Journal of Prevention, Diagnosis and Management of Human Diseases
Vol: 06, No. 02, July-Dec 2026
ISSN: 2799-1202, DOI: 10.55529/jpdmhd.62.43.53

Research Paper

Automated white blood cell differential from blood-smear images: Segmentation accuracy and cross-laboratory generalisation on public datasets

Dr. Anuj Gaba[1*], Dr. Dhara Andipara[2], Atabekov Tenirberdi[3]
[1*]Department of Forensic Medicine and Toxicology, AIIMS Kalyani, West Bengal, India.
[2]Department of Dentistry, Geetanjali Dental and Research Institute, Udaipur, Rajasthan, India.
[3]Department of Public Health and Family Medicine, Jalal-Abad International University, 720906, 24 Kurmanbek Str., Manas, Kyrgyzstan.

Article Info

Article History:
Received: 28 March 2026
Revised: 15 June 2026
Accepted: 22 June 2026
Published: 08 August 2026

Keywords:
White Blood Cell Differential
Leucocyte Classification
Domain Shift
Stain Normalisation
Haematology
Diagnostic Imaging

ABSTRACT

Background: The white blood cell (WBC) differential is central to the diagnosis of infection, allergy and haematological disease, and automated image-based systems promise to make it available where trained morphologists are scarce. Their reliability across laboratories is rarely tested.
Objective: To evaluate the within-dataset and cross-dataset performance of automated nucleus segmentation and WBC classification, and to test whether common remedies restore transferability.
Methods: Two public datasets of real stained blood-smear images with expert nucleus and cytoplasm annotations, acquired with different staining and imaging systems, were used (299 and 97 cells of four types: neutrophils, lymphocytes, monocytes and eosinophils). Five unsupervised nucleus segmentation methods were compared with expert masks. An interpretable random forest on 14 morphological and colour features and a compact convolutional neural network (CNN) were evaluated within each dataset by cross-validation and across datasets by training on one and testing on the other. Colour augmentation, scale-harmonised crops and Reinhard stain normalisation were tested as remedies. Results: K-means clustering in CIELAB space segmented nuclei with a mean Dice coefficient of 0.947 and 0.960 on the two datasets, outperforming all thresholding methods ($p < 10^{-6}$ in most comparisons). Within a dataset, the CNN reached a balanced accuracy of up to 0.95 and the random forest 0.86; both reproduced the differential within 1–2 percentage points. Across datasets, CNN balanced accuracy fell to 0.30–0.46 (chance 0.25), with systematic errors such as reporting 1% neutrophils in a population with 59%; none of the remedies restored performance. The morphology-based random forest transferred better (balanced accuracy 0.49 and 0.71) but still mis-estimated the differential by 13–14 percentage points on average.

Journal of Prevention, Diagnosis and Management of Human Diseases (JPDMHD)

Conclusions: High within-dataset accuracy does not predict performance in a different laboratory setting. Automated WBC differentials should be validated externally and locally before diagnostic use. Interpretable morphological features provide a transparent and more transferable baseline.

Corresponding Author:
Dr. Anuj Gaba
Department of Forensic Medicine and Toxicology, AIIMS Kalyani, West Bengal, India
Email: anuj.gaba989@gmail.com

Copyright © 2026 The Author(s). This is an open access article distributed under the Creative Commons Attribution License, (http://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

The white blood cell (WBC) differential count, the proportion of neutrophils, lymphocytes, monocytes, eosinophils and basophils in peripheral blood, is one of the most frequently requested haematological investigations. Abnormal proportions point towards bacterial or viral infection, allergic and parasitic disease, haematological malignancy and drug toxicity, and a blood smear review is the standard way to confirm flagged results from automated analysers [1], [2]. Manual review is slow and requires trained staff, which limits its availability in primary care and in low-resource settings. Digital morphology analysers that locate and classify leucocytes automatically are now used in many laboratories [3], and research systems based on classical image analysis [4], [5] and deep learning [6], [7] have reported high accuracy.

Most of these results, however, are obtained by training and testing on images from a single laboratory, with the same staining protocol, microscope and camera. Blood smears from different sites differ in stain colour, illumination, magnification and image quality, and models trained on data from one site can fail at another [8], [9]. For a diagnostic aid intended for prevention and early diagnosis in diverse settings, this cross-site behaviour matters more than within-site accuracy [10]. Two common remedies are colour (stain) normalisation [11], [12] and colour augmentation during training [8], but their benefit for leucocyte classification on real multi-source images has rarely been quantified.

This study uses two public datasets of real stained blood-smear images with expert-drawn nucleus and cytoplasm masks and cell-type labels [13], acquired with different staining and imaging systems. It addresses three questions: (i) how accurately simple, unsupervised colour-based methods segment the leucocyte nucleus compared with expert masks; (ii) how well an interpretable morphology-based classifier and a small convolutional neural network (CNN) identify the four main leucocyte types within each dataset; and (iii) how much accuracy is lost when a model trained on one dataset is applied to the other, and whether stain normalisation or colour augmentation recovers it.

2. METHODODOLOGY

2.1 Image Data

Both datasets were released by Zheng et al. [13] and are freely available [14]. Dataset 1 (D1) contains 300 images of 120 × 120 pixels, each centred on one leucocyte, acquired with a Motic Moticam Pro 252A camera on an N800-D motorised microscope after a rapid staining protocol; the background is typically yellow. Dataset 2 (D2) contains 100 images of 300 × 300 pixels from the CellaVision blog, stained in purple tones and showing more surrounding red cells. For every image, domain experts marked the

nucleus, cytoplasm and background, and each cell is labelled as neutrophil, lymphocyte, monocyte, eosinophil or basophil. Because only four basophils were available (one in D1, three in D2), basophils were excluded from classification, leaving 299 images in D1 and 97 in D2 (Table 1). Figure 1 shows real examples of each class from both datasets with the expert annotations. No patient information is associated with the images.

Table 1. Composition of the Two Public Leucocyte Image Datasets Used in the Study

Dataset	Source and Staining	Image Size (Pixels)	Neutrophil	Lymphocyte	Monocyte	Eosinophil	Basophil*	Total Used
D1	Jiangxi Tecom; rapid stain, yellow background	120 × 120	176	53	48	22	1	299
D2	CellaVision blog; purple stain, red cells visible	300 × 300	33	37	15	12	3	97

*Excluded from classification because of insufficient numbers.

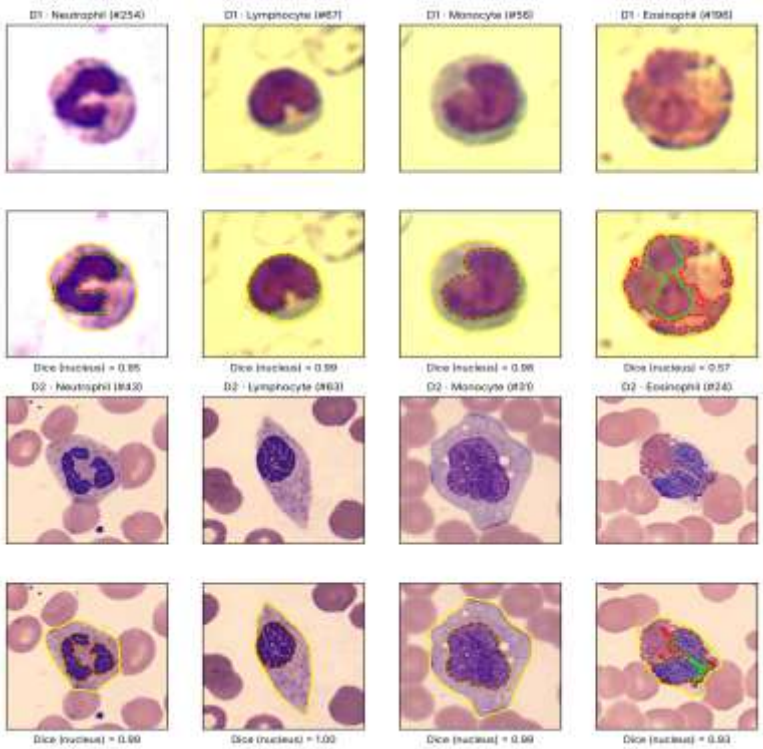


Figure 1. Real Blood-Smear Images from Dataset 1 (Top Two Rows) and Dataset 2 (Bottom Two Rows), One Randomly Selected Cell per Class. In the Second Row of Each Pair, the Expert Nucleus (Green) and Cell (Yellow) Boundaries Are Shown Together with the Automatic K-Means Nucleus Segmentation (Red, Dashed) and its Dice Agreement with the Expert Nucleus

2.2 Unsupervised Nucleus Segmentation

Leucocyte nuclei stain more intensely than the cytoplasm, red cells and background, which allows unsupervised segmentation. Five colour-based methods that require no training were compared: Otsu thresholding [15] of (i) the grey-level image, (ii) the HSV saturation channel, (iii) the magenta channel of the CMYK representation and (iv) the haematoxylin channel obtained by colour deconvolution [16]; and (v) K-means clustering [17] of the CIELAB colour values into three clusters, taking the cluster with the lowest lightness as nucleus. All masks were post-processed identically: morphological opening with a disc of radius 1, removal of objects smaller than 0.3% of the image, retention of the largest object and of any other object within 2.5 equivalent radii of it (to keep separate nuclear lobes) and filling of small holes. Agreement with the expert nucleus mask was measured by the Dice coefficient, $2|A \cap B| / (|A| + |B|)$.

2.3 Interpretable Morphology-Based Classifier

For each cell, 14 features were computed from the automatically segmented nucleus (K-means) and from a perinuclear ring obtained by dilating the nucleus: the number of nuclear lobes, solidity, eccentricity, circularity ($4\pi A/P2$), extent, area relative to the convex hull, relative nuclear size ($\sqrt{A}$ divided by the image side), the mean CIELAB colour of the nucleus and of the perinuclear ring, and the standard deviation of nuclear lightness as a texture measure. These features correspond to the criteria used by morphologists: lobulation for neutrophils, a round dense nucleus for lymphocytes, a large indented nucleus for monocytes and orange-red granules for eosinophils [1]. A random forest [18] with 500 trees, balanced class weights and a minimum leaf size of 2 was used. For comparison, the same features were also computed from the expert nucleus mask, which gives an upper bound for the effect of segmentation errors.

2.4 Convolutional Neural Network

A compact CNN was trained from scratch, since no pretrained weights were used: four convolutional blocks (16, 32, 64 and 96 channels, 3 × 3 kernels, batch normalisation, ReLU and 2 × 2 max pooling), global average pooling, dropout (0.3) and a linear output layer for the four classes. Images were resized to 96 × 96 pixels. The network was trained with Adam [19] (learning rate 2 × 10−3, weight decay 10−4, cosine schedule) for 60 epochs with batch size 32 and a class-weighted cross-entropy loss, using random flips and 90° rotations. Three input variants were compared: (a) the whole image with geometric augmentation only; (b) the whole image with additional colour augmentation (random per-channel gain of ±25%, offset of ±0.1 and conversion to greyscale with probability 0.2) [8]; and (c) a scale-harmonised input in which a square crop centred on the automatically segmented nucleus, with a side of 3.2 nuclear equivalent radii, was resized to 96 × 96 pixels, used in colour and in greyscale. Variant (c) was designed to remove the difference in magnification between the datasets. Models were implemented in PyTorch [20]; segmentation and features used scikit-image [21] and scikit-learn [22].

2.5 Evaluation Design

Within-dataset performance was estimated by stratified five-fold cross-validation (repeated three times for the random forest). Cross-dataset performance was estimated by training on all images of one dataset and testing on all images of the other (D1 → D2 and D2 → D1); CNNs were trained with three random seeds and results averaged. For the random forest, cross-dataset tests were also run after Reinhard stain normalisation [11], which matches the mean and standard deviation of each CIELAB channel of a test image to those of the training dataset, and with shape features only (no colour). The primary metric was balanced accuracy (the mean of per-class recalls), which is robust to the unequal class sizes; macro-averaged F1 and per-class recall were also reported. Because the clinical output of a differential is a set of proportions, the absolute error between predicted and true class percentages in each test set was also computed. Segmentation methods were compared with the Wilcoxon signed-rank test. No confidence intervals were computed for the classification metrics; given the size of the two datasets, the cross-dataset estimates in particular should be read as indicative of the direction and rough magnitude of the drop rather than as precise values.

3. RESULTS AND DISCUSSION

3.1 Nucleus Segmentation

K-means clustering in CIELAB space gave the most accurate and most consistent nucleus segmentation on both datasets, with a mean Dice coefficient of 0.947 (median 0.967) on D1 and 0.960 (median 0.983) on D2 (Table 2 and Figure 2). It was significantly better than every thresholding method on both datasets (Wilcoxon $p < 10-6$, except HSV saturation on D2, $p = 0.022$). The thresholding methods were less robust to the change in staining: HSV saturation and haematoxylin deconvolution performed well on the purple-stained D2 (0.942 and 0.930) but poorly on the yellow-background D1 (0.785 and 0.825), and grey-level Otsu thresholding failed on both. Only 4.3% of D1 and 3.0% of D2 cells had a K-means Dice below 0.80. Eosinophils were the most difficult class (mean Dice 0.81 in D1 and 0.91 in D2), because their dense red-orange granules have a colour similar to the nucleus (Figure 1).

Table 2. Agreement between Unsupervised Nucleus Segmentation and the Expert Nucleus Masks (Dice Coefficient, Mean and Median)

Method	D1 Mean	D1 Median	D2 Mean	D2 Median
Otsu, grey level	0.662	0.631	0.546	0.389
Otsu, HSV saturation	0.785	0.880	0.942	0.978
Otsu, CMYK magenta	0.811	0.897	0.772	0.958
Otsu, haematoxylin (colour deconvolution)	0.825	0.867	0.930	0.962
K-means, CIELAB (darkest cluster)	0.947	0.967	0.960	0.983

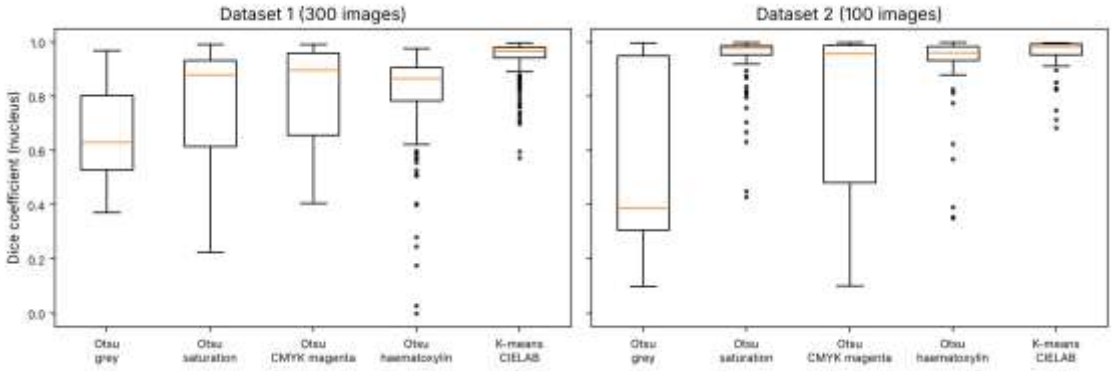


Figure 2. Distribution of the Dice Coefficient between Automatic and Expert Nucleus Masks for Each Unsupervised Method on Dataset 1 (300 Images) and Dataset 2 (100 Images)

3.2 Within-Dataset Classification

When training and test images came from the same dataset, both approaches identified the four leucocyte types well (Table 3). On D1, the CNN reached a balanced accuracy of 0.949 with whole images and 0.951 with nucleus-centred crops, and the morphology-based random forest 0.859 ± 0.010. On the smaller D2 (97 images), the CNN reached 0.797–0.830 and the random forest 0.659 ± 0.058. The random forest built on automatic segmentation performed as well as one built on the expert masks on D1 (0.859 vs. 0.831), which confirms that the K-means segmentation was adequate for feature extraction; on D2 the expert masks gave a higher score (0.752). The most important features were nuclear circularity, the ratio of nuclear area to convex area, solidity and relative nuclear size, followed by the redness (a*) of the perinuclear region, which separates eosinophils. Colour augmentation reduced within-dataset accuracy, markedly so on D2 (0.471). Estimated differential counts from cross-validated predictions were close to the true proportions, with a mean absolute error of 1.0–1.5 percentage points for the random forest and the whole-image CNN (Table 4).

Table 3. Balanced Accuracy and Macro F1 for Four-Class Leucocyte Classification within Each Dataset (Cross-Validation) and Across Datasets (Train on One, Test on the Other). Cross-Dataset CNN Results Are Mean ± SD over Three Seeds; Random-Forest Within-Dataset Results Are Mean ± SD over Three Repetitions. These Ranges Describe Variability over Seeds or Data Splits, Not Sampling Uncertainty; because the Test Sets Contain 299 and 97 Cells, the Sampling Uncertainty of Every Estimate is Substantially Wider and Confidence Intervals Are Not Reported

Model / Input	Within D1 BA	Within D2 BA	D1 → D2 BA	D2 → D1 BA	D1 → D2 F1	D2 → D1 F1
RF, morphology (expert nucleus mask)	0.831 ± 0.023	0.752 ± 0.025	—	—	—	—
RF, morphology (automatic nucleus mask)	0.859 ± 0.010	0.659 ± 0.058	0.485	0.707	0.473	0.572
RF, shape features only	—	—	0.509	0.634	0.508	0.586
RF, + Reinhard stain normalisation	—	—	0.456	0.602	0.458	0.451
CNN, whole image	0.949	0.797	0.379 ± 0.038	0.347 ± 0.047	0.257	0.239
CNN, whole image + colour augmentation	0.843	0.471	0.325 ± 0.031	0.461 ± 0.026	0.294	0.414
CNN, nucleus-centred crop	0.951	0.819	0.303 ± 0.054	0.317 ± 0.033	0.153	0.222
CNN, nucleus-centred crop, greyscale	0.881	0.830	0.323 ± 0.041	0.396 ± 0.089	0.221	0.299

3.3 Cross-Dataset Classification

Performance fell sharply when a model trained on one dataset was applied to the other (Table 3 and Figure 3). The CNN, which was the best model within each dataset, dropped to a balanced accuracy of 0.30–0.46 across datasets, close to the chance level of 0.25 for four classes. It did not fail randomly but systematically: trained on D1 and tested on D2, the whole-image CNN assigned 71% of cells to the monocyte class, and trained on D2 and tested on D1 it classified only 1% of cells as neutrophils, although neutrophils made up 59% of D1 (Figure 4). Neither colour augmentation nor scale-harmonised, nucleus-centred crops (in colour or greyscale) restored cross-dataset performance, although colour augmentation and greyscale crops gave a modest improvement for D2 → D1 (0.461 and 0.396 vs. 0.347).

The interpretable random forest was considerably more robust. It retained a balanced accuracy of 0.707 when trained on D2 and tested on D1, and 0.485 in the opposite direction, outperforming every CNN variant in both directions. Restricting it to colour-free shape features gave similar results (0.509 and 0.634), whereas Reinhard stain normalisation of the test images reduced accuracy (0.456 and 0.602). The errors in the resulting differential counts were nevertheless large for all models (Table 4): the mean absolute error across the four classes was 12.9–14.0 percentage points for the random forest and 17.4–28.9 points for the

whole-image CNNs with or without colour augmentation, with errors of up to 58 points for a single cell type.

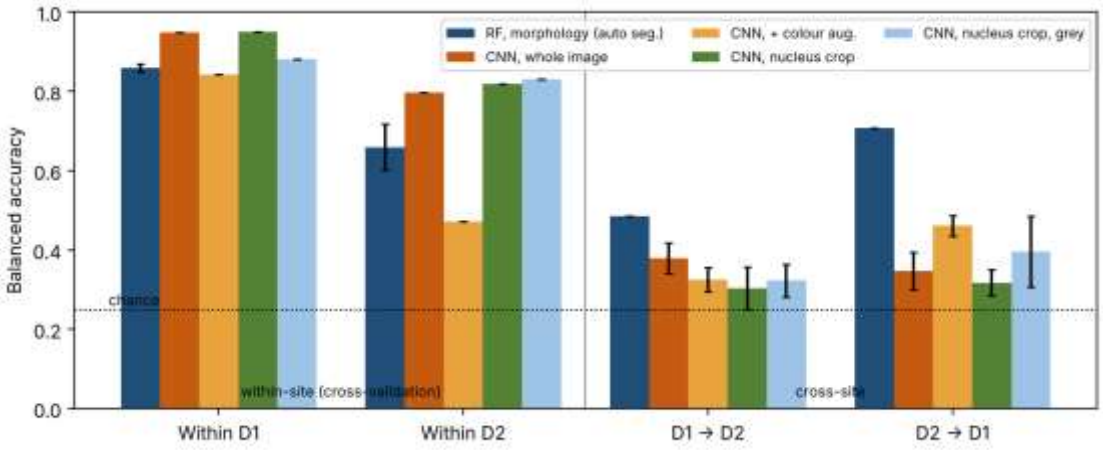


Figure 3. Balanced Accuracy of the Morphology-Based Random Forest (RF) and Four CNN Variants within Each Dataset and Across Datasets. Error Bars Show the Standard Deviation over Repetitions or Seeds Where Available. The Dotted Line Marks Chance Level for Four Classes

Table 4. Leucocyte Differential Estimated by the Classifiers Compared with the Manual (True) Differential (Percentage of Cells). MAE, Mean Absolute Error over the Four Classes in Percentage Points

Setting	Model	Neutrophil	Lymphocyte	Monocyte	Eosinophil	MAE
D1 (true)	—	58.9	17.7	16.1	7.4	—
Within D1	RF / CNN	57.2 / 56.5	17.1 / 18.4	17.7 / 16.1	8.0 / 9.0	1.2 / 1.2
D2 → D1	RF	31.8	27.4	15.1	25.8	14.0
D2 → D1	CNN (whole image)	1.0	41.1	31.4	26.4	28.9
D2 (true)	—	34.0	38.1	15.5	12.4	—
Within D2	RF / CNN	34.0 / 33.0	35.1 / 39.2	18.6 / 16.5	12.4 / 11.3	1.5 / 1.0
D1 → D2	RF	51.5	17.5	23.7	7.2	12.9
D1 → D2	CNN (whole image)	24.7	4.1	71.1	0.0	27.8

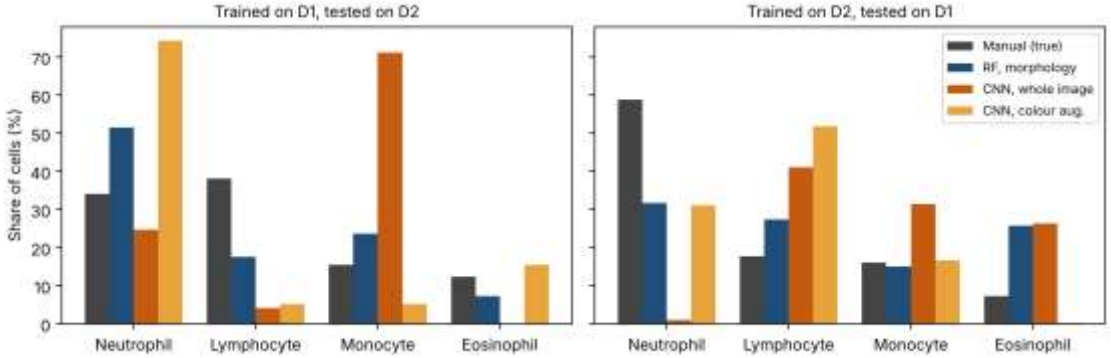


Figure 4. True (Manual) and Predicted Leucocyte Differentials When Models Are Trained on One Dataset and Applied to the Other

3.4 Discussion

This study evaluated automatic leucocyte segmentation and classification on real stained blood-smear images from two sources with expert annotations. Three findings are relevant to the use of such

systems for diagnosis. First, unsupervised K-means clustering in CIELAB space segmented leucocyte nuclei with a Dice coefficient of about 0.95 on both datasets, without any training data or parameter tuning, and was more robust to staining differences than thresholding of single colour channels. This agrees with earlier work showing that colour-based unsupervised segmentation can be effective for leucocytes [5], [13] and offers a simple, transparent first step for low-cost systems.

Second, within a single source, both a small CNN trained from scratch and a random forest on a handful of morphological features classified the four major leucocyte types with high balanced accuracy and reproduced the differential count to within about 1–2 percentage points. These results are consistent with the high accuracies reported for single-site datasets [4], [6], [7].

Third, and most importantly, within-site accuracy was a poor predictor of performance at another site. The CNN, the best model within each dataset, fell to near-chance accuracy and produced differentials that would be clinically misleading, for example reporting an almost complete absence of neutrophils in a population in which they were the majority. Such an error could mask neutrophilia in bacterial infection or simulate neutropenia, with direct consequences for patient management [1], [2]. The datasets differ not only in stain colour but also in magnification, image resolution, focus and the amount of surrounding red cells. Colour augmentation, greyscale input and scale-harmonised crops each addressed part of this difference but none was sufficient, which suggests that the CNN learned dataset-specific texture and context cues [9]. Reinhard colour normalisation even degraded the feature-based classifier, probably because matching global colour statistics alters the relative colour of the nucleus and cytoplasm when the proportion of background and red cells differs between images. In contrast, morphological features computed from a robust segmentation transferred better, because the shape of the nucleus, the most important diagnostic criterion for leucocyte classification, is largely independent of staining. Even this approach, however, produced differential errors of 13–14 percentage points across sites, which is too large for diagnostic use without local validation.

For the prevention, diagnosis and management of disease, these results have practical implications. Automated differential systems should be validated on images from the site and staining protocol where they will be used, and reported accuracy should include an external test set rather than only cross-validation [3], [10]. Interpretable, segmentation-based features provide a useful, more transferable baseline and allow errors to be traced to specific morphological criteria. Local recalibration with a small number of annotated cells, multi-site training data and stain-invariant representations are likely to be needed before such tools can support decisions in primary care and resource-limited laboratories.

The study has limitations. The datasets are small (299 and 97 cells), contain only single, centred cells and are imbalanced, and basophils could not be evaluated. Both datasets come from one publication, and D2 consists of curated example images rather than consecutive clinical samples. No pretrained networks were used, so larger CNNs pretrained on natural or haematological images might transfer better. The clinical consequences of the observed errors were inferred from the predicted proportions, not assessed on patients. Finally, the scale-harmonisation step relied on the automatic segmentation and inherits its errors, particularly for eosinophils.

4. CONCLUSION

Using real blood-smear images from two public sources with expert annotations, this study showed that unsupervised K-means colour clustering segments leucocyte nuclei with a Dice coefficient of about 0.95, and that both a small CNN and an interpretable morphology-based random forest classify the four major leucocyte types accurately within a single source (balanced accuracy up to 0.95 and 0.86) and reproduce the differential count within 1–2 percentage points. Across sources, however, the CNN fell to near-chance accuracy and produced grossly incorrect differentials, while the morphology-based classifier was more robust but still erred by 13–14 percentage points. Colour augmentation, scale harmonisation and stain normalisation did not solve the problem. Automated leucocyte differentials intended for diagnosis should therefore be validated externally and locally before clinical use, and interpretable morphological

features offer a transparent and more transferable baseline. Future work will use larger multi-centre datasets, pretrained models and domain adaptation with small local calibration sets.

Author Contribution Statement

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Dr. Anuj Gaba		✓	✓		✓				✓		✓			✓
Dr. Dhara Andipara		✓	✓	✓		✓		✓				✓		
Atabekov Tenirberdi	✓		✓		✓		✓		✓		✓		✓	✓

C: **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

Acknowledgments

The authors acknowledge Jiangxi Tecom Science Corporation, China, and the CellaVision blog for the images in Datasets 1 and 2, X. Zheng and colleagues for making the annotated datasets publicly available, and the developers of scikit-image, scikit-learn and PyTorch, and the researchers whose works have been cited in this study.

Funding Information

This research has received no external funding.

Conflict of Interest Statement

No conflict of interest.

Informed Consent

Not Applicable. The study used only publicly available, anonymised images.

Ethical Approval

Not Applicable. The study used only publicly available, anonymised images and involved no human participants.

Data Availability

This study uses the publicly available WBC image datasets of Zheng et al. [14]. No new datasets were created. The analysis code is available from the corresponding author on reasonable request.

Use of Generative AI

Generative AI tools were not used to generate the scientific content, data, analyses or results of this study. Any AI-assisted language editing was reviewed and revised by the authors, who take full responsibility for the content of this article.

REFERENCES

[1] B. J. Bain, "Diagnosis from the blood smear," New England Journal of Medicine, vol. 353, no. 5, pp. 498–507, 2005. doi.org/10.1056/NEJMra043442

[2] P. W. Barnes, S. L. McFadden, S. J. Machin, and E. Simson, "The international consensus group for hematology review: Suggested criteria for action following automated CBC and WBC differential analysis," Laboratory Hematology, vol. 11, no. 2, pp. 83–90, 2005. doi.org/10.1532/LH96.05019

[3] A. Kratz et al., "Digital morphology analyzers in hematology: ICSH review and recommendations," International Journal of Laboratory Hematology, vol. 41, no. 4, pp. 437–447, 2019. doi.org/10.1111/ijlh.13042

[4] S. H. Rezatofighi and H. Soltanian-Zadeh, "Automatic recognition of five types of white blood cells in peripheral blood," Computerized Medical Imaging and Graphics, vol. 35, no. 4, pp. 333–343, 2011. doi.org/10.1016/j.compmedimag.2011.01.003

[5] L. Putzu, G. Caocci, and C. Di Ruberto, "Leucocyte classification for leukaemia detection using image processing techniques," Artificial Intelligence in Medicine, vol. 62, no. 3, pp. 179–191, 2014. doi.org/10.1016/j.artmed.2014.09.002

[6] C. Matek, S. Schwarz, K. Spiekermann, and C. Marr, "Human-level recognition of blast cells in acute myeloid leukaemia with convolutional neural networks," Nature Machine Intelligence, vol. 1, no. 11, pp. 538–544, 2019. doi.org/10.1038/s42256-019-0101-9

[7] A. Acevedo, A. Merino, S. Alférez, Á. Molina, L. Boldú, and J. Rodellar, "A dataset of microscopic peripheral blood cell images for development of automatic recognition systems," Data in Brief, vol. 30, Art. no. 105474, 2020. doi.org/10.1016/j.dib.2020.105474

[8] D. Tellez et al., "Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology," Medical Image Analysis, vol. 58, Art. no. 101544, 2019. doi.org/10.1016/j.media.2019.101544

[9] J. R. Zech, M. A. Badgeley, M. Liu, A. B. Costa, J. J. Titano, and E. K. Oermann, "Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study," PLoS Medicine, vol. 15, no. 11, Art. no. e1002683, 2018. doi.org/10.1371/journal.pmed.1002683

[10] C. J. Kelly, A. Karthikesalingam, M. Suleyman, G. Corrado, and D. King, "Key challenges for delivering clinical impact with artificial intelligence," BMC Medicine, vol. 17, Art. no. 195, 2019. doi.org/10.1186/s12916-019-1426-2

[11] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color transfer between images," IEEE Computer Graphics and Applications, vol. 21, no. 5, pp. 34–41, 2001. doi.org/10.1109/38.946629

[12] M. Macenko et al., "A method for normalizing histology slides for quantitative analysis," in Proc. IEEE Int. Symp. Biomedical Imaging (ISBI), 2009, pp. 1107–1110. doi.org/10.1109/ISBI.2009.5193250

[13] X. Zheng, Y. Wang, G. Wang, and J. Liu, "Fast and robust segmentation of white blood cell images by self-supervised learning," Micron, vol. 107, pp. 55–71, 2018. doi.org/10.1016/j.micron.2018.01.010

[14] X. Zheng, "WBC image dataset (segmentation_WBC)," GitHub repository. Accessed: Oct. 2026. github.com/zxaoyou/segmentation_WBC

[15] N. Otsu, "A threshold selection method from gray-level histograms," IEEE Transactions on Systems, Man, and Cybernetics, vol. 9, no. 1, pp. 62–66, 1979. doi.org/10.1109/TSMC.1979.4310076

[16] A. C. Ruifrok and D. A. Johnston, "Quantification of histochemical staining by color deconvolution," Analytical and Quantitative Cytology and Histology, vol. 23, no. 4, pp. 291–299, 2001. pubmed.ncbi.nlm.nih.gov/11531144/

[17] S. Lloyd, "Least squares quantization in PCM," IEEE Transactions on Information Theory, vol. 28, no. 2, pp. 129–137, 1982. doi.org/10.1109/TIT.1982.1056489

[18] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001. doi.org/10.1023/A:1010933404324

[19] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Proc. Int. Conf. Learning Representations (ICLR), 2015. doi.org/10.48550/arXiv.1412.6980

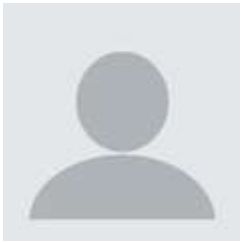
[20] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in Advances in Neural Information Processing Systems (NeurIPS), vol. 32, 2019, pp. 8024–8035. doi.org/10.48550/arXiv.1912.01703

[21] S. van der Walt et al., "scikit-image: Image processing in Python," PeerJ, vol. 2, Art. no. e453, 2014. doi.org/10.7717/peerj.453

[22] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011. jmlr.org/papers/v12/pedregosa11a.html

How to Cite: Dr. Anuj Gaba, Dr. Dhara Andipara, Atabekov Tenirberdi. (2026). Automated white blood cell differential from blood-smear images: Segmentation accuracy and cross-laboratory generalisation on public datasets. Journal of Prevention, Diagnosis and Management of Human Diseases (JPDMHD), 6(2), 43-53. https://doi.org/10.55529/jpdmhd.62.43.53

BIOGRAPHIES OF AUTHORS

	**Dr. Anuj Gaba**, is a Junior Resident (Academic) in the Department of Forensic Medicine and Toxicology at AIIMS Kalyani, West Bengal, India, with a foundation in clinical medicine and a growing specialisation in medico-legal practice. Dr. Gaba has experience in detailed medico-legal examinations, post-mortem evaluations and forensic documentation, with an emphasis on accuracy, ethics and evidence-based interpretation, and works with clinical teams, law enforcement and the judiciary. Dr. Gaba is engaged in teaching, case discussions and research aimed at advancing forensic standards and medico-legal education. Email: anuj.gaba989@gmail.com
	**Dr. Dhara Andipara**, is a Bachelor of Dental Surgery (BDS) graduate with 2.5 years of experience as a staff member at Geetanjali Medical College and Hospital, along with clinical experience. She has two patents and one copyright to her credit. She has also completed an online fellowship in medical oncology through AIIMS, New Delhi, and has one research paper published in dentistry.
	**Atabekov Tenirberdi** is a lecturer in public health and family medicine at Jalal-Abad International University in Manas, Kyrgyzstan. On this study, Atabekov Tenirberdi worked alongside co-authors from AIIMS Kalyani and Geetanjali Dental and Research Institute in India to assess automated white blood cell differentials and how far their accuracy holds when a model is moved between laboratories.