Decoding Cellular Complexity: A Deep Learning Expedition in Blood Cell Image Classification

Shivani Battu[1], Likhitha Marrapu[1], Chandini Karrothu[1], Anuranjani Thota[1], Hanan Muhajab[2], Areej Muhajab[3], Kambiz Ghazinour[4], Stacy Miner[5], Betis Baheri[6], and Safa Shubbar[6]

[1] Department of Data Science, Kent State University, Kent OH 44242 USA, sbattu1, lmarrapu, ckarroth, athota3@kent.edu

[2] Department of Computer Science, College of Engineering and Computer Science, Jazan University, Jazan, Saudi Arabia
hmuhajab@jazanu.edu.sa

[3] Computers and Information Technology College, Taif University, Taif, Saudi Arabia
areej.n@tu.edu.sa

[4] Department of Cybersecurity, State University of New York (SUNY), Canton, NY, USA
ghazinourk@canton.edu

[5] College of Nursing, Kent State University, Kent OH 44242 USA, sminer5@kent.edu

[6] Department of Computer Science, Kent State University, Kent OH 44242 USA, sshubbar, bbaheri@kent.edu

Abstract. Accurate classification of blood cell images is pivotal for advancing medical diagnostics. Our study introduces a robust framework that harnesses advanced deep learning techniques to enhance diagnostic accuracy and enable timely interventions in healthcare. We meticulously preprocessed a large dataset sourced from the Hospital Clinic of Barcelona, ensuring its suitability for model training. Employing an ensemble of cutting-edge deep learning architectures, including Convolutional Neural Networks (CNNs), U-Net, ResNet-50, VGG-16, DenseNet-201, Supervised Contrastive Learning, MobileNet-V2, InceptionV3, Xception, and a novel hybrid InceptionV3-Xception model. Comprehensive training and evaluation were conducted across these architectures. The results demonstrated significant improvements in blood cell image classification, with the hybrid InceptionV3-Xception model achieving a remarkable test accuracy of 98.51%. Furthermore, the study includes a detailed analysis of the effects of data shuffling on model performance, offering critical insights into the robustness and generalization capabilities of the proposed framework. This research contributes substantially to the field of medical image analysis, providing scalable solutions for accurate blood cell classification and advancing the prospects of healthcare diagnostics.

Keywords: Deep learning, Blood cell classification, Ensemble learning, Supervised contrastive learning, medical imaging, Hybrid models, Convolutional Neural Networks (CNNs)

1 Introduction

Access to affordable healthcare services is critical to achieve optimal population health outcomes. The heavy price of advanced healthcare equipment is burdensome for our healthcare institutions and impedes access to services, especially in low socio-economic and rural areas. With the help of advanced Machine Learning and Deep Learning techniques, technology has the potential to provide cost-effective solutions for evaluating blood smears. These methods evaluate blood images to classify cells into their respective types quickly and accurately determine blood counts. This information can be utilized to inform treatment decisions.

The 3 main categories of blood cells, Red Blood Cells(RBCs), White Blood Cells(WBCs), and Platelets, collectively constitute 40% of the blood circulating in the human body. WBCs can be further classified into five types - Lymphocytes, Monocytes, Eosinophils, Basophils and Neutrophils. The total number of cells, as well as the total of each category and WBC type, provide essential information regarding a patient's health and cardiovascular status. Diseases such as leukemia, AIDS, autoimmune disorders, immune deficiencies, and blood diseases can be diagnosed based on the number of WBCs (citation). Laboratory tests that use a cell counter or a flow cytometer can perform a complete blood count (CBC) rapidly. Nevertheless, a manual blood smear inspection is still required. These results are compiled for routine testing, for disease identification, and to monitor patients undergoing certain therapies. The blood cell evaluation assesses and describes the cells appearance as well as any abnormalities. This manual analysis is lengthy and repetitive, and the subjective result is prone to human error. In contrast, the proposed method, using machine image processing techniques, is entirely automated. The system can be trained to achieve high accuracy and easily identify the typical, and the unusual, differences in conditions of blood smear images obtained.

Deep learning methodology is superior to traditional image processing methods in literature. In addition, traditional methods require the appearance of the whole object to be able to recognize objects. Contrary to traditional methods, convolutional neural networks (CNN), a deep learning architecture, can extract features from a part of an object and perform object recognition. In this case, a CNN-based system shows a higher performance in recognizing partially visible cells for reasons such as overlap or only partial visibility of the image [1]. Therefore, it has been the motivation of this study to increase the performance of existing blood test devices with deep learning methods.

2 Literature Review

This research paper utilizes deep learning techniques, specifically convolutional neural networks (CNNs), for the automatic detection and classification of blood cells in peripheral blood smear images. At the initial phase, blood images are pre-processed using Open CV and resized for uniformity. The model built has

be trained on various classifiers to achieve notable accuracy in blood cell classification.

In the research [1], regional convolutional neural networks (R-CNN), Fast R-CNN, and Faster R-CNN algorithms were employed for object detection and classification. Transfer learning was applied to fine-tune pre-trained CNN architectures including AlexNet, VGG16, GoogLeNet, and ResNet50 [2]. The proposed method achieved remarkable accuracy rates in identifying different types of WBCs: Lymphocyte with 99.52% accuracy, Monocyte with 98.40% accuracy, Basophil with 98.48% accuracy, Eosinophil with 96.16% accuracy, and Neutrophil with 95.04% accuracy. The system demonstrated 100% success in determining the total number of all WBC types. These results showcase the effectiveness of deep learning techniques, particularly CNN-based architectures, in automated WBC detection and classification from blood smear images.

Convolutional Neural Network(CNN) techniques have been employed to classify white blood cells (WBCs) based on microscopic images, offering a significant advancement in disease diagnosis and treatment. By leveraging components such as convolution layers, stride, ReLU activation, pooling layers, and padding, CNNs excel in extracting intricate features from input images, thereby enhancing their capability to discern spatial hierarchies and patterns crucial for accurate classification. This innovative approach discussed in [3] empowers hematologists and health practitioners to efficiently categorize WBC subtypes, aiding in the identification of underlying health conditions with remarkable accuracy- exceeding 90 percent on the test set. Such breakthroughs not only streamline diagnostic processes but also pave the way for more reliable and efficient disease detection methodologies compared to traditional laboratory techniques.

Segmentation methods played a pivotal role in partitioning images into meaningful regions, crucial for accurate diagnosis and treatment planning. Various techniques, ranging from traditional clustering algorithms like K-means to advanced deep learning architectures such as convolutional encoder-decoder networks and U-Net, were employed to achieve precise segmentation results.

In the zest to utilize image processing techniques for create a system that can perform early identification of Leukaemia based on blood smear images, analysts in [4] developed a hybrid model of InceptionV3 and Xception using weighted averaging ensembling which achieved an accuracy of 91.71% in classifying the images. This proved to be highly efficient when compared to earlier tested models such as Transfer Learning.

The resurgence of contrastive learning techniques in self-supervised representation learning has garnered significant attention in recent years, culminating in remarkable performance gains for unsupervised training of deep image models. One such study [5] by Prannay Khosla, et al extends the application of self-supervised batch contrastive learning to the fully-supervised domain, enabling exploitation of label information. The approach entails pulling together clusters of points belonging to the same class in the embedding space while simultaneously pushing apart clusters of samples from different classes. The analysis delves into two versions of the supervised contrastive (SupCon) loss, discerning the most

effective formulation. Notably, on the ImageNet dataset, utilizing ResNet-200 architecture, the proposed method achieved a top-1 accuracy of 81.4%.

In the research [6] transfer learning with pre-trained CNN models, such as VGG16, VGG19, ResNet, InceptionV3, MobileNetV2, and DenseNet-20, have been explored for automatic blood cell classification. Achieving accuracies ranging from 91.375% to 94.72% on the PBC dataset, these models have inspired the development of novel CNN-based frameworks for enhanced accuracy. A proposed CNN model achieved remarkable accuracy of 99.91% on the PBC dataset, demonstrating competitive performance compared to prior literature. This underscores the potential of CNN-based frameworks in automating blood cell classification with high accuracy and efficiency.

In the study, "Stratification of White Blood Cells Using Optimized DenseNet201 Model" [7], Convolutional Neural Network (CNN) models based on Deep Learning were proposed for the classification of these cell types, leveraging pre-trained models. Performance evaluation was conducted using pre-trained DenseNet121, DenseNet201, EfficientNetB0, EfficientNetB7, ResNet50, and ResNet152V2 models. Comparative analysis revealed DenseNet201 as the most effective model demonstrated superior performance compared to other pre-trained models, achieving an accuracy of 86% with Adam optimizer, 30 epochs, and a batch size of 32, underscoring its efficacy in white blood cell classification.

Given the computational complexity associated with Deep Learning models when trained on large datasets, the adoption of lightweight models like MobileNetV2 is advocated to mitigate processing time. [2]. In a comparative analysis "A Deep Learning Model for Human Blood Cells Classification" by M. Pramodha, S. Ansith, et al on the same for classification of blood cells using various transfer learning approaches(such as VGG16, VGG19, ResNet50, and MobileNetV2), with a specific emphasis on the MobileNetV2 model tailored for accurate multi-classification of blood cells with performance evaluation metrics such as Accuracy, Precision, Recall, and F-Score; yielded MobileNetV2 as the top-performing model, achieving an accuracy of 97.89%, showcasing its potential for clinical applications.

U-net architecture was used for segmenting Plasmodium parasites within red blood cells [8] resulting in high accuracy levels of 99.40%, 99.36%, and 99.47% in RGB, HSV, and GGB color spaces respectively.

3 Methodology

3.1 Dataset

To conduct blood cell image classification, we leveraged advanced deep learning techniques as well as a meticulously curated dataset. A labelled dataset, comprising 17,092 high-resolution images of individual normal blood cells, was obtained from the Core Laboratory at the Hospital Clinic of Barcelona using the CellaVision DM96 analyzer. These images as shown in (Fig.1) encompass a variety of all types of blood cells, including neutrophils, eosinophils, basophils, lymphocytes, monocytes, immature granulocytes (promyelocytes, myelocytes, and

metamyelocytes), erythroblasts, and platelets or thrombocytes. Expert clinical pathologists annotated each image, ensuring high-quality labeling. Notably, the dataset featured images from individuals devoid of infection, hematologic or oncologic disease, and pharmacologic treatment at the time of blood collection, ensuring a representative sample of typically appearing (or healthy) peripheral blood cells.

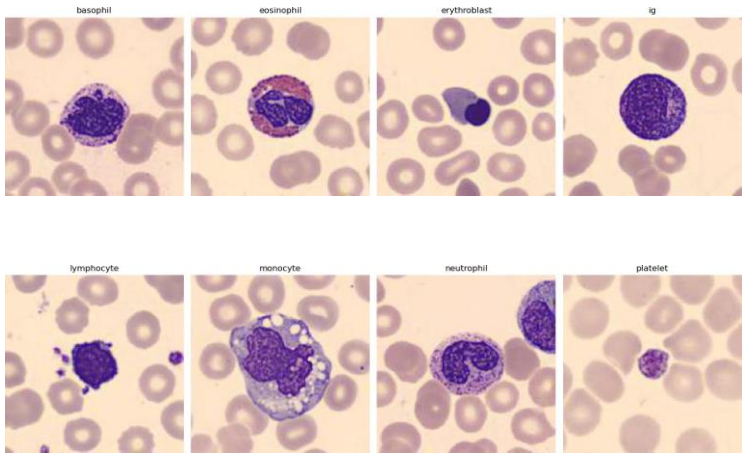


Fig. 1: Images in the Dataset

3.2 Data Pre-processing

As part of pre-processing, all the images were resized to a standardized format of 224 x 224 pixels using nearest neighbor interpolation to maintain image fidelity and converted them to NumPy arrays. This resizing ensured uniformity across the dataset while preserving essential morphological details. Secondly, the dataset was split into training and testing sets using a 70-30 split with stratification, ensuring that the class distribution in both sets remains similar to the original dataset. During training, 20% of the data was used for validation to monitor the model's performance. Below, (Fig.2) is the class distribution of training and testing sets for each of the 8 classes using Stratified sampling technique. This technique justifies the proportion of images chosen for the testing set to be uniform to the size of the class.

3.3 Process-Flow

A dataset is first imported into the classification process as conjured up in (Fig.3), after which all the dataset's annotations are retrieved and used to separate cells from images. Then, the pre-processing is carried out to convert images

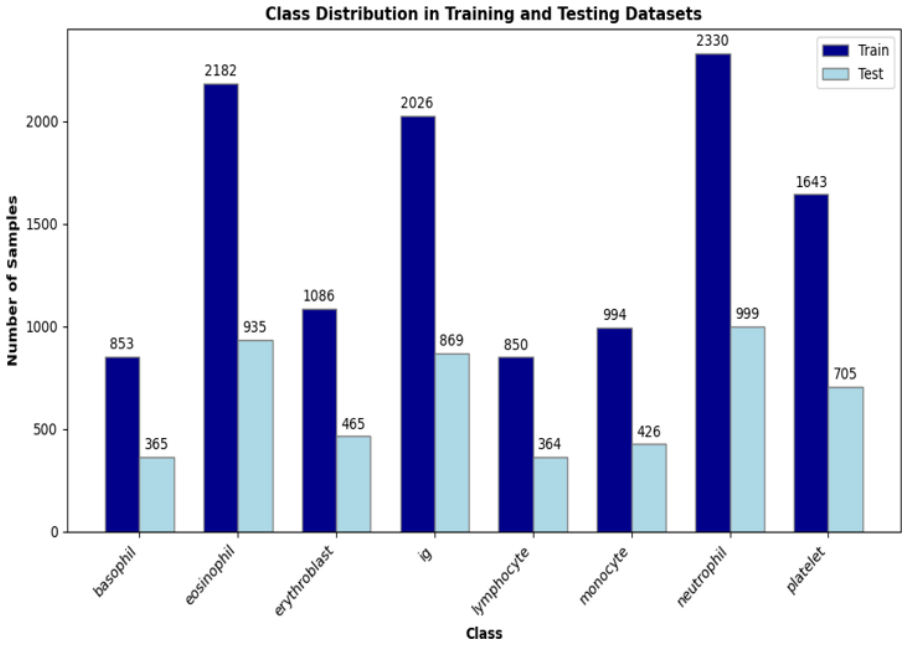


Fig. 2: Class Distirbution in the Training and Testing Data

into the necessary common format so that the model can be trained on them with ease. In addition to data pre-processing, we employed multiple state-of-the-art deep learning architectures for blood cell image classification and model comparison. Specifically, we utilized Convolutional Neural Networks (CNNs), U-Net, ResNet-50, DenseNet-201, VGG-16, Supervised Contrastive Learning, MobileNet-V2, Incpetion-V3, Xception, Hybrid Model(Inception-V3 and Xception).

These architectures offer distinct advantages for image analysis tasks, including robust feature extraction, precise localization, and effective gradient propagation. By leveraging these techniques and the meticulously curated dataset, we aimed to achieve accurate and reliable classification of blood cell images, thereby facilitating advancements in medical diagnostics and treatment monitoring.

The classifier models are employed on the dataset, which is split into 70% training data and 30% testing data for training and prediction of the results on testing. The predicted outcomes are analyzed in accordance with the classification reports from all the models.

3.4 Model Fitting

Convolutional Neural Network(CNN) Model We developed an image classification model that has been marked by iterative enhancements and thought-

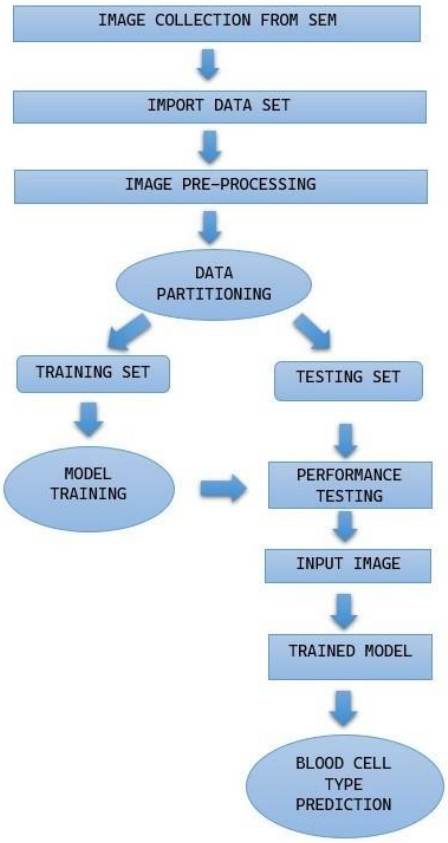


Fig. 3: Workflow illustrating deep learning techniques for blood cell image classification

ful adjustments. Initially, we crafted a Convolutional Neural Network (CNN) model [9] for image classification, employing a Sequential architecture with convolutional and dense layers. The model was trained using a batch size of 32 for 10 epochs, achieving a test accuracy of approximately 86.6%. Evaluation metrics including precision, recall, and F1-score were calculated, indicating overall satisfactory performance. The confusion matrix revealed the model's ability to distinguish between different classes, with notable accuracies across various categories. To enhance the robustness and reliability of our image classification model, we implemented a shuffling technique on our dataset before training. By randomizing the order of samples while preserving the distribution of classes, we aimed to reduce overfitting and improve the generalization ability of our model. Evaluation on the test set revealed promising performance metrics and a test accuracy of 91.08%, indicating the effectiveness of our approach in mitigating biases and enhancing model generalization. In addition to implementing

the shuffling technique, we employed 5-fold cross-validation to further validate the performance and robustness of our image classification model. This technique involved partitioning the dataset into five subsets, with each subset used once as a validation set while the model was trained on the remaining four subsets. During each fold of the cross-validation process, we trained a CNN model using the specified architecture and optimization techniques. By iteratively repeating this process and averaging the evaluation metrics across all folds, we obtained a comprehensive assessment of the model's overall performance. Finally, we evaluated the model's performance on the test set, calculating metrics such as accuracy, precision, recall, and F1-score. This approach allowed us to validate the effectiveness of our CNN model in handling image classification tasks. However, we encountered a notable challenge. Across different folds, our model struggled to accurately predict the platelet class due to class imbalances, as revealed by the confusion matrix. Implementing SMOTE addressed this, but memory constraints led us to reduce image size to 128x128, a compromise that did not compromise performance with a test accuracy of 94.10%. Table 1 summarizes the performance metrics obtained from our CNN model with 5-fold cross-validation and SMOTE augmentation. Despite this adjustment, our model maintained commendable accuracy, underscoring its scalability and reliability for real-world applications.

Table 1: CNN: 5-Fold Cross Validation (SMOTE) Results

Metric	**Performance (%)**
Training Accuracy	98.11
Test Accuracy	94.10
Precision	94.21
Recall	94.10
F1 Score	94.10

U-Net Model We developed a U-Net model (Fig.4) [10] which is a specialized image classification model tailored for blood cell analysis. U-Net architecture is a Convolutional Neural Network (CNN) renowned for its prowess in image segmentation tasks. Our objective was to optimize model performance through meticulous and iterative refinement. Initially, the model exhibited promising results, achieving a test accuracy of approximately 94.04%. However, a comprehensive evaluation revealed suboptimal precision, recall, and F1 score values, indicating potential shortcomings in real-world applicability. In response, we implemented a dataset shuffling strategy to mitigate overfitting and enhance generalization capabilities. This proven technique yielded substantial improvements, elevating the test accuracy to 94.66%. Despite this advancement, modest enhancements were observed in precision, recall, and F1 score metrics. This suggests that while the model demonstrates proficiency on training data, its translational efficacy,

particularly in the domain of blood cell analysis, may be limited. Furthermore, to strike a balance between computational efficiency and model performance, we adopted a pragmatic approach by scaling images to 128x128 pixels, alleviating memory constraints. Our findings underscore the scalability and reliability of the U-Net-based model for blood cell image classification tasks. Nonetheless, its adaptability to diverse and intricate datasets may encounter challenges, as evidenced by the noted limitations in accuracy, recall, and F1 score metrics. To surmount these deficiencies and bolster the model's utility in research and medical diagnostics, further exploration of alternative methodologies is warranted. This includes investigating novel techniques to enhance precision, recall, and F1 score values, thereby bolstering the model's efficacy in real-world applications.

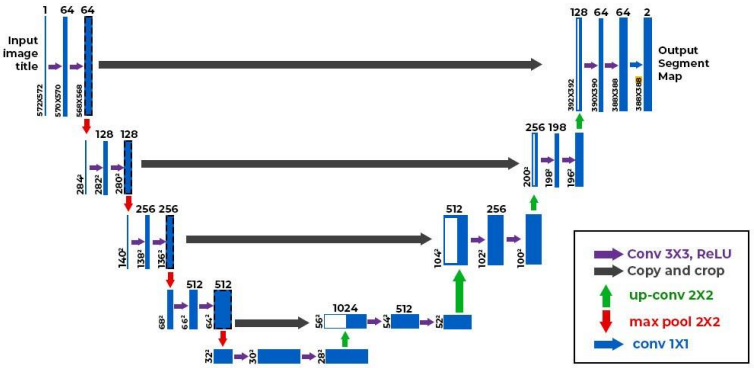


Fig. 4: UNet [10]

ResNet50 Choosing ResNet50 architecture as the base model for its deep layers, we transformed the labels into numeric format for model training after which we added custom layers on top of the base ResNet50 architecture, including global average pooling and dense layers, to extract features and perform classification. The model was then compiled using the Stochastic Gradient Descent (SGD) optimizer initially, and later using RMSprop optimizer for comparison. This involved utilization of Categorical cross-entropy loss function for optimization. Training the model on pre-processed data by employing 20 epochs and a batch size of 32 with a validation split of 20% to monitor performance and prevent overfitting, we followed a two-step approach of considering the original training data followed by a shuffled set for comparative analysis. Subsequently, we evaluated metrics including train accuracy, test accuracy, precision score, and built a confusion matrix, to assess the model's performance comprehensively. Based on our testing, RMSprop optimizer demonstrated superior test-accuracy compared to SGD optimizer, with RMSprop achieving 92.39% and SGD achiev-

ing 89.88%. The results indicate that the RMSprop optimizer outperformed the SGD optimizer, achieving higher accuracy in image classification tasks.

DenseNet-201 After pre-processing, we encoded class labels into numeric format using Label Encoder. For data with shuffling, we randomly shuffled to maintain consistency between images and labels. Then, we split the dataset into training and testing sets. Initializing the DenseNet201 model (Fig.5) [11] by removing its top classification layer and adding custom classification layers, we compiled the model using the Adam optimizer, sparse categorical crossentropy loss function, and accuracy metric. Training started on the training data with a batch size of 32 and 20 epochs, optionally utilizing a validation split for monitoring. Post-training, we evaluated the model on the test data, calculating test accuracy and additional metrics using sklearn.metrics functions. Finally, we assessed the model's performance, potentially visualizing it using a confusion matrix or other relevant metrics. Upon comparing the two approaches it is safe to conclude that the model was able to achieve 87.73% test-accuracy after shuffling in contrast to the baseline prediction of 90.17%.

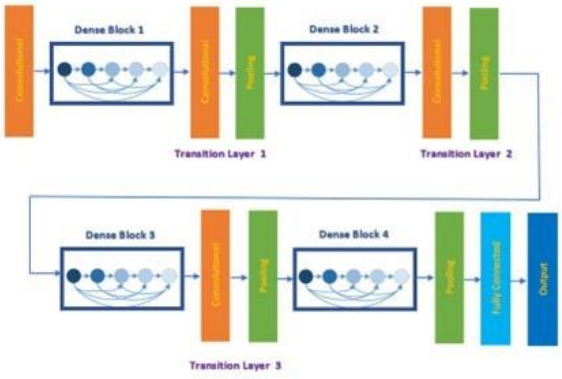


Fig. 5: DenseNet-201 [11]

Supervised Contrastive Learning Contrastive learning (Fig.6) [12] aims to learn representations by pulling similar instances together and pushing dissimilar instances apart in a latent space, without using explicit labels. In contrast, supervised contrastive learning incorporates class labels into the learning process. By leveraging labeled data, it ensures that representations from the same class are brought closer together while maximizing the separation between representations from different classes. This approach enhances the model's ability to learn discriminative features that are relevant for classification tasks, such as accurately classifying blood cell images. In our model development of Supervised

Contrastive Learning, we drew inspiration from Khosla et al. (2020) to devise a comprehensive approach for enhancing the discriminative power of our image classification model. Through careful implementation, including the use of data augmentation techniques and a pre-trained ResNet50V2 encoder, we aimed to extract high-level features from blood cell images efficiently. By incorporating a projection head into the encoder model, we projected high-dimensional feature representations into a lower-dimensional space, aiding in learning semantically meaningful representations optimized for the classification task. This projection head, achieved by adding a dense layer to reduce the dimensionality of the features, played a crucial role in our model's ability to discriminate between different classes of blood cell images. Furthermore, we utilized a supervised contrastive loss function during training, considering both similarities between samples of the same class and differences between samples from different classes. By leveraging labeled data in the contrastive learning framework, our model learned to produce representations that not only discriminate between classes but also capture their semantic relationships. This enabled the model to generalize well to unseen data, leading to improved classification performance. After training and evaluation without shuffling the dataset for 20 epochs, our model demonstrated an impressive accuracy of 97.33% on the test set. To further enhance the discriminative power of our image classification model, we employed supervised contrastive learning with shuffling for an additional 20 epochs. This approach allowed our model to learn to extract high-level features from blood cell images efficiently, achieving a remarkable accuracy of 96.35% on the test set. The evaluation metrics including precision, recall, and F1-score demonstrated the reliability and effectiveness of our model in accurately classifying blood cell images. These results underscore the potential of our approach in medical image classification tasks.

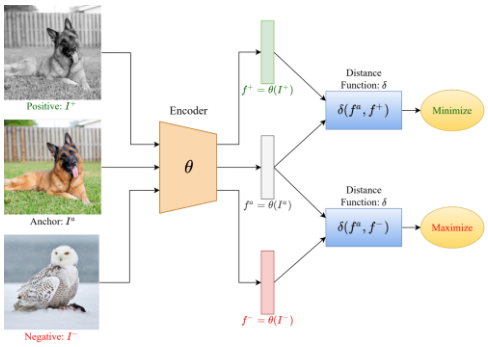


Fig. 6: Contrastive Learning [12]

In the absence of data shuffling, the supervised contrastive learning epochs exhibit notable trends in accuracy metrics. From the initial epochs to the later stages of training, both the training and validation accuracies demonstrate an

encouraging rise as in the (Fig.7), indicative of the model's progressive refinement in classifying blood cell images accurately. This steady improvement underscores the effectiveness of supervised contrastive learning in discerning intricate patterns within the data, leading to enhanced classification performance over successive training iterations. Moreover, the consistency between training and validation accuracies underscores the model's ability to generalize well to unseen data, bolstering its reliability in real-world applications of blood cell image analysis.

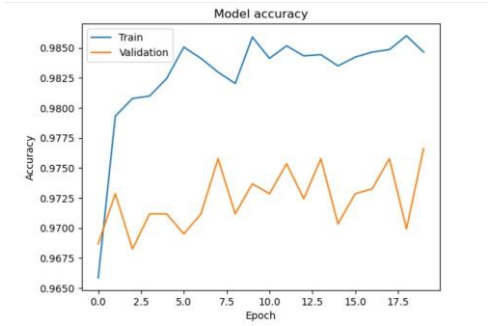


Fig. 7: Supervised Contrastive Learning Without Shuffling - Accuracy Plot

Across the 20 epochs of supervised contrastive learning with shuffling, the model's performance in terms of both training and validation accuracies demonstrated notable consistency and improvement as shown in (Fig.8). The sparse categorical accuracy, which measures the proportion of correctly classified samples, exhibited a steady increase throughout the training process, reflecting the model's ability to learn and generalize effectively. This upward trend in accuracies indicates that the model became progressively more adept at accurately classifying blood cell images over successive epochs. Despite minor fluctuations, particularly in the early stages of training, the overall trajectory suggests robust learning and convergence towards optimal performance. The validation accuracy, which assesses the model's generalization to unseen data, closely mirrored the training accuracy, further validating the model's efficacy in classifying blood cell images accurately and reliably.

Throughout the 20 epochs of supervised contrastive learning with shuffling, a consistent pattern emerged in the training and validation losses as depicted in (Fig.9a). Both the training and validation losses exhibited a gradual decrease over successive epochs, indicating that the model effectively learned to minimize discrepancies between predicted and actual labels. This downward trend in losses suggests that the model's performance improved steadily over time, capturing more nuanced features and refining its classification capabilities. Despite fluctuations in loss values, particularly in the early epochs, the overall trajectory

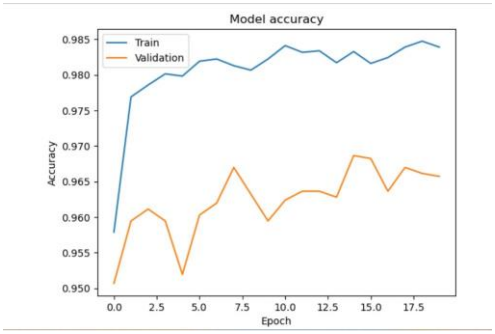
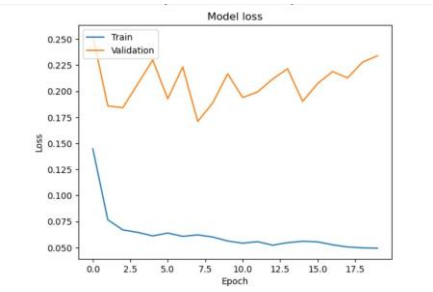


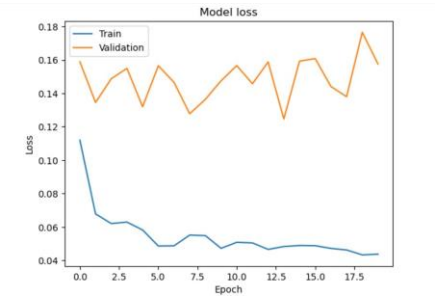
Fig. 8: Supervised Contrastive Learning With Shuffling - Accuracy Plot

demonstrates the effectiveness of the training process in optimizing the model for accurate blood cell image classification.

In the context of supervised contrastive learning without shuffling, the epochs reveal a consistent decrease in both training and validation losses as in the (Fig. 9b). This decline signifies the model's ability to effectively learn from the labeled training data and generalize well to unseen validation samples. The diminishing gap between the training and validation losses underscores the model's capacity to capture relevant features without overfitting to the training set. This convergence indicates the robustness of the supervised contrastive learning approach in extracting discriminative representations for accurate classification, even without shuffling the data.



(a) Supervised Contrastive Learning With Shuffling - Loss Plot



(b) Supervised Contrastive Learning Without Shuffling

Fig. 9: Comparison of Supervised Contrastive Learning Loss Plots

Mobilenet-V2 (Figure 10) [13] is a lightweight convolutional neural network (CNN) architecture, has garnered substantial attention in computer vision for

its efficiency and effectiveness in diverse tasks, notably image classification. Tailored to strike a balance between computational complexity and accuracy, it finds particular utility in resource-constrained devices like mobile phones and embedded systems. Leveraging transfer learning, our model capitalized on the features learned by MobileNetV2 from the expansive ImageNet dataset, initially achieving a commendable test accuracy of around 81.45%. This initial success underscores the model's adeptness in generalizing to unseen data, further substantiated by precision, recall, and F1-score metrics, alongside the confusion matrix, providing insights across various blood cell classes. In crafting a robust blood cell classification model, we embraced a sequential layer architecture, selectively freezing layers within the MobileNetV2 base, and fine-tuned hyperparameters, including the learning rate. This strategy streamlined model development, preserving crucial features from ImageNet while mitigating overfitting risks and enhancing generalization. To bolster performance and resilience, we implemented SMOTE for class imbalance and applied data augmentation during training. Following dataset shuffling and 30 epochs of training, our MobileNetV2 model exhibited an improved test accuracy of approximately 85.61%, surpassing the previous iteration. This enhancement underscores the efficacy of our approach, illustrating the model's adaptability to augmented and balanced data, ultimately enhancing blood cell image classification performance.

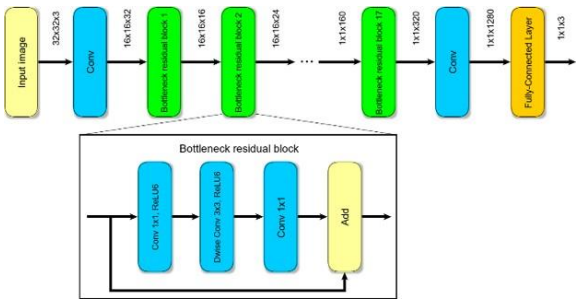


Fig. 10: MobileNet-V2 [13]

Inception-V3 Employing InceptionV3, (Fig.11) [14] a pre-trained convolutional neural network (CNN) model as the base model, we encoded the labels into a numeric format to prepare the data for training. Custom layers for classification were added on top of the InceptionV3 base to extract features and perform classification. The custom layers included a Global Average Pooling layer for feature extraction, a dense layer with 1024 units and ReLU activation for feature processing, and a dropout layer with a rate of 0.5 to prevent overfitting. A final dense layer with softmax activation was added for predicting the probability distribution of the classes. The model was compiled using the Adam optimizer with a learning rate of 0.001 and categorical cross-entropy loss function. The

model was trained on the pre-processed data for 10 epochs with a batch size of 32. The model achieved a test accuracy of 91.71%. These metrics, along with the confusion matrix, comprehensively evaluate the model's performance on the test set. Test accuracy measures the proportion of correctly classified samples, showcasing the effectiveness and reliability of the InceptionV3 model for blood cell image classification.

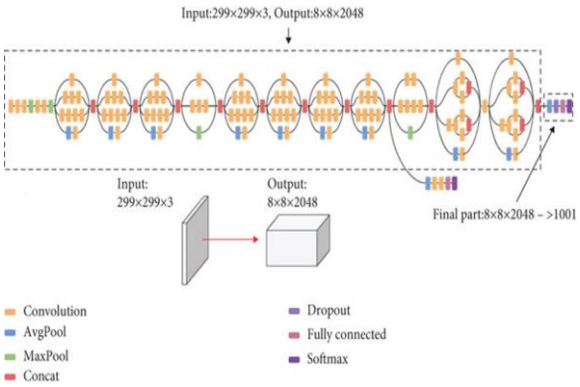


Fig. 11: Inception-V3 [14]

Xception: Employing Xception, (Fig.12) [15] , a pre-trained convolutional neural network (CNN) model as the base model for its effectiveness in capturing complex patterns, we encoded the labels into a numeric format for model training. Custom layers for classification were added on top of the Xception base, including a Global Average Pooling layer for feature extraction, a dense layer with 1024 units and ReLU activation for feature processing, and a dropout layer with a rate of 0.5 to prevent overfitting. A final dense layer with softmax activation was added for predicting the probability distribution of the classes. The Xception model was compiled using the Adam optimizer with a learning rate of 0.001 and categorical cross-entropy loss function. Training was conducted for 10 epochs with a batch size of 32. The model's performance was assessed using various metrics, including test accuracy, precision score, recall score, F1 score, and confusion matrix. Test accuracy measures the proportion of correctly classified samples. The Xception model exhibited robust performance, achieving a remarkable test accuracy of 97.89%. Additionally, the precision score stood at 97.92%, indicating the model's ability to correctly classify positive instances. The recall score of 97.89% reflects the model's capability to identify all relevant instances, minimizing false negatives. Furthermore, the F1 score, a harmonic mean of precision and recall, reached 97.89%, showcasing a balanced performance between precision and recall. These metrics, along with the confusion

matrix, provide comprehensive insights into the Xception model's effectiveness in classifying blood cell images.

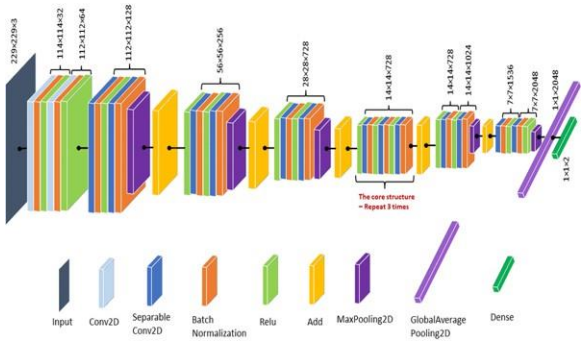


Fig. 12: Xception [15]

VGG16 We trained a robust image classification model tailored for blood cell analysis, leveraging the renowned Visual Geometry Group's VGG16 [16] architecture, a Convolutional Neural Network (CNN) acclaimed for its efficacy in image recognition tasks. Through iterative refinement and thoughtful adjustments, we aimed to optimize model performance. The initial model yielded promising results, achieving a test accuracy of approximately 80.36% without shuffling the data. To enhance the model's generalization capabilities and mitigate overfitting, we introduced a shuffling technique to the dataset, a method previously proven effective in bolstering robustness. The incorporation of shuffling yielded notable improvements, elevating the test accuracy to 79.37%. To alleviate memory constraints, we resized the images to 224x224 pixels, striking a balance between computational efficiency and model performance. Our study underscores the scalability and reliability of the VGG16-based model for real-world applications in blood cell image classification. It highlights the model's adaptability in handling challenges such as class imbalance while maintaining robust performance, furthering its potential for medical diagnostics and research.

Hybrid Model with Incpetion-V3 and Xception We developed a hybrid model (Fig. 13) by combining the InceptionV3 and Xception architectures. The outputs of the second last layers from both models were concatenated. Custom layers were added for classification, including a Dense layer with 512 units and ReLU activation for feature processing, followed by a Dropout layer with a rate of 0.5 to prevent overfitting. A final Dense layer with softmax activation was used to predict the probability distribution of the classes. The model was compiled using the Adam optimizer with a learning rate of 0.0001 and categorical

crossentropy loss. Initially, the hybrid model was trained for 10 epochs on the training data. The model's performance was assessed using various metrics, including test accuracy, precision score, recall score, F1 score, and a confusion matrix. Subsequently, the hybrid model structure with InceptionV3 and Xception was re-created, and the model was re-compiled with the same optimizer and loss function. The updated hybrid model was then retrained for 5 additional epochs on the training data. Upon evaluation, the hybrid model demonstrated strong performance, achieving an impressive test accuracy of 98.51%. The precision score of 98.53% indicated the model's accuracy in identifying positive instances, while the recall score of 98.51% showcased its ability to detect relevant instances while minimizing false negatives. The F1 score, reaching 98.52%, reflected a balanced performance between precision and recall. These metrics, combined with the confusion matrix, provided a comprehensive understanding of the hybrid model's efficacy in classifying blood cell images. Moreover, test images with their predicted and actual labels was presented for visual examination as shown in (Fig.14), offering insights into the model's predictions. By repeating the procedure after shuffling the dataset, we ensured a thorough assessment of the hybrid model's performance under varied data distributions. This process facilitated comparisons between the original and updated models, providing valuable insights into the effects of shuffling on model training and testing.

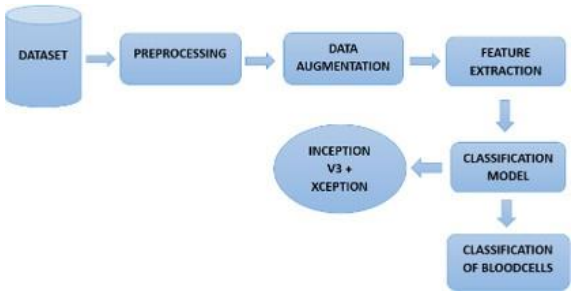


Fig. 13: Hybrid Model - Inception V3 and Xception

4 Results

In Table 2, the performance evaluation of various deep learning architectures for blood cell image classification reveals significant differences in accuracy and other metrics. This table summarizes the classification accuracies obtained without shuffling the dataset across different models. Each model exhibits distinct strengths and weaknesses, with the hybrid model (Inception V3 and Xception) demonstrating exceptional performance, achieving a test accuracy of 98.57%. Additionally, the Xception model attained a commendable test accuracy of 89.84%,

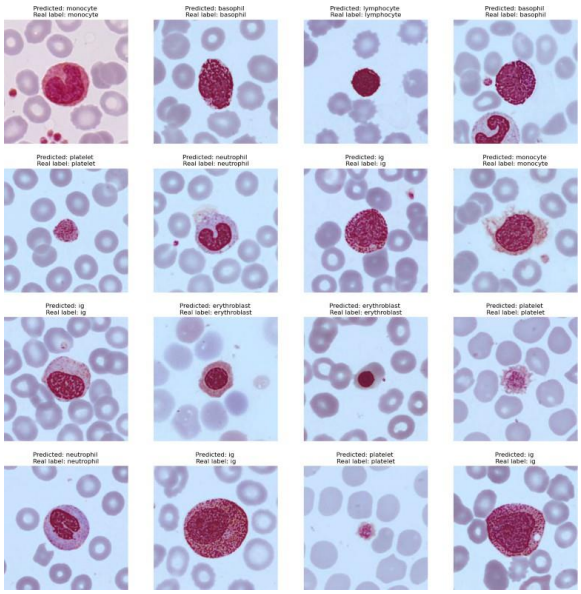


Fig. 14: Classification by the Hybrid Model

while Supervised Contrastive Learning also yielded promising results, with accuracies reaching up to 98.57%. These findings underscore the importance of selecting an appropriate architecture tailored to the specific requirements of blood cell analysis tasks.

Table 2: Classification Accuracies Without Shuffling

Model	Train Accuracy	Test Accuracy
CNN	97.15%	86.60%
U-Net	96.88%	94.04%
ResNet50-SGD	92.01%	90.76%
ResNet50-RMSProp	96.78%	89.22%
DenseNet-201	93.77%	90.17%
VGG-16	82.10%	80.36%
Supervised Contrastive Learning	98.46%	97.33%
MobileNet-V2	89.13%	85.39%
Inception-V3	97.93%	89.84%
Xception	99.05%	89.84%
Hybrid Model	99.89%	98.57%

In Table 3, substantial variations in accuracy and other evaluation metrics are observed when examining the effectiveness of different deep learning architectures on shuffled datasets. This table illustrates the classification accuracies

achieved by various models when applied to shuffled datasets. Notably, the hybrid model, a fusion of Inception V3 and Xception, achieves an outstanding test accuracy of 98.51%. Additionally, both the Xception model and the Supervised Contrastive Learning approach demonstrate commendable accuracy rates of 97.89% and 96.35%, respectively. These findings highlight the crucial roles played by data shuffling techniques and careful model selection in attaining optimal performance levels in blood cell image classification tasks.

Table 3: Classification Accuracies - With Shuffling

Model	Train Accuracy	Test Accuracy
CNN	95.76%	91.08%
U-Net	96.88%	94.66%
ResNet50-SGD	92.81%	89.88%
ResNet50-RMSProp	96.57%	92.40%
DenseNet-201	89.81%	87.09%
VGG-16	81.32%	79.36%
Supervised Contrastive Learning	98.39%	96.35%
MobileNet-V2	90.54%	85.61%
Inception-V3	98.17%	91.71%
Xception	99.34%	97.89%
Hybrid Model	99.75%	98.51%

5 Conclusion

In conclusion, our study marks a significant leap forward in medical diagnostics by introducing a robust framework for precise blood cell image classification. Through the integration of advanced deep learning techniques with a meticulously curated dataset, we have achieved remarkable strides in accuracy, with our hybrid InceptionV3-Xception model leading the pack with an exceptional test accuracy of 98.51%. Our comprehensive analyses, including the impact of preprocessing methods and data shuffling, underscore the reliability and scalability of our approach. Beyond academic discourse, our findings hold profound implications for clinical practice, promising to streamline decision-making processes and expedite interventions in healthcare settings. Looking ahead, continued research in medical image analysis stands to further refine our methodologies and drive transformative advancements in healthcare technology and patient care. This study represents a pivotal contribution to the quest for more accurate and efficient blood cell image classification, signaling a paradigm shift in the landscape of healthcare diagnostics.

References

1. Kutlu, H., Avci, E., and Özyurt, F., "White blood cells detection and classification based on regional convolutional neural networks," *Medical hypotheses*, vol. 135, p. 109472, 2020.
2. Pramodha, M., Ansith, S., Benifa, J. B., Al-Sarem, M., Hanumanthappa, J., Bini, A., Ndagijimana, E., Saeed, F., Heyat, M. B. B., Alqarafi, A. *et al.*, "A deep learning model for human blood cells classification," in *The International Conference of Advanced Computing and Informatics*. Springer, 2022, pp. 410–418.
3. Indrani, N. and Rao, C. S., "White blood cell image classification using deep learning," *vol*, vol. 7, pp. 7–12, 2021.
4. Rajeswari, S., Vasanth, C. S., Bhavana, C., and Chowdary, K. S. S., "Detection and classification of various types of leukemia using image processing, transfer learning and ensemble averaging techniques," in *2022 2nd Asian Conference on Innovation in Technology (ASIANCON)*. IEEE, 2022, pp. 1–6.
5. Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., and Krishnan, D., "Supervised contrastive learning," *Advances in neural information processing systems*, vol. 33, pp. 18 661–18 673, 2020.
6. Asghar, R., Kumar, S., and Mahfooz, A., "Classification of blood cells using deep learning models," *arXiv preprint arXiv:2308.06300*, 2023.
7. Gill, K. S., Anand, V., Gupta, S., and Figat, P., "Stratification of white blood cells using optimized densenet201 model," in *International Conference on Advanced Communication and Intelligent Systems*. Springer, 2022, pp. 31–47.
8. Nautre, A., Nugroho, H. A., Frannita, E. L., and Nurfauzi, R., "Detection of malaria parasites in thin red blood smear using a segmentation approach with u-net," in *2020 3rd international conference on biomedical engineering (IBIOMED)*. IEEE, 2020, pp. 55–59.
9. Jung, C., Abuhamad, M., Mohaisen, D., Han, K., and Nyang, D., "White blood cells (wbc) images classification using cnn-based networks," 2021.
10. .S, A., Priyanka, A., Kavitha, G., and K, B., "Automatic detection of acute lymphoblastic leukemia using unet based segmentation and statistical analysis of fused deep features," *Applied Artificial Intelligence*, vol. 35, pp. 1–18, 10 2021.
11. Attallah, O., "Mb-ai-his: Histopathological diagnosis of pediatric medulloblastoma and its subtypes via ai," *Diagnostics*, vol. 11, p. 359, 02 2021.
12. Kundu, R., "The beginner's guide to contrastive learning," https://www.v7labs.com/blog/contrastive-learning-guide, 2022, accessed: [insert date accessed].
13. Seidaliyeva, U., Akhmetov, D., Ilipbayeva, L., and Matson, E., "Real-time and accurate drone detection in a video with a static background," *Sensors*, vol. 20, p. 3856, 07 2020.
14. Ramaneswaran, S., Srinivasan, K., Vincent P M, D., and Chang, C.-Y., "Hybrid inception v3 xgboost model for acute lymphoblastic leukemia classification," *Computational and Mathematical Methods in Medicine*, vol. 2021, pp. 1–10, 07 2021.
15. Mehmood, A., "Efficient anomaly detection in crowd videos using pre-trained 2d convolutional neural networks," *IEEE Access*, vol. PP, pp. 1–1, 10 2021.
16. Deif, M. A., Attar, H., Amer, A., Elhaty, I. A., Khosravi, M. R., Solyman, A. A. *et al.*, "Diagnosis of oral squamous cell carcinoma using deep neural networks and binary particle swarm optimization on histopathological images: an aiomt approach," *Computational Intelligence and Neuroscience*, vol. 2022, 2022.