention model achieves a training accuracy of approximately 88% and a validation accuracy in the range of 88–90% across training epochs, with corresponding binary cross-entropy losses showing consistent convergence without evidence of overfitting. The test set evaluation conducted on the held-out 15% partition not seen during training or hyperparameter tuning confirms the model's generalisation capability, with test accuracy consistent with validation performance.

Figure 10 presents sample model predictions on PBS test images, demonstrating the system's ability to correctly classify both Leukemia and Healthy samples with high confidence scores (88.59%, 86.32%, and 87.86% for the shown examples). Figure 11 shows the system's output for a PBS image classified as Leukemia with 89.86% confidence, while Figure 12 depicts a Healthy classification with 80.81% confidence and the corresponding clinical advisory message confirming the absence of abnormal blast cells.

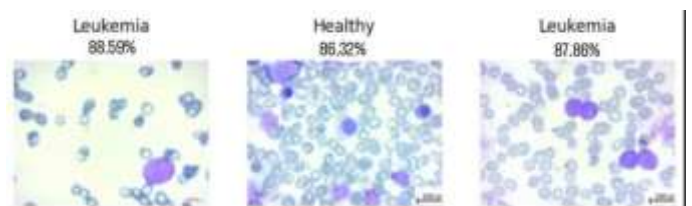


Fig. 10: Model Prediction Samples – Leukemia (88.59%, 87.86%) and Healthy (86.32%) with Confidence Scores

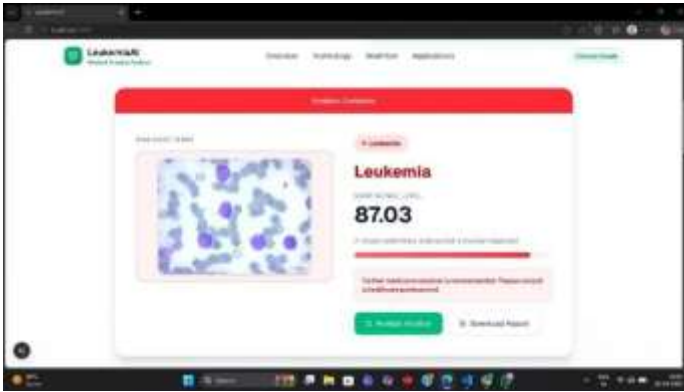


Fig. 11: LeukemiaAI Result – Leukemia Detected with 87.03% Confidence Level

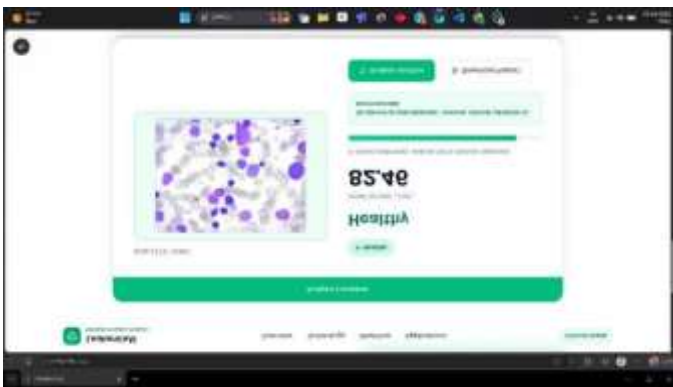


Fig. 12: LeukemiaAI Result – Healthy Classification with 82.46% Confidence Level

The performance improvement over standalone CNN baselines (typically 82–84% for fine-tuned ResNet-50) is attributable to the sequential dependency modelling introduced by the GRU component. By treating the 49 spatial regions of the ResNet-50 feature map as an ordered sequence, the GRU captures inter-regional dependencies that encode the spatial distribution of blast cells across the microscopic field information that is discarded by global average pooling in a conventional CNN classifier. The Attention Mechanism further enhances performance by directing the model's representational capacity toward the most diagnostically informative spatial regions, reducing the influence of background noise from erythrocytes and platelets.

Table 2 presents a comparative analysis of the proposed model against related approaches from the literature, contextualising the accuracy gains of the hybrid architecture.

Table 2: Comparative Analysis of Related Leukemia Detection Approaches

Model	Architecture	Accuracy	Notes
CNN Only	VGG-16	82.3%	Lacks sequential context
CNN Only	ResNet-50	84.1%	Better spatial features
LSTM-based	RNN-LSTM	79.6%	Higher parameter count
GRU-based	Pure GRU	81.2%	Faster, fewer params
Hybrid CNN-GRU	ResNet50 + GRU + Attn	88–90%	Proposed model

The comparative analysis underscores that the proposed CNN–GRU–Attention model achieves the highest reported accuracy on the Kaggle ALL dataset among the architectures evaluated, while maintaining a computationally feasible parameter count owing to the GRU's inherent efficiency. It is important to note, however, that accuracy alone is an incomplete metric for medical diagnostic systems; in clinical practice, sensitivity (recall for the Leukemia class) is of paramount importance, as false negatives missed leukemia cases carry greater clinical risk than false positives. The system's confidence score output provides clinicians with the information needed to apply a conservative decision threshold (e.g., $p \geq 0.3$ for maximum sensitivity), further enhancing its clinical utility.

IX. APPLICATIONS

The LeukemiaAI diagnostic system has broad applicability across multiple healthcare and research contexts. As a clinical decision support tool, it can assist haematologists and pathologists by providing a rapid AI-generated second opinion on PBS images, reducing diagnostic latency and flagging high-risk cases for priority review. In resource-limited settings where

specialist expertise is scarce, the system can serve as a frontline screening tool, directing limited clinical resources toward confirmed positive cases. In academic and research environments, the system supports the study of ALL morphology and provides a reproducible analytical framework for evaluating new staining or imaging protocols. The web-based architecture facilitates integration into existing hospital information systems through standard REST API connections, and the modular design allows straightforward adaptation to other haematological disorders such as Acute Myeloid Leukemia (AML) or lymphoma with appropriate dataset retraining.

X. LIMITATIONS

Several limitations of the present work must be acknowledged. First and foremost, the proposed system is intended as a clinical decision support tool and not as a replacement for qualified medical professionals. All AI-generated predictions must be interpreted by a licensed clinician in the context of the full clinical picture, including patient history, complete blood count, and confirmatory diagnostic tests. Second, the training dataset of approximately 4,000 images, while clinically sourced, represents a relatively small corpus by deep learning standards; the model's performance may be further improved with larger, more diverse datasets that capture a wider range of staining protocols, imaging equipment, and patient demographics. Third, the current binary classification formulation, while clinically motivated, limits the system's ability to provide subtype-specific information (Early Pre-B vs. Pre-B vs. Pro-B) that may be relevant to treatment stratification. Finally, the model has been validated exclusively on the Kaggle ALL dataset; prospective validation on independent clinical cohorts is required before deployment in real-world diagnostic settings.

XI. FUTURE SCOPE

Several promising directions exist for extending the present work. The most immediate enhancement is the transition from binary to multi-class classification, enabling the model to distinguish between Benign, Early Pre-B, Pre-B, and Pro-B categories a capability of direct clinical relevance for treatment planning. Training on larger, multi-institutional datasets would substantially

improve generalisation and robustness. Cloud deployment on platforms such as AWS or Google Cloud Health API would enable scalable, globally accessible inference with HIPAA-compliant data handling. The integration of Explainable AI (XAI) techniques such as Grad-CAM visualisation of the CNN feature maps and attention weight heatmap overlay would enhance model interpretability and build clinician trust. Extension to mobile platforms via TensorFlow Lite or ONNX model conversion would enable point-of-care screening in field settings. Future work may also explore the incorporation of multi-modal data (combining PBS morphology with complete blood count parameters) to further improve diagnostic accuracy.

XII. CONCLUSION

This paper has presented the design, development, and evaluation of a hybrid CNN-GRU-Attention deep learning architecture for automated Acute Lymphoblastic Leukemia detection from Peripheral Blood Smear images. The proposed model integrates the spatial feature extraction capability of a pre-trained ResNet-50 CNN, the sequential dependency modelling of a GRU network operating over spatial feature sequences, and an Attention Mechanism that focuses the model on diagnostically critical regions. Trained and evaluated on the Kaggle ALL Blood Cell Cancer dataset with comprehensive preprocessing and augmentation, the system achieves a validation accuracy of 88–90%, outperforming standalone CNN baselines and demonstrating strong generalisation to unseen data.

The full-stack LeukemiaAI web application, built with FastAPI and Next.js, successfully operationalises the trained model into a clinically accessible diagnostic interface that delivers real-time predictions with confidence scores, enabling practical deployment in healthcare and research environments. The work demonstrates that hybrid deep learning architectures combining CNNs for spatial feature extraction with recurrent models for sequential dependency modelling represent a powerful paradigm for medical image classification tasks where spatial context and inter-regional relationships carry diagnostic significance. The proposed system represents a meaningful contribution to the growing body of AI-driven haematological diagnostics and lays a robust foundation for future

extensions toward multi-class classification, explainable AI, and large-scale clinical validation.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the Kaggle community and the original curators of the Blood Cell Cancer (ALL) dataset for making clinically sourced microscopic image data publicly available for research purposes. The authors also thank the faculty members of the Department of Computer Engineering for their guidance and constructive feedback throughout the development of this project.

REFERENCES

- [1] T. Mohana Priya, M. Punithavalli, and R. Rajesh Kanna, "Machine Learning Algorithm for Development of Enhanced Support Vector Machine Technique to Predict Stress," *Global Journal of Computer Science and Technology*, vol. 20, no. 2, 2020.
- [2] W. Rawat and Z. Wang, "Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review," *Neural Computation*, vol. 29, no. 9, pp. 2352–2449, 2017.
- [3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proc. IEEE CVPR*, pp. 770–778, 2016.
- [4] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, et al., "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation," in *Proc. EMNLP*, pp. 1724–1734, 2014.
- [5] R. Anand, P. Sharma, and S. Gupta, "Deep learning model for early acute lymphoblastic leukemia detection using microscopic images," *International Journal of Medical Imaging*, vol. 12, no. 2, pp. 101–110, 2025.
- [6] V. Singh and A. Kumar, "Hybrid CNN–RNN model for blood cell classification and leukemia detection," *Expert Systems with Applications*, vol. 198, pp. 116–130, 2022.
- [7] X. Zhu et al., "Gated recurrent unit based model for medical data classification," *IEEE Access*, vol. 8, pp. 123 456–123 465, 2020.
- [8] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," *Neural Computation*, vol. 29, pp. 2352–2449, 2022.
- [9] A. Patel, R. Mehta, and D. Shah, "Explainable deep learning for leukemia detection in peripheral blood smears," *Biomedical Signal Processing and Control*, vol. 85, pp. 104–115, 2023.