ERU Engineering Research Unit University of Moratuwa Sri Lanka

Evaluation of Key Focus Measures for Smear Analysis

D.C Hewavitharana
Department of Mechanical Engineering
University of Moratuwa
Katubedda, SriLanka
dinethh@uom.lk

W.A.D.M Jayathilaka
Department of Mechanical Engineering
University of Moratuwa
Katubedda, SriLanka
dumithj@uom.lk

V.P.C Dassanayake
Department of Mechanical Engineering
University of Moratuwa
Katubedda, SriLanka
palitha@uom.lk

Y.W.R Amarasinghe
Department of Mechanical Engineering
University of Moratuwa
Katubedda, SriLanka
ranama@uom.lk

Keywords— Autofocus algorithm, Focus measures, Microscopy, Blood film analysis.

I. INTRODUCTION

Reliable focus assessment is crucial in microscopic smear analysis, as automated systems' diagnostic accuracy is heavily reliant on image sharpness. Despite the availability of numerous focus measures (FMs), their applicability across biological domains remains unknown. This study employs a weighted, raw value-based generalization approach to evaluate fifty single focus measures across five publicly available Z-stack microscopy datasets. Ten quantitative criteria were used to assess focus curve fidelity and resilience. The raw-value-based generalization score combines weighted mean performance and inter-dataset stability to provide a complete measure of cross-domain reliability.

Many FMs have been proposed, including gradient, statistical-, texture-, and transform-based operators. However, their efficacy varies greatly depending on specimen type, staining process, and optical arrangement. Prior comparative studies [4], [5] typically focused on one or two datasets or imaging modalities, limiting the extent to which their recommendations may be applied to larger smear-analysis contexts.

Recent research has also offered deep learning-based focusing algorithms for digital pathology and whole-slide imaging [5]. While effective, these systems typically require task-specific training data, significant computational resources, and function as black box predictors rather than interpretable attention scores. In contrast, traditional hand-crafted FMs remain appealing for incorporation with resource-constrained or older autofocus systems that require transparency, minimal latency, and ease of deployment.

Our study presents a comprehensive evaluation of fifty single-focus measures from five publicly available z-stack microscopy datasets that include haematological and microbiological smears. A raw-value-based generalization methodology is suggested to measure both performance and inter-dataset stability, with the goal of discovering a limited set of FMs that are consistently dependable across a wide range of smear types and imaging conditions. The collected findings assist researchers and developers in identifying the best focus operators for future biomedical imaging applications.

II. LITERATURE REVIEW

Autofocus algorithms for computer microscopy have been studied for decades. Sun et al. [4] examined a variety of traditional FMs, including gradient-based and statistical operators, and addressed criteria for selecting optimal focus methods in general microscopy. More recently, Bian et al. [5] examined autofocus solutions for whole slide imaging, including optical hardware strategies and software-based approaches like deep learning. But most published studies have focused on small datasets, often from a single laboratory or modality, and have not rigorously assessed cross-dataset robustness in smear analysis. Furthermore, the emphasis has frequently been on assessing FMs within a single experiment rather than measuring generalization across diverse biological and optical settings. This inspires a broader, dataset-aware benchmark for smear microscopy, where staining, background texture, and cellular morphology vary significantly.

III. PROPOSED METHOD

The z-stack image collections were standardized to grayscale, scaled to 128×128 pixels, and aligned to 10 focal planes per stack. The data sets cover five biological and imaging areas as shown in Table I.

TABLE I. Z-STACK DATASETS USED

Dataset	Sample Type	Magnification	Step (µm)	Purpose
WBC [1]	Peripheral blood	50×	0.4	Multi-focus WBC analysis
PBS [2]	Thin smear	100× oil	0.5	Hematology diagnostics
BMA [2]	Bone marrow	100× oil	0.5	Leukemia diagnostics
TBF [2]	Thick smear	100× oil	0.5	Malaria detection
TBI [3]	Sputum smear	100× oil	0.5	Tuberculosis screening

Each dataset was weighted based on its clinical relevance and sample size: WBC (25%), TBF (20%), PBS (15%), BMA (20%), and TBI. Each focus measure creates a z-axis focus curve $F(z)$ for each image stack. Ten quantitative descriptors were calculated, focusing on curve accuracy, stability, and sharpness. Accuracy, False Maxima, Range, FWHM, and Curvature at Peak were the primary measures, with Steep

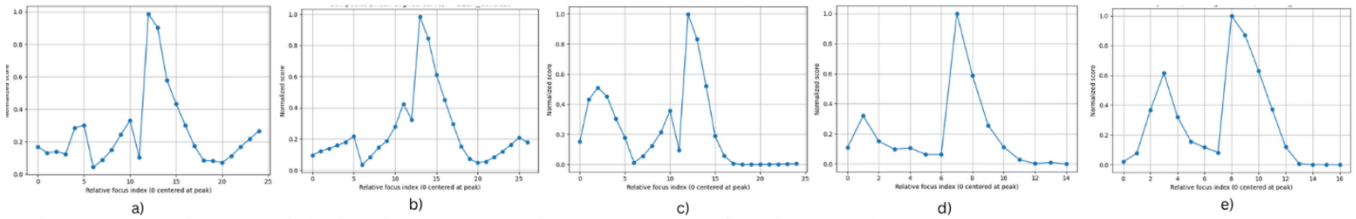

Fig 1. Mean focus curve behavior of GLCM Contrast focus measure across five microscopy datasets. (a–e) shows GLCM Contrast evaluated on (a) Bone Marrow Aspirate (BMA), (b) Peripheral Blood Smear (PBS), (c) Thick Blood Film (TBF), (d) Tuberculosis Smear (TBI), and (e) White Blood Cell (WBC-MF) datasets.

Slope Width (Ws), Steep-to-Gradual Ratio (Rsg), Noise Level, RRMSE, and Execution Time serving as auxiliary metrics. The weighted generalization score was derived using dataset k, focus measure i, and metric j as:

$$G_i = \alpha \overline{V_i} + (1 - \alpha) SD_i$$

$\overline{V_i}$ is the weighted mean of raw metric values, whereas SD_i is the standard deviation across datasets. Pilot evaluations were used to optimize the trade-off coefficient α =0.7 and metric weights, resulting in a balanced sensitivity between performance magnitude and stability across datasets. Lower scores indicate greater generalization and stability. Weights have been determined empirically (Accuracy = 0.20, Range = 0.15, False Maxima = 0.10, FWHM = 0.10, Noise Level = 0.10, Rsg = 0.10, Execution Time = 0.10, Ws = 0.05, Curvature = 0.05, and RRMSE = 0.05). This underlines the diagnostic value and operational viability of real-time microscopy systems.

IV. RESULTS AND DISCUSSION

Using the raw-value-based scoring approach, the top ten focus measures for smear analysis are shown in Table II.

TABLE II. TOP TEN FOCUS MEASURES BY RAW-VALUE-BASED GENERALIZATION SCORE

Focus Measure	Weighted Mean	Weighted Std	Gen Score	Rank
GLCM Contrast	3.406079637	0.8515953	2.639734336	1
Sum Modified Laplacian	3.346297081	1.060808755	2.660650583	2
Brenner Gradient	3.410329164	0.965270815	2.676811659	3
Variance of Laplacian	3.373512969	1.091594163	2.688937327	4
Wavelet W1	3.438397737	1.026267825	2.714758764	5
Wavelet W3	3.43873263	1.026379278	2.715026624	6
Curvelet Transform Sharpness Index	3.43884933	1.026318211	2.715089994	7
Wavelet W2	3.438931114	1.026370808	2.715163022	8
Fourier High Frequency Energy Ratio	3.454895311	0.996059671	2.717244619	9
DCT Focus Measure	3.435590213	1.062837235	2.72376432	10

GLCM Contrast surpassed all others, with a generalization score of 2.639, indicating greater stability across datasets and strong discriminative strength for texture-rich smears. Sum Modified Laplacian (SML) and Brenner Gradient came in a close second and third, indicating that edge-derived sharpness operators are highly effective in capturing defocus. In high-noise datasets like TBF and TBI, gradient-based measures (SML and Brenner) produced relatively consistent generalization scores, indicating resilience to staining and contrast variations. Texture measurements (GLCM) performed well in PBS and BMA, where the morphological features are rich in localized contrast. This suggests that

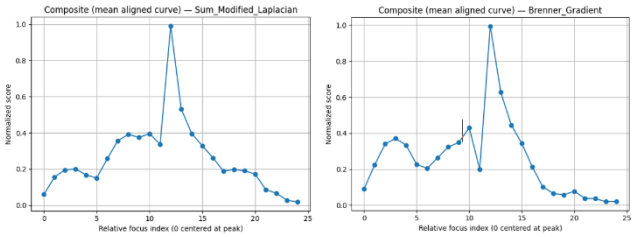

Fig 1. Mean focus curves of f) SML, g) Brenner Gradient

texture-domain operators are particularly effective for heterogeneous cellular samples, whereas gradient-domain operators provide strong focus sensitivity through contrast edge detection.

Multiscale, gradient-, and texture-based focus measures, particularly GLCM Contrast and edge-derived operators such as SML demonstrated strong overall stability across smear datasets. These findings highlight the combined importance of texture-, gradient-, and frequency-domain information in accurate morphological focus evaluation and provide quantitative justification for selecting domain specific focus measures in smear microscopy.

V. CONCLUSION

It is found that five focus metrics, GLCM Contrast, Sum Modified Laplacian, Brenner Gradient, Variance of Laplacian, and Wavelet W1, are the most generalizable and reliable for smear analysis across several microscopy modalities. The findings show that texture and gradient-domain operators outperform classic intensity-based metrics in terms of stability, discriminability, and resistance to noise and staining variability.

REFERENCES

[1] S. Park et al., "A large multi-focus dataset for white blood cell classification," Scientific Data, vol. 11, no. 331, pp. 1–11, 2024. [Online]. Available: https://www.nature.com/articles/s41597-024-03938-1.

[2] A. Smith et al., "High Magnification Z-Stacks from Blood Films," UCL Research Data Repository, University College London, 2021. DOI: 10.5522/04/13402301.

[3] R. da Silva Ferreira et al., "TbImages: A Smear Microscopy Image Dataset to Support the Development of Automated Bacilli Detection in the Diagnosis of Tuberculosis," IEEE DataPort, 2020. DOI: 10.21227/y9mf-0320.

[4] Y. Sun, S. Duthaler, and B. J. Nelson, "Autofocusing in computer microscopy: Selecting the optimal focus algorithm," *Microsc Res Tech*, vol. 65, no. 3, pp. 139–149, 2004, doi: 10.1002/jemt.20118.

[5] Z. Bian *et al.*, "Autofocusing technologies for whole slide imaging and automated microscopy," Dec. 01, 2020, *Wiley-VCH Verlag*. doi: 10.1002/jbio.202000227.