HematoFusion: A Weighted Residual-Vision Transformer Ensemble for Automated Classification of Haematologic Disorders in Microscopic Blood Images

Mouna SAADALLAH[1*], Latefa OULLADJI[1] and Farah BEN-NAOUM[1]

[1]Evolutionary Engineering and Distributed Information Systems Laboratory, Department of Computer Science, Djillali Liabes University, Sidi Bel Abbes, Algeria
mouna.saadallah@univ-sba.dz (*Corresponding author*), latifa.oulladji@univ-sba.dz, farah.bennaoum@univ-sba.dz

1- Related Work on Cancer Classification

Cancer Type	Dataset	Method	Authors	Classes	Accuracy (%)
RBC Abnormalities	Not shared	Neural Network	Kim et al. (2000)	15 RBC, 5 WBC	91
WBC Subtypes	Not shared	Feed forward back propagation neural network	Othman et al. (2017)	5	96
WBC Subtypes	Online datasets	VGG16-ViT	Ali et al. (2025)	5	98.99, 99.95
WBC Subtypes	Kaggle (Mooney, 2018)	DenseNet121	Bozkurt (2021)	5	98
WBC Subtypes	ImageNet, BCCD (Shenggan, 2019)	Two-Module CNN	Yao et al. (2021)	5	95.7, 94.5, 91.6
Leukemia	ALL-IDB, ASH ImageBank, JUST	CNN (AlexNet, DenseNet, ResNet, VGG16)	Areen et al. (2024)	N/A	94
ALL	ALL-IDB1, ALL-IDB2	DeepLeukNet (CNN)	Saeed et al. (2024)	N/A	99.61
Leukemia Subtypes	ALL-IDB, Munich AML Morphology	CNN, RandomForest, SVM, XGBoost	Kasim et al. (2025)	N/A	88
ALL	N/A	Vision Transformer	Swain et al. (2025)	N/A	99.67
ALL	ALL image dataset, Kaggle	Vision Transformer	Prasad et al. (2024)	2	98.01
ALL	N/A	ResNet50-ViT	Tanwar et al. (2025)	15	99
Lymphoma (CLL, FL, MCL)	Andrewmvd, Kaggle	ML/DL Ensemble	Ozgur et al. (2024)	3	94, 92, 82
Malaria	Kassim et al. (2020)	LeNet5, DRNet	Harahap et al. (2021)	2	95.7, 95
RBC Morphology	N/A	CNN with ECOC	Alzubaidi et al. (2020)	3	92.06
RBC Abnormalities	Open-source	AlexNet, SVM	Aliyu et al. (2018)	5	33 (CNN), 100 (SVM)
RBC Abnormalities	Open-source	SVM with Canny Edge	Syahputra et al. (2017)	3	83.3
ALL	ISBI 2019 data set	ViT-CNN Ensemble	Jiang et al. (2021)	2	99.03

Table S1: Summary of related work on cancer classification using deep learning approaches.

2.1- Data Split Strategy

The three curated datasets of microscopic blood cell images, focusing on Leukemia, Lymphoma, and Red Blood Cell Morphology, were divided into training (80%), validation (10%), and test (10%) subsets to evaluate the individual and ensemble models (Table S2). The split was performed using scikit-learn's train_test_split function, while image groups, including original images and their augmentations, were shuffled using Python's random.shuffle to ensure random distribution across subsets. Images from the same patient, *including augmentations*, were assigned to a single subset using *patient IDs* from filenames, preventing overlap across subsets.

RBC	Class	Train	Test	Validation
	Acanthocyte	1004	296	132
	Cigar Cell	1332	360	200
	Hypochromia	820	304	160
	Normal	2303	635	354
	Schistocyte	329	80	44
	Spherocyte	1772	548	320
	Stomatocyte	1268	348	154
	Ovalocyte	2785	576	552
	Target Cell	2120	1240	775
	Teardrop	5547	1626	374
	Burr Cell	660	227	95
Lymphoma	CLL	354	45	44
	FL	420	54	52
	MCL	373	48	46
Leukemia	ALL (L1)	847	114	172
	ALL (L2)	3199	37	359
	AML (m1)	1360	170	170
	AML (m0)	797	101	99
	CLL	4517	566	564
	CML	2992	375	374

Table S2 : Overview of Class Distribution Across Training, Validation, and Test Sets

Following the dataset split into training, validation, and test subsets, a manual validation was conducted to ensure that there was no data leakage among the sets. This was achieved by ensuring the absence of images belonging to the same patient or source sample in more than one subset. Hence, the image distribution among classes is not optimally equal for subsets, but reflects deliberate adjustments performed to preserve the integrity of the evaluation.

2.2- Data Augmentation Details

Data augmentation was performed using a series of transformation techniques to increase the variability of the training set. Each original image was flipped horizontally using OpenCV's cv2.flip function with a flip code of 1, rotated 90 degrees clockwise using cv2.rotate with the ROTATE_90_CLOCKWISE flag, and rotated 180 degrees using the ROTATE_180 flag. Additionally, Gaussian blur of kernel size (5, 5) and sigma=0 was applied, allowing OpenCV to automatically determine the standard deviation. These processes were done to simulate real-world variation and make the model more robust to other orientations and small distortions.

3- Optimal Ensemble Weights

For the datasets used in this work, the optimal ensemble weights were selected based on the configuration that achieved the highest accuracy on each dataset. The following weight combinations between ViT and CNN were found to achieve the best results and were used in the final model evaluations:

- RBC dataset: Best ensemble accuracy of 96.96% achieved with weights (0.7 for ViT, 0.3 for CNN).
- Leukemia dataset: Best ensemble accuracy of 99.85% achieved with weights (0.3 for ViT, 0.7 for CNN).
- Lymphoma dataset: Best ensemble accuracy of 96.60% achieved with weights (0.6 for ViT, 0.4 for CNN).

These configurations were consistently observed to outperform other weight combinations in their respective datasets and were adopted in the final evaluation of the HematoFusion ensemble model.

4- Classification Report of the Weighted Ensemble for Each Class Across the three Datasets

Class	Precision	Recall	F1-Score	Support
Acanthocyte	1.00	0.97	0.98	296
Burr Cell	1.00	0.92	0.96	227
Ovalocyte	0.91	0.86	0.88	576
Normal	0.95	0.96	0.95	635
Hypochromia	0.86	0.91	0.88	304
Schistocyte	0.99	0.96	0.97	80
Spherocyte	0.94	0.97	0.95	548
Stomatocyte	0.97	0.98	0.97	348
Target Cell	1.00	0.97	0.98	1240
Cigar Cell	0.94	0.98	0.96	360
Teardrop	0.99	1.00	0.99	1626

Table S3. Classification Report – RBC Dataset using Weighted Ensemble Model

Class	Precision	Recall	F1-Score	Support
CLL	0.96	1.00	0.98	45
FL	1.00	0.93	0.96	54
MCL	0.94	0.98	0.96	48

Table S4. Classification Report – Lymphoma Dataset using Weighted Ensemble Model

Class	Precision	Recall	F1-Score	Support
CML	1.00	1.00	1.00	375
CLL	1.00	1.00	1.00	566
ALL (L1)	1.00	1.00	1.00	114
ALL (L2)	1.00	0.97	0.99	37
AML (m0)	0.99	1.00	1.00	101
AML (m1)	0.99	1.00	1.00	170

Table S5. Classification Report – Leukemia Dataset using Weighted Ensemble Model

5- AUC Scores and ROC Curves

Dataset	Class	AUC score	
		ResNet	ViT
RBC	Acanthocyte	1.0	1.0
	Burr Cell	1.0	1.0
	Ovalocyte	0.99	1.0
	Normal	1.0	1.0
	Hypochromia	0.99	0.99
	Schistocyte	0.99	1.0
	Spherocyte	1.0	1.0
	Stomatocyte	0.99	1.0
	Target Cell	1.0	1.0
	Cigar Cell	1.0	1.0
	Teardrop	1.0	1.0
Leukemia	CLL	1.0	1.0
	CML	1.0	1.0
	ALL (L1)	1.0	1.0
	ALL (L2)	1.0	1.0
	AML (m0)	1.0	1.0
	AML (m1)	1.0	1.0
Lymphoma	CLL	0.97	1.0
	FL	0.97	1.0
	MCL	0.96	1.0

Table S6: AUC scores for individual classes across the three datasets. High AUC values indicate strong discriminatory performance.

Table S6 illustrates the AUC scores for individual classes across the datasets, using the models ResNet50V2 and ViT that were selected for the ensemble model.

- **AUC = 1.0** means *perfect classification*: the model predicted all positive and negative examples correctly at all thresholds.
- **AUC = 0.99** suggests *near-perfect performance*, with very few misclassifications.

The extremely high AUC values (≥ 0.99) on all datasets show that the HematoFusion ensemble model yields *near-perfect* separability. Such performance, especially consistent across various datasets, suggests a strong and well-generalizing classifier.

For better visualization, Figure S1 presents the ROC curves for the ResNet50V2 and ViT models across each of the three datasets. The models show high separability with AUC values close to 1.0 across all classes.

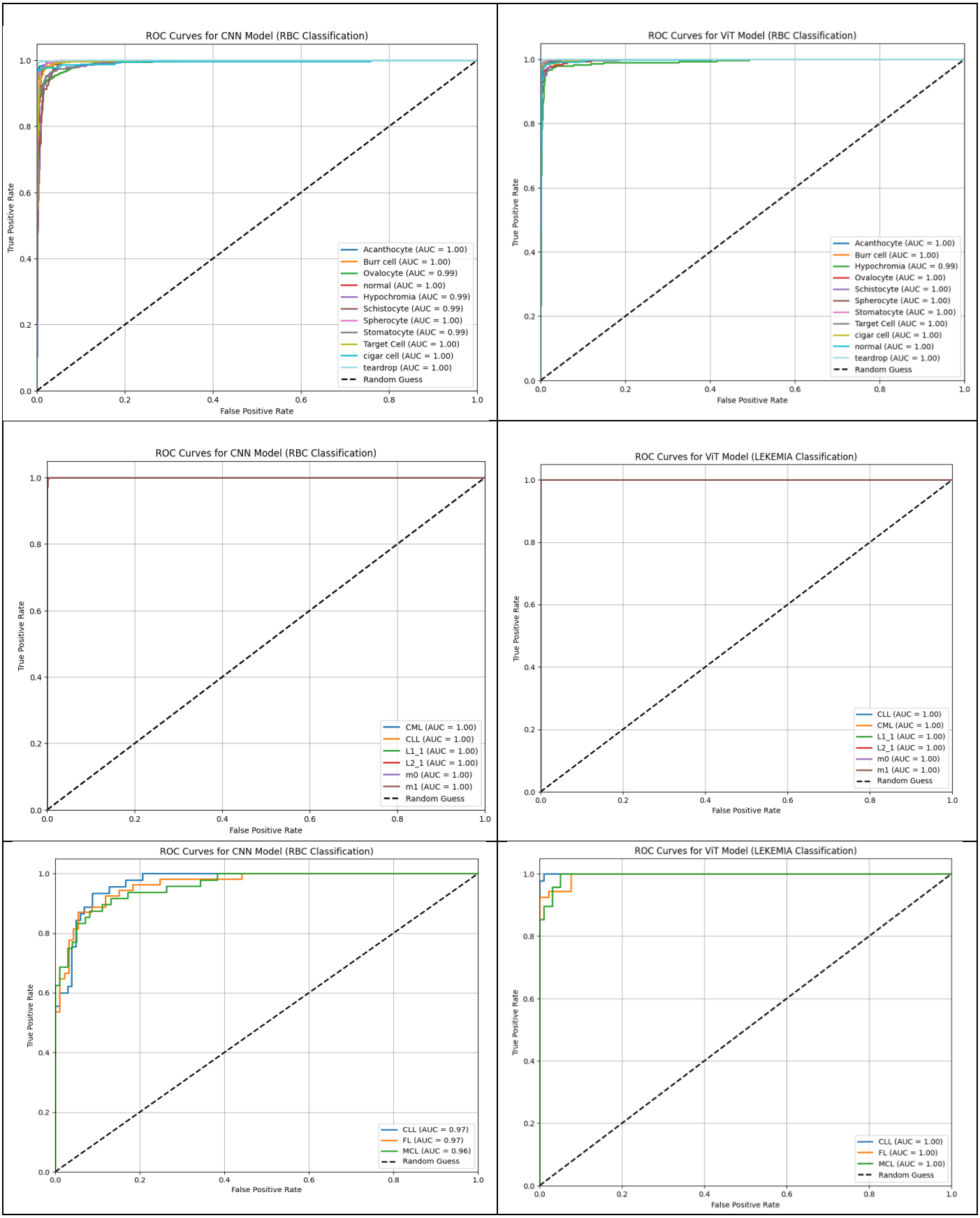


Figure S1: ROC curves for each class on the RBC, Leukemia, and Lymphoma datasets, respectively, (CNN on the left and ViT on the right).

6- Bootstrap Evaluation Results

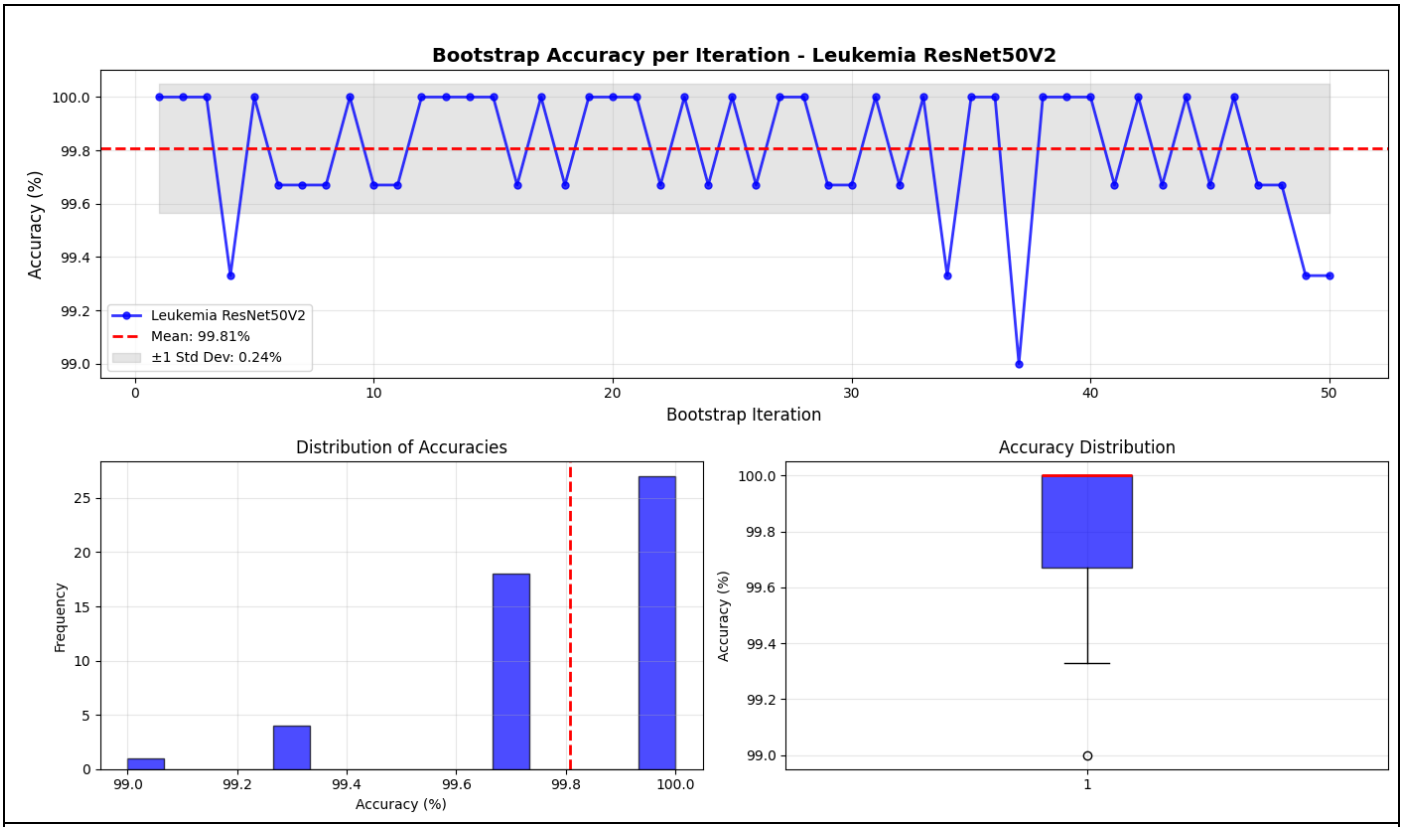
This section presents bootstrapping results for the performance evaluation of the two models (ResNet50V2 and ViT) on the three datasets: Dataset 1 (Leukemia), Dataset 2 (RBC), and Dataset 3 (Lymphoma). For each model-dataset combination, the visualization of accuracy distributions over multiple resampling iterations is provided to illustrate performance variability.

Table S7 reports the mean accuracy along with the 95% confidence intervals (CI) and the standard deviation of bootstrapped accuracy results for each model and dataset combination.

Figures S2-S4 illustrate the visualization of bootstrap accuracy across 50 iterations for each model-dataset combination.

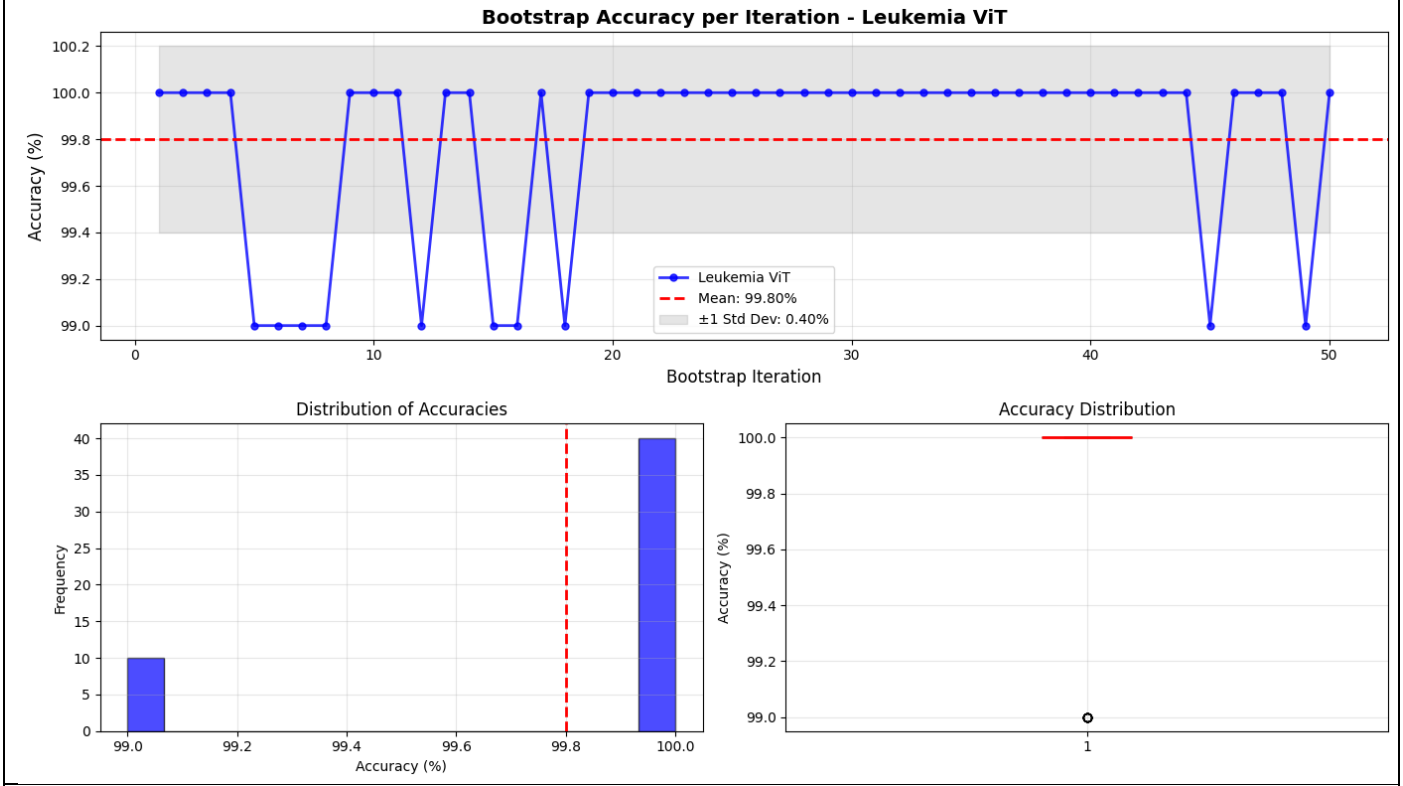
Dataset	Model	Mean Accuracy	95% CI Lower	95% CI Upper	Standard Deviation
Leukemia	ResNet50V2	99.81%	99.33%	100.00%	0.24%
	ViT	99.80%	99.00%	100.00%	0.40%
RBC	ResNet50V2	93.73%	93.26%	94.36%	0.29%
	ViT	96.90%	91.45%	99.78%	1.99%
Lymphoma	ResNet50V2	85.35%	79.74%	91.84%	3.31%
	ViT	95.40%	91.22%	98.78%	2.14%

Table S7: Summary of bootstrapped accuracy results for each model and dataset combination.



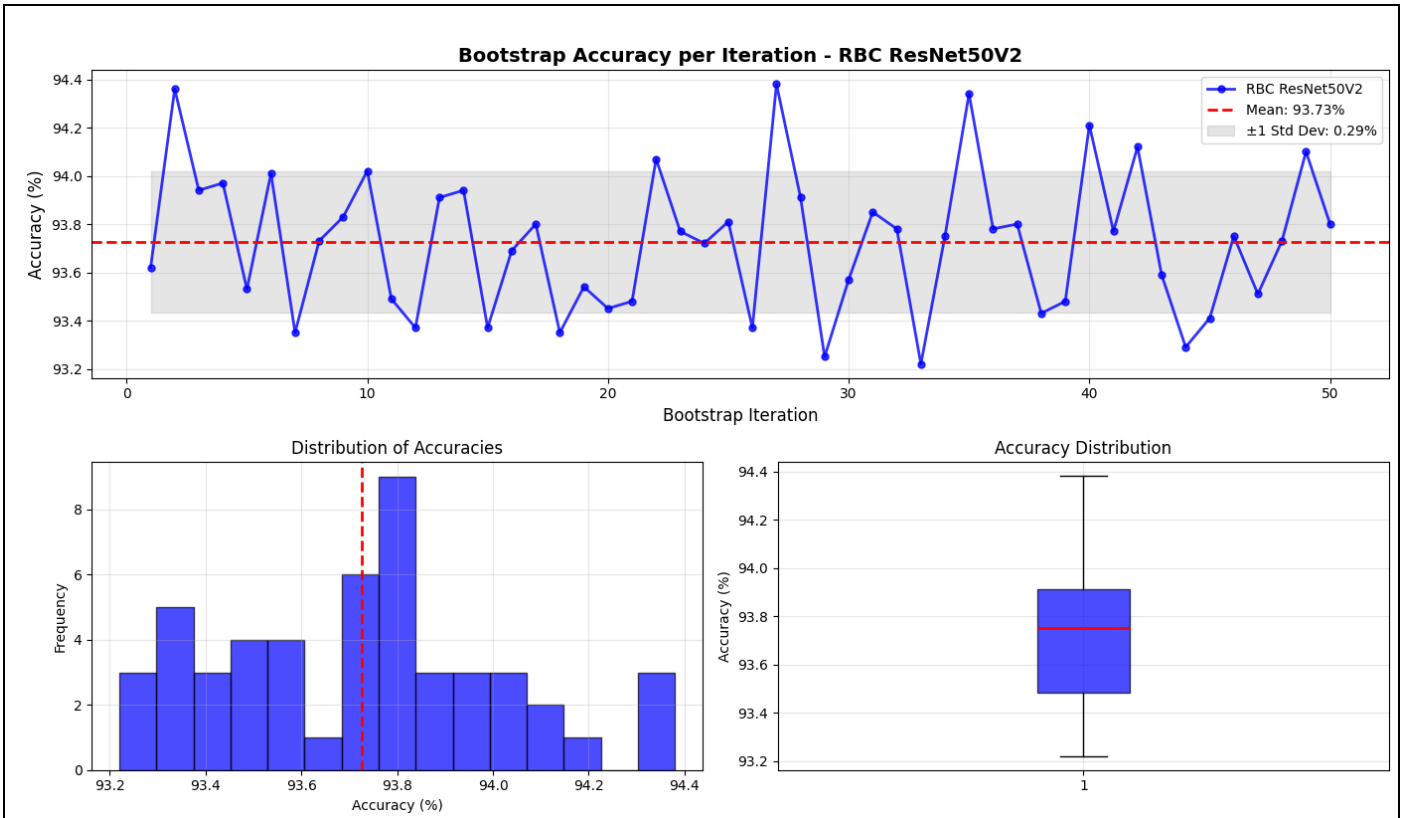
(a)

Figure S2: Bootstrapping distribution for the (a) ResNet50V2 and (b) ViT models on Leukemia dataset. The plot shows the distribution of accuracy over 50 iterations.



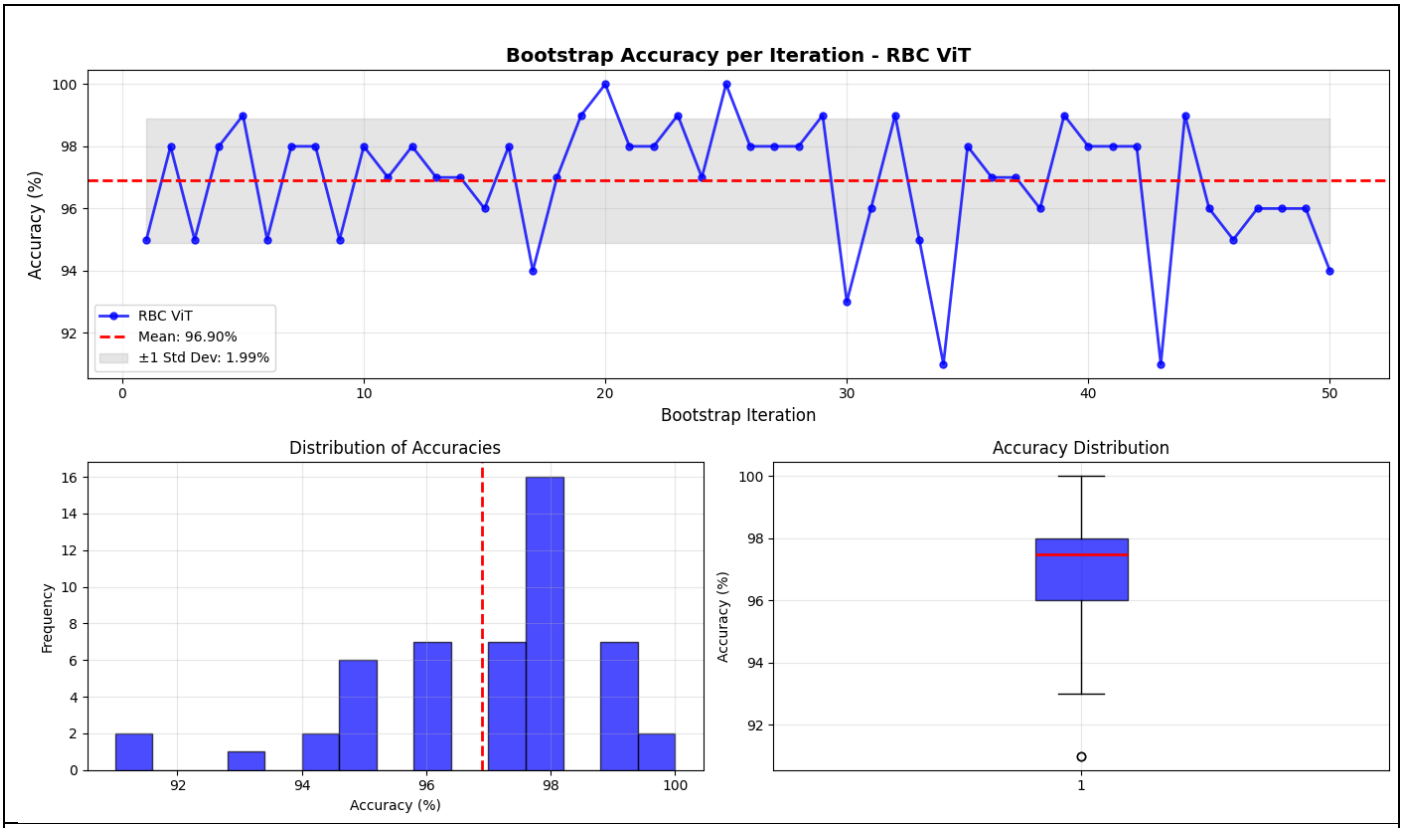
(b)

Figure S2 (Continued)



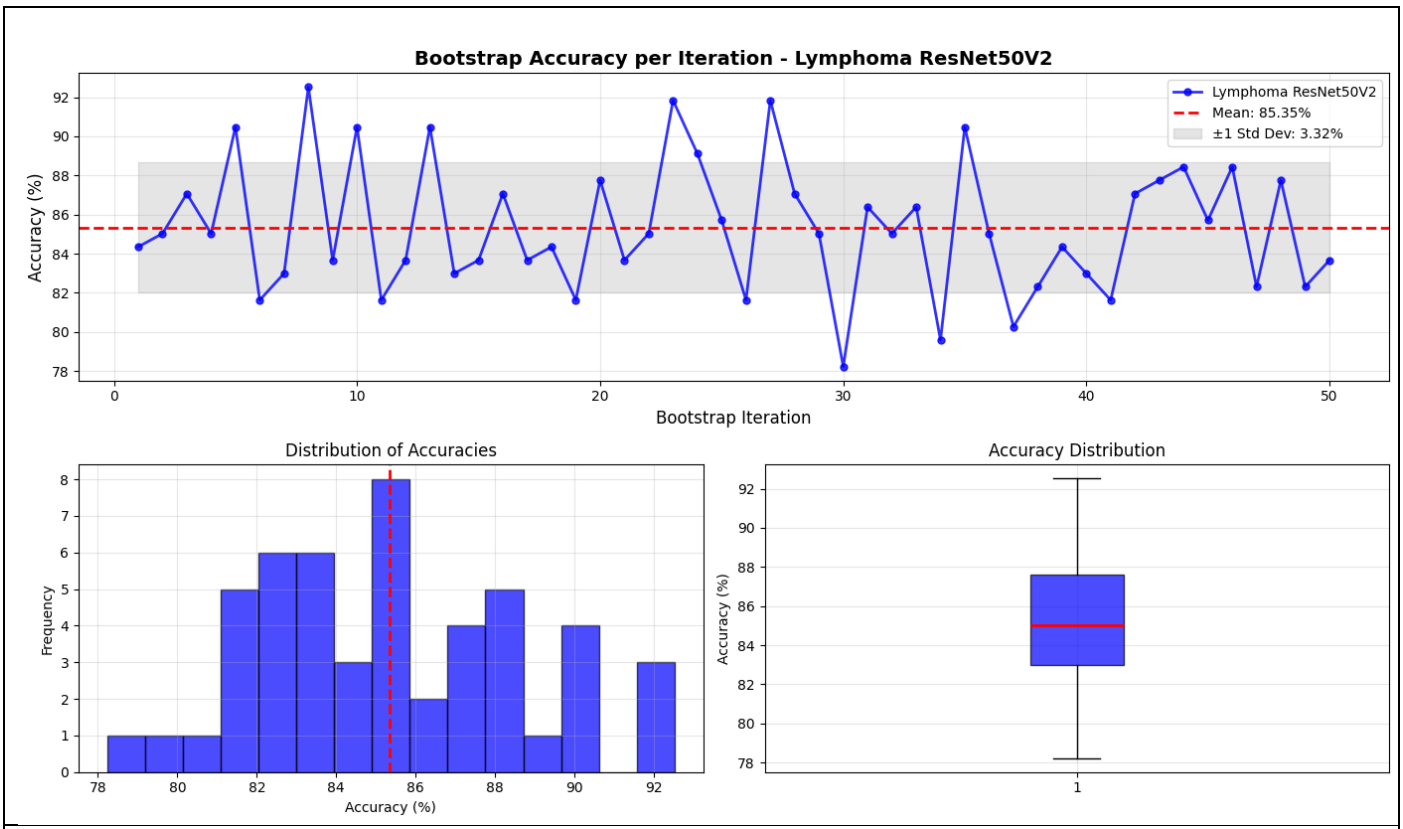
(a)

Figure S3: Bootstrapping distribution for the (a) ResNet50V2 and (b) ViT models on RBC dataset. The plot shows the distribution of accuracy over 50 iterations.



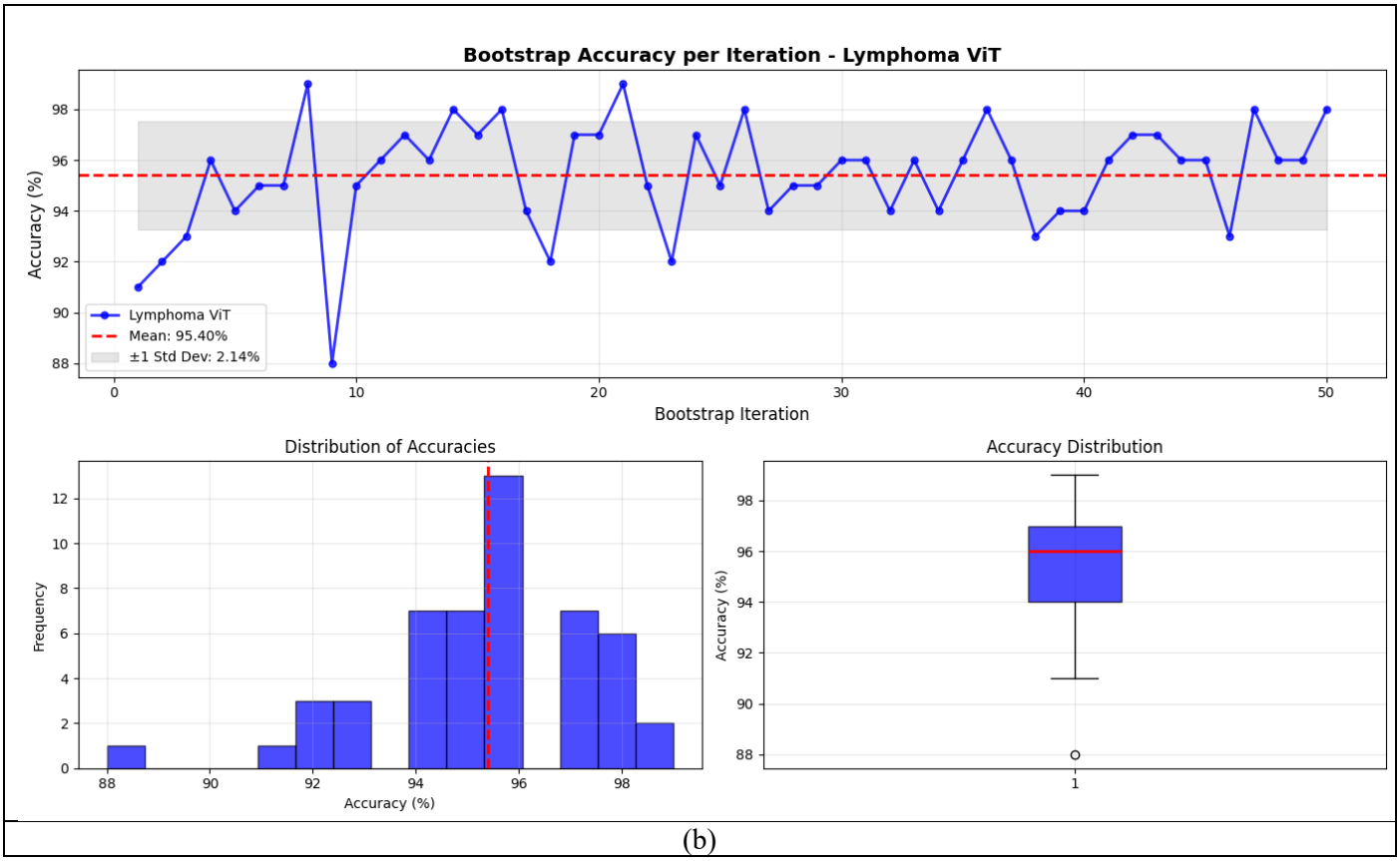
(b)

Figure S3 (Continued)



(a)

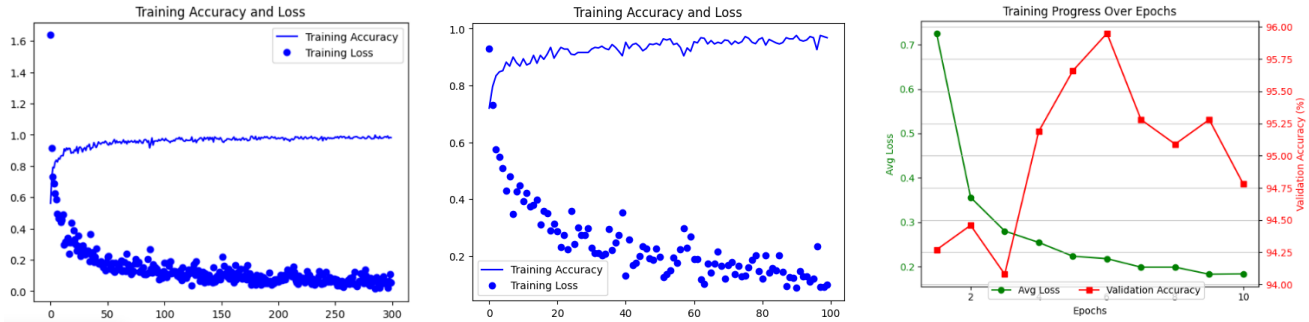
Figure S4: Bootstrapping distribution for the (a) ResNet50V2 and (b) ViT models on Lymphoma dataset. The plot shows the distribution of accuracy over 50 iterations.



(b)

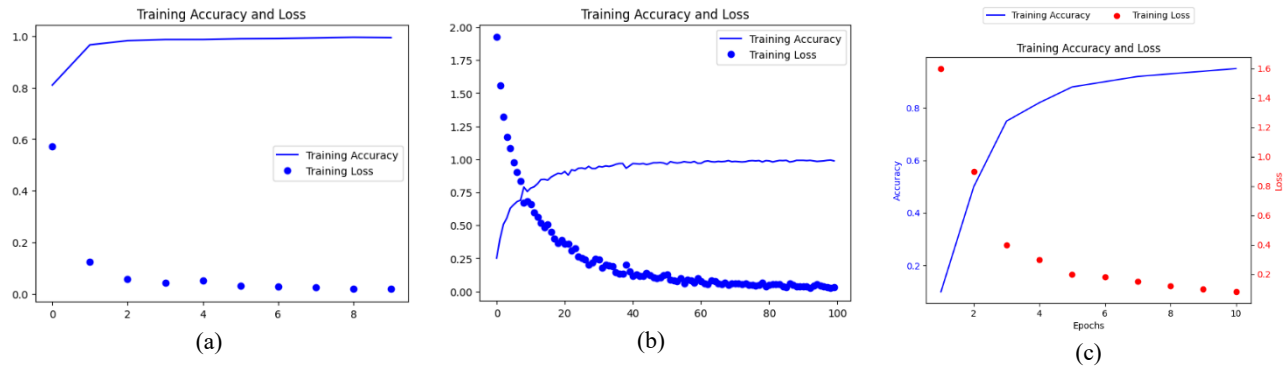
Figure S4 (Continued)

7- Training Curves



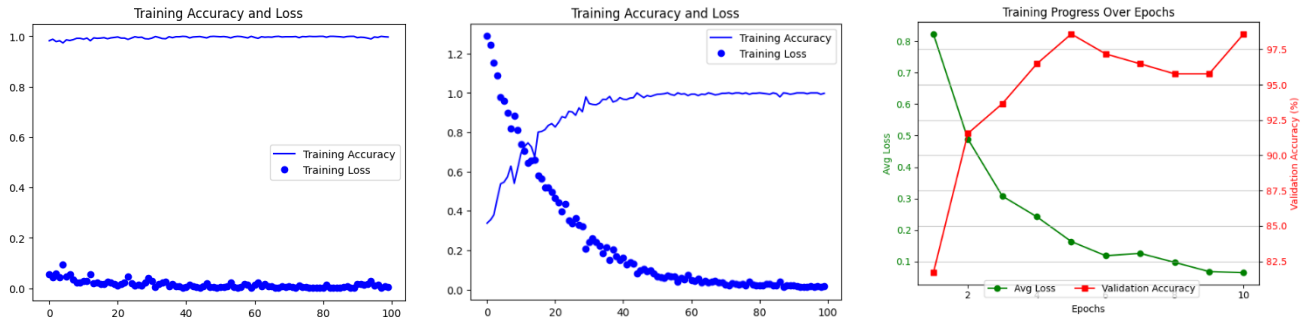
(a) (b) (c)

Figure S5: The training curves of the individual models on the RBC dataset: (a) The EfficientNetB3 model, (b) the ResNet50V2 model, and (c) the pre-trained ViT model



(a) (b) (c)

Figure S6: The training curves of the individual models on the Leukemia dataset: (a) The EfficientNetB3 model, (b) the ResNet50V2 model, and (c) the pre-trained ViT model



(a) (b) (c)

Figure S7: The training curves of the individual models on the Lymphoma dataset: (a) The EfficientNetB3 model, (b) the ResNet50V2 model, and (c) the pre-trained ViT model

8- Comparative with the Literature

Cancer Type	Dataset	Method	Authors	Classes	Accuracy (%)
RBC Abnormalities	Not shared	Neural Network	Kim et al. (2000)	15 RBC, 5 WBC	91
Leukemia	ALL-IDB, ASH ImageBank, JUST	CNN (AlexNet, DenseNet, ResNet, VGG16)	Areen et al. (2024)	N/A	94
ALL	ALL-IDB1, ALL-IDB2	DeepLeukNet (CNN)	Saeed et al. (2024)	N/A	99.61
Leukemia Subtypes	ALL-IDB, Munich AML Morphology	CNN, RandomForest, SVM, XGBoost	Kasim et al. (2025)	N/A	88
ALL	N/A	Vision Transformer	Swain et al. (2025)	N/A	99.67
ALL	ALL image dataset, Kaggle	Vision Transformer	Prasad et al. (2024)	2	98.01
ALL	N/A	ResNet50-ViT	Tanwar et al. (2025)	15	99
Lymphoma (CLL, FL, MCL)	Andrewmvd (Kaggle)	ML/DL Ensemble	Ozgur et al. (2024)	3	94, 92, 82
RBC Morphology	N/A	CNN with ECOC	Alzubaidi et al. (2020)	3	92.06
RBC Abnormalities	Open-source	AlexNet, SVM	Aliyu et al. (2018)	5	33 (CNN), 100 (SVM)
RBC Abnormalities	Open-source	SVM with Canny Edge	Syahputra et al. (2017)	3	83.3
Lymphoma (CLL, FL, MCL)	Multi Cancer Dataset (Kaggle)	Hybrid system between three DL models, XGBoost algorithms and Decision Tree algorithms	Hamdi et al. (2023)	3	96

Table S8 : HematoFusion Comparison with Literature

Cancer Type	Dataset	Method	Authors	Classes	Accuracy (%)
RBC Abnormalities	Not shared	Neural Network, SVM, KNN (Maximum voting ensemble)	Lotfi et al.	3	98
ALL	ISBI 2019 data set	ViT-CNN Ensemble	Jiang et al. (2021)	2	99.03
RBC	*RBC dataset*	*HematoFusion-RBC*	*Saadallah et al.*	*11*	*96*
Leukemia	*Raabin-Leukemia dataset*	*HematoFusion-Leukemia*	*Saadallah et al.*	*6*	*99*
Lymphoma	*Malignant Lymphoma Classification dataset*	*HematoFusion-Lymphoma*	*Saadallah et al.*	*3*	*96*

Table S8 (Continued)

To contextualize the performance of HematoFusion, an extended comparison with recent SOTA models across several hematologic malignancies is presented. As indicated in Supplementary Table S8 (Section 7), our ensemble model achieves competitive results for all three classification tasks. For RBC morphology, HematoFusion achieves 96% accuracy, which is better than traditional approaches such as ECOC-based CNNs (92.06%) and SVM-based approaches, and is also quite approaching the 98% reported by Lotfi et al. using an ensemble of traditional classifiers. In Leukemia classification, HematoFusion is 99% accurate, on par with or slightly underperforming top Vision Transformer-based architectures (e.g., Swain et al. at 99.67%, Saeed et al. at 99.61%), but with added generalizability to six classes. In Lymphoma, our model reaches an accuracy of 96%, outperforming existing works such as Ozgur et al. (with class-wise accuracies between 82–94%) and is comparable to Hamdi et al.'s hybrid model. Notably, HematoFusion performs outstandingly across different datasets in both diagnostic features and class distributions, which demonstrates its flexibility and stability in multi-class blood disease classification contexts.

9- Pseudo-Code Summary of The Pipeline

```
# Step 1: Dataset Preparation

# RBC Dataset Curation
RBC_mini_dataset = load("RBC_mini_dataset")
RBC_thalassemiaPBS_dataset =load("RBC_thalassemiaPBS_dataset")
Chula_RBC12_dataset=load("Chula_RBC12_dataset")
RBC_combined_dataset = combine(RBC_thalassemiaPBS_dataset, RBC_mini_dataset,
Chula_RBC12_dataset)

# Load other datasets
Leukemia_dataset = load("Raabin-Leukemia")
Lymphoma_dataset = load("Malignant_Lymphoma_Classification")

# Step 2: Preprocessing & Augmentation

for dataset in [RBC_combined_dataset, Leukemia_dataset, Lymphoma_dataset]:
    dataset = apply_preprocessing(dataset)
    dataset = apply_data_augmentation(dataset)
    dataset = stratified_split(dataset, train=0.8, val=0.1, test=0.1)
```

```
    dataset = rescale_images(dataset, target_size=(224, 224))

# Step 3: Model Training (for each dataset)

for dataset in [RBC_combined_dataset, Leukemia_dataset, Lymphoma_dataset]:
    # Train EfficientNetB3
    model_efficient = load_pretrained("EfficientNetB3", input_shape=(224,224,3))
    train(model_efficient, dataset, epochs=10, val_steps=316)

    # Train ResNet50V2
    model_resnet = load_pretrained("ResNet50V2", input_shape=(224,224,3))
    train(model_resnet, dataset, epochs=300, val_steps=10)

    # Train ViT
    model_vit = load_pretrained("ViT-Base-Patch16-224-in21k", input_shape=(224,224,3))
    train(model_vit, dataset, epochs=10)

# Step 4: Ensemble Inference (HematoFusion)

for dataset in [RBC_combined_dataset, Leukemia_dataset, Lymphoma_dataset]:
    y_pred_resnet = predict(model_resnet, dataset.test)
    y_pred_vit = predict(model_vit, dataset.test)

    # Weighted Average Ensemble
    weights = (0.7, 0.3)  # Example: 0.7 for ViT, 0.3 for ResNet
    ensemble_output = weighted_average([y_pred_vit, y_pred_resnet], weights)

    evaluate(ensemble_output, dataset.test.labels)
```