International Journal of Laboratory Hematology
The Official journal of the International Society for Laboratory Hematology

Letter to the Editor

Examination of peripheral blood smears: performance evaluation of a digital microscope system using a large-scale leukocyte database

Sir, The analysis of blood morphology is of great diagnostic importance to the clinician. Manual morphological assessment using the microscope has been considered the gold standard for years but can be vulnerable to interobserver variability, is labor intensive, and requires highly and continuously trained personnel [1–3]. An exciting development in the field is the introduction of digital microscope (DM) systems. A DM ensures the constant presence of a morphological expert in the routine laboratory and enables the automated recognition of (pathological) cell types [3–5].

It was previously shown that the classification performance of the DM is equal to manual performance when classifying the five main peripheral blood cell classes (neutrophils, lymphocytes, monocytes, eosinophils, and basophils) [3–9]. However, these studies either used a low number of samples and cells or did not include a combination of normal and abnormal peripheral blood smears (PBS) [3–9]. Several studies also compared the postclassification results (which include manual interference) with manual analysis preventing a clear view on the ability of the DM to correctly classify cells without manual interference [4, 6, 7]. Here, we present a large-scale database of about 1.4 million leukocytes from both normal and abnormal PBS, pitting the DM's preclassification performance against the gold standard.

Methods

Patient samples

Venous blood was collected using K3-EDTA as anticoagulant and stored at room temperature until further analysis. Within 4 h of collection, blood smears were prepared and stained according to Romanowsky (May-Grunwald/Giemsa/Wright), using an automated slide preparation unit (SP-100, Sysmex, Kobe, Japan). The number of cells analyzed per slide was set at 200, both for manual assessment and assessment by the DM. Manual assessment, set as the gold standard, was defined as analysis of a slide by an experienced morphology expert by reviewing the digital images provided by the DM. The preclassification performance, defined as the initial classification by the DM without manual intervention, was compared to this gold standard. The samples were selected from our laboratory which handles routine samples from both general practitioners and hospitals, including a hemato-oncology ward.

Automated microscopy system

For this study, the DM96 (CellaVision, Lund, Sweden), described in reference 3, was used as DM system, operated with the Cellavision Blood differential module (Version 2.0).

Statistics

Statistical analysis was carried out using Statistical Package for Social Sciences (SPSS) version 18 and MedCalc version 15, both for Windows. The study design was based on the National Committee for Clinical Laboratory Standard document H20A. To determine accuracy, the percentage of each class per sample found by the DM was compared to the percentage found by the manual assessment. Results were analyzed according to Bland and Altman [10] and by determining the Pearson product–moment correlation coefficient (R). To generate Bland–Altman plots, the difference (percentage DM – percentage expert) per sample was shown as a function of the mean result of both DM and expert. Mean difference and limits of agreement (mean difference $\pm$ 1.96*standard deviation) were determined and added to the plots as references lines. Constant and proportional biases were assessed using Deming regression.

Results

A total of 6945 PBS were analyzed with approximately 1.4 million classified cells. When possible, 200 cells per sample were analyzed. In 214 cases (3.1%), the DM did not reach the number of 200 cells due to leukopenic samples. In 57 cases (0.8%), the number of counted cells was below 100 (minimum 16 cells). As the number of cells in each class was expressed as a percentage of the

total number of counted cells, these samples were included in the analysis.

The results for the different leukocyte classes and NRBCs can be found in Table 1 and Figure 1. Preclassification performance has an excellent accuracy for the five main blood cell classes and nucleated red blood cells (NRBCs). The DM software is currently unable to recognize promonocytes, prolymphocytes, hairy cells, and cleaved cells; these were only counted by the experts. Proportional bias was found to be present for the five main classes and metamyelocytes and ranged from 0.5% for neutrophils to 23.3% for metamyelocytes. Constant bias was found to be present for nine classes and ranged from 0.04% for basophils to 0.39% for blast cells.

For blast cells, the limit for a positive finding was set at 0 for the DM, defining a percentage above 0.0% as positive. With this limit, the DM achieved a blast cell sensitivity of 100% with a specificity of 67%.

Discussion

Before digital microscopy can be accepted as an improvement over the current manual method and used as a standardized diagnostic tool, it is necessary to establish that the systems are as reliable as manual assessment [4]. A large database of leukocytes was used to compare the preclassification performance of the DM to morphological experts.

The detection of blast cells is essential for the correct and early diagnosis of patients with hematological malignancies. It is therefore of extreme importance that the DM does not miss any blast cells (i.e., displays a high sensitivity). In the current study, the DM achieved a blast cell sensitivity of 100% and a specificity of 67%. The rather low specificity indicates that, quite often, the DM classifies a leukocyte as a blast cell, while the cell actually belongs to a different class. In several cases, the expert found a higher percentage of blast cells compared with the DM. While the DM shows excellent sensitivity for blast cells, it does remain necessary for the operator to routinely check the exact percentage.

Excellent accuracy was found for the five main cell classes as shown by the small limits of agreement approaching 0, in combination with the high correlation. The mean differences shown in the Bland–Altman plots were small enough to be considered clinically insignificant [10]. However, rare individual cases may show large differences between the results obtained by the DM and by the expert. In these cases, the DM generally classified cells into one of the five main classes, while the expert classified the cells into one or several of the less common classes.

A statistically significant proportional bias was present for the five main classes but was too small to be considered clinically relevant, as determined by an experienced hematologist. Constant bias was found for nine classes, including four of the five main classes, but was small enough to be considered negligible. However, when analyzing a database of this size, even small biases are statistically significant.

The DM's performance for less common classes ranged from adequate (blast cells, correlation = 0.840) to poor (promyelocytes, correlation = 0.432), with small mean differences. However, the majority of samples did not con-

Table 1. Overview of the correlation, mean difference, limits of agreement, regression equation, and the corresponding 95% confidence intervals for the intercept and the slope of the regression equation. The classes promonocytes, prolymphocytes, hairy cells, and cleaved cells cannot yet be detected by the DM and were therefore not included in this table. The mean difference and limits of agreement were added to the Bland–Altman plots as reference lines. The regression equation and the corresponding 95% confidence intervals were used to assess constant and proportional bias. Significant biases are highlighted with an asterisk

Cell classes	Correlation	Mean difference	Limits of agreement	Regression equation	95% CI intercept	95% CI slope
Neutrophils (band and segmented)	0.997	−0.147	−3.385–3.091	0.16 + 0.99 x	0.05–0.27*	0.99–0.99*
Lymphocytes (also variant)	0.995	−0.461	−4.767–3.845	0.38 + 0.97 x	0.31–0.46*	0.97–0.98*
Monocytes	0.933	−0.151	−4.140–3.838	0.21 + 0.95 x	−0.10–0.51	0.90–0.99*
Eosinophils	0.978	−0.099	−1.339–1.141	0.09 + 0.91 x	0.07–0.11*	0.90–0.92*
Basophils	0.928	−0.062	−0.864–0.740	0.04 + 0.84 x	0.01–0.07*	0.80–0.89*
Blast cells	0.840	0.367	−3.298–4.032	0.39 + 0.93 x	0.36–0.41*	0.79–1.07
Promyelocytes	0.432	0.027	−1.361–1.415	0.05 + 0.60 x	−0.01–0.10	−0.99–2.18
Myelocytes	0.808	0.139	−1.707–1.985	0.16 + 0.94 x	0.08–0.25*	0.68–1.21
Metamyelocytes	0.802	−0.021	−2.037–1.995	0.09 + 0.77 x	0.02–0.15*	0.59–0.95*
Plasma cell	0.576	0.114	−1.121–1.349	0.13 + 0.67 x	0.09–0.16*	−0.61–1.94
NRBC	0.958	0.333	−1.660–2.326	0.32 + 1.03 x	0.30–0.34*	0.99–1.06

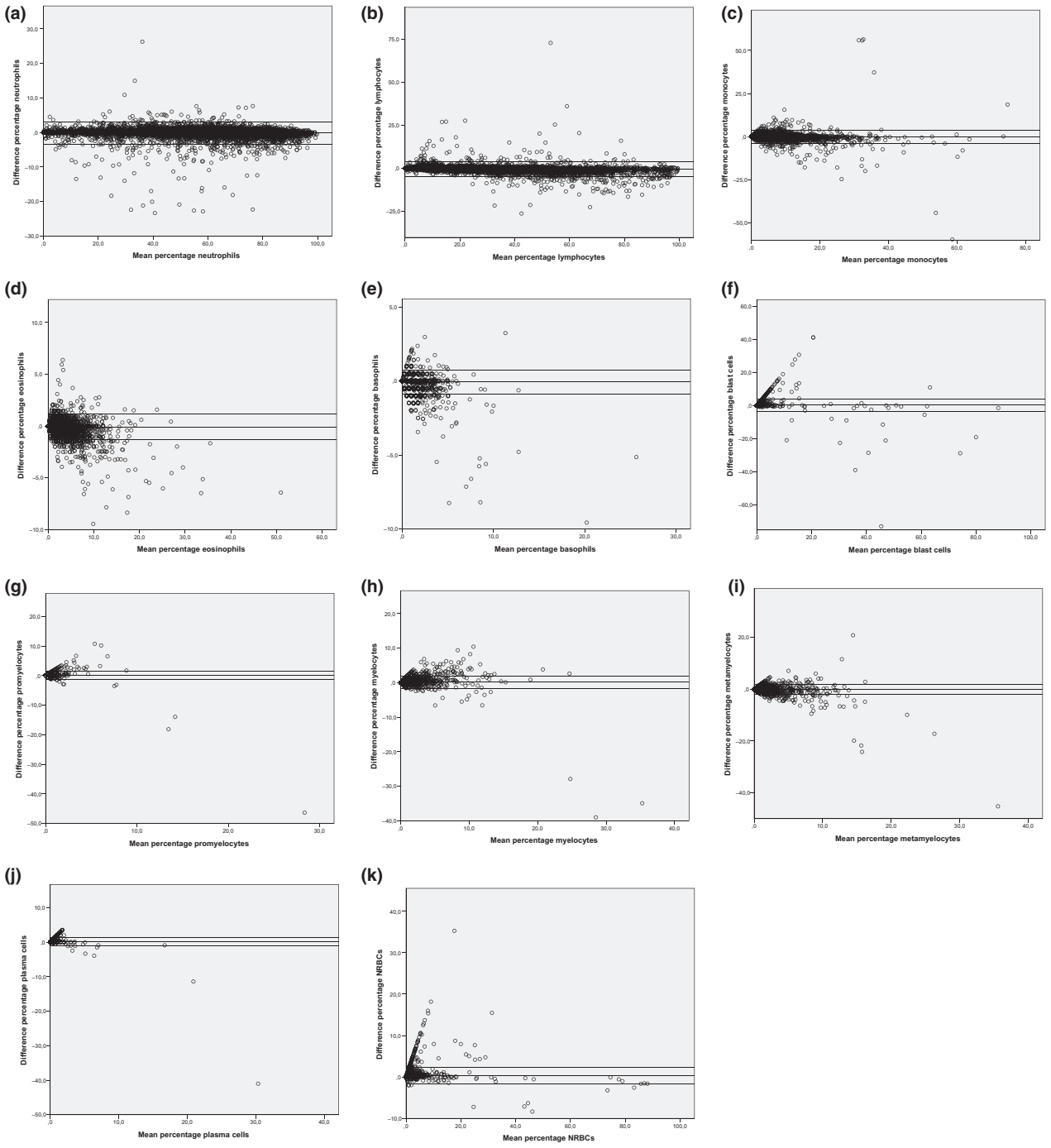

Figure 1. Bland–Altman plots for (a) neutrophils (b) lymphocytes (c) monocytes (d) eosinophils (e) basophils (f) blast cells (g) promyelocytes (h) myelocytes (i) metamyelocytes (j) plasma cells, and (k) NRBCs.

tain these classes according to both the DM and the expert. The calculated difference was 0 in these cases, markedly lowering the mean. For these classes, Bland–Altman plots may not be the most fitting method for the analysis of the results. Metamyelocytes showed a clinically significant proportional bias. However, the algorithms used by the

DM for this class, and for the other less common classes, are still in development. Further work is needed to enhance the DM's performance for these classes.

Four rare but clinically significant classes – promonocytes, prolymphocytes, hairy cells, and cleaved cells – cannot yet be detected by the DM. The operator can use the overview option of the DM to review the complete slide after preclassification to check the DM's performance and detect these rare classes. In practice, the DM can be used as a screening tool for peripheral blood smears, saving time, and reducing workload. The presence of the operator is still required to ensure the proper classification of the less common and rare cell classes.

The next step in morphology will be automated assessment of blood samples, which will allow a blood sample to be processed by a cell counter, an automated preparation unit, and a DM system without any manual intervention, while the results obtained can be sent to the laboratory information system without manual confirmation. This will decrease labor costs (an important issue in today's healthcare system), minimize interobserver variability, and reduce reporting time for morphological assessment of PBS. The database described here will next be used to assess the possibility of autovalidation of a DM system, which will, if successful, exclude manual interference.

In conclusion, the DM is capable of an excellent performance for the five main blood cell classes and blast cells, but at this moment, manual intervention remains necessary to 'help' the system with the less common classes and the occasional outlier. The algorithms used by the DM to classify the less common classes do require further refinement to improve the DM's preclassification performance. Nonetheless, the current preclassification performance of DM systems is a significant step toward the acceptance of DM systems as the standard diagnostic tool for morphological assessment.

ACKNOWLEDGEMENTS

We thank Annemiek de Pijper-van Poppel, Barbara van de Breevaart- Ardonne, Carla Jongmans, Linda de Bruin, Peggy Brosens, Pleunie de Lint- van Driel, Ria de Jong – Kwikkers, and Ria Teuns for their technical assistance and for performing morphological analysis, Joost van Rosmalen for his advice on statistics and Rebecca Buis for critical reading of the manuscript. JR and WG designed the study and were two of the participating morphological experts. KS analyzed the data. KS and JR wrote the paper. WG and MD critically revised the paper. All authors read and approved the final version of the manuscript. Competing interests: the authors have no competing interests.

K. Stouten*,†, J. A. Riedl*, M.-D. Levin†, W. van Gelder*

*Department of Clinical Chemistry, Albert Schweitzer Hospital, Dordrecht, The Netherlands
†Department of Internal Medicine, Albert Schweitzer Hospital, Dordrecht, The Netherlands
E-mail: j.riedl@asz.nl

K. Stouten and J. A. Riedl contributed equally to this study.

doi: 10.1111/ijlh.12391

References

1. Rümke CL. Imprecision of ratio-derived differential leukocyte counts. Blood Cells 1985;11:311–4.
2. Bentley SA. Automated differential white cell counts: a critical appraisal. Baillieres Clin Haematol 1990;3:851–69.
3. Ceelie H, Dinkelaar RB, van Gelder W. Examination of peripheral blood films using automated microscopy; evaluation of Diffmaster Octavia and Cellavision DM96. J Clin Pathol 2007;60:72–9.
4. Kratz A, Bengtsson HI, Casey JE, Keefe JM, Beatrice GH, Grzybek DY, Lewandrowski KB, van Cott EM. Performance evaluation of the Cellavision DM96 system, WBC differentials by automated digital image analysis supported by an artificial neural network. Am J Clin Pathol 2005;124:770–81.
5. Briggs C, Longair I, Slavik M, Thwaite K, Mills R, Thavaraja V, Foster A, Romanin D, Machin SJ. Can automated blood film analysis replace the manual differential? An evaluation of the Cellavision DM96 automated image analysis system. Int Jnl Lab Hem 2009;31:48–60.
6. Lee LH, Mansoor A, Wood B, Nelson H, Higa D, Naugler C. Performance of CellaVision DM96 in leukocyte classification. J Pathol Inform 2013;4:14.
7. Cornet E, Perol JP, Troussard X. Performance evaluation and relevance of the CellaVision™ DM96 system in routine analysis and in patients with malignant hematological diseases. Int Jnl Lab Hem 2008;30:536–42.
8. Rollins-Raval MA, Raval JS, Contis L. Experience with CellaVision DM96 for peripheral blood differentials in a large multi-center academic hospital system. J Pathol Inform 2012;3:29.
9. Billard M, Lainey E, Armoogum P, Alberti C, Fenneteau O, Da Costa L. Evaluation of the CellaVision™ DM automated microscope in pediatrics. Int Jnl Lab Hem 2010;32:530–8.
10. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods to clinical measurement. Lancet 1986;1:307–10.