2026 JULY 19–22
VANCOUVER, BC
JOINT AAPM | COMP MEETING

Ultrasound Uniformity: Correlating a Hybrid Quantitative Metric with Expert Assessment

Asher Shertok[1]; Angela Snyder, PhD[1]; Jacqueline Gallet, PhD[2]
[1]Atirix Medical Systems, [2]Department of Radiology, UT Southwestern Medical Center

ABSTRACT

Purpose: Visual assessment of ultrasound uniformity is subjective and prone to inter-observer variability. This study validates a novel, automated metric combining weighted intensity statistics and texture analysis to provide an objective, continuous scale for uniformity, correlated with expert visual ranking.

Methods: Building on previous work establishing algorithms sensitivity to simulated degradation, this study utilizes clinical hardware. The hybrid metric integrates: (1) a Spatially Weighted Standard Deviation using a sigmoid function to prioritize darker regions where non-uniformities are clinically critical, and (2) Local Binary Pattern (LBP) analysis to quantify local texture consistency. Images are processed in normalized patches, where the final index is calculated as the relative difference between the best and worst patches.
Images for visual assessment were chosen from a data set across 6 machines and 45 transducers, spanning 5 years from an academic breast imaging center. Eight images were selected based on their suspected gradient of uniformity. An ultrasound medical physicist expert independently ranked the images on a scale of 1 – 5, with 5 representing a more uniform image. Analysis was then performed to calculate correlation between the algorithm's uniformity index and expert grade.

Results: Across eight ultrasound images with expert grades (3–5), the hybrid uniformity index closely tracked the physicist's visual scoring and improved with increased SWHM weighting. The strongest agreement occurred at α = 0.75, yielding Spearman ρ = 0.945 (p = 0.0004) and Kendall τ = 0.866 (p = 0.005), indicating a robust monotonic relationship between automated uniformity and expert assessment. Lower α values (greater LBP χ^2 influence) reduced concordance.

Conclusion: A hybrid SWHM–LBP metric provides an objective ultrasound uniformity score that closely matches expert visual grading, with α = 0.75 identified as the optimal weighting for this dataset.

Asher Shertok
asher.shertok@atirix.com
651-529-7503

Angela Snyder, PhD
angela.snyder@atirix.com
952-674-4730

Atirix Medical Systems
10201 Wayzata Boulevard, Suite 310
Minneapolis, MN 55305

INTRODUCTION

Ultrasound quality assurance is essential for identifying image artifacts caused by transducer damage, element dropout, or system degradation. Image uniformity assessment is a key part of B-mode ultrasound QC because nonuniform regions, streaks, or shadowing artifacts can indicate degraded probe or scanner performance[1,2].

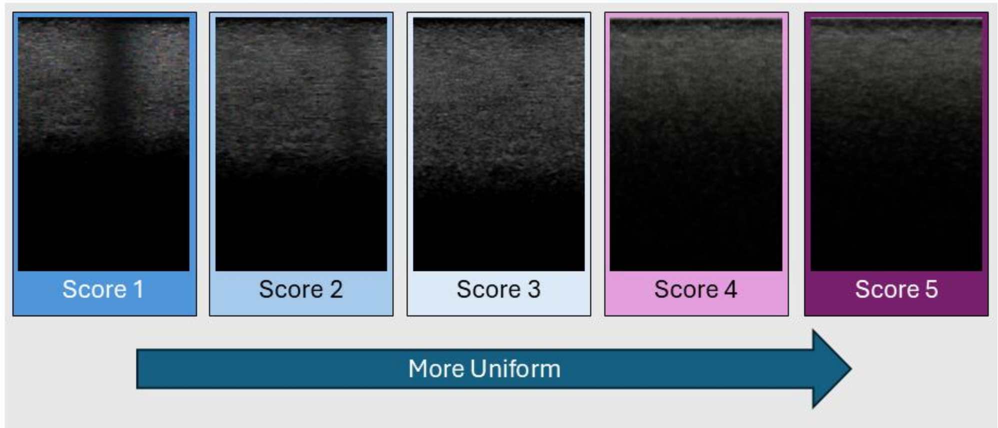

Figure 1. Representative ultrasound transducer uniformity examples arranged from severe non-uniformity appearance to uniform.

Although ultrasound QC programs can be successfully implemented in busy clinical environments, they must remain simple, efficient, and standardized for routine use[3]. Current QC guidance supports routine ultrasound assessment, but uniformity is often judged visually and may miss early subtle defects[2,4]. Automated and texture-based methods may improve objective detection of B-mode nonuniformities[5]. This study evaluates an automated hybrid metric combining intensity variation and local texture analysis to better match expert visual assessment. This algorithm has been previously shown to be sensitive to artificially degraded images, for both brightness and texture non-uniformities.

METHODS AND MATERIALS

The hybrid automated analysis, previously established, combined a spatial weighted histogram to prioritize darker regions through the application of a sigmoid weight function[6], and a local binary pattern (LBP) function. LBP is used to determine how local texture differs across the image. The two metrics are combined through an alpha parameter (α=1, pure SWHM; α=0, pure LBP).

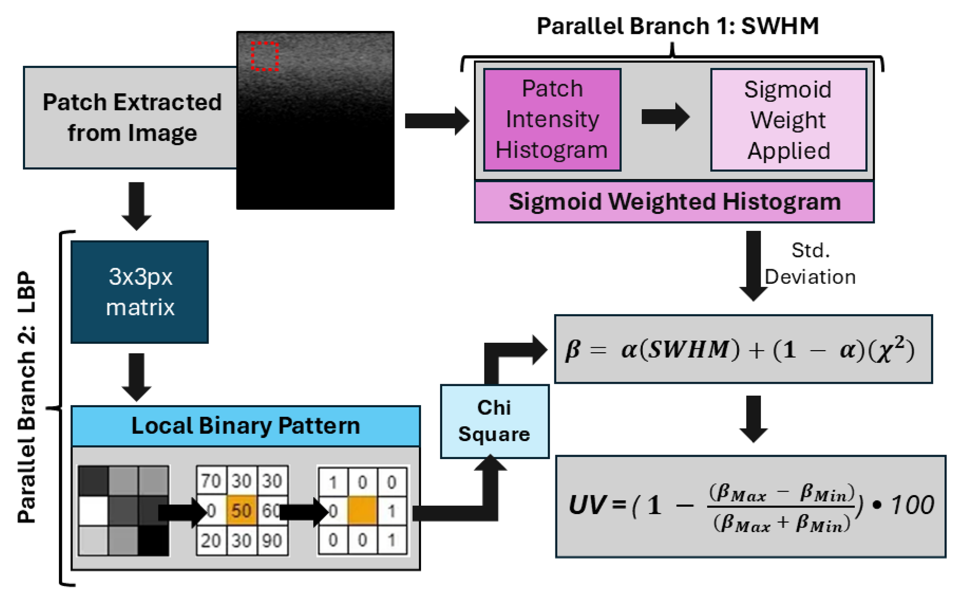

Figure 2. Hybrid ultrasound uniformity metric workflow combining SWHM and LBP features.

Images for analysis were chosen from 6 machines with 45 transducers over 5 years from an academic breast imaging center. Expert ultrasound medical physicists and novice readers ranked 10 classified images on a scale from 1 – 5 (5=high uniformity; 1=low uniformity). Spearman rank-order correlation and Kendall correlation were performed on the two results to correlate and compare visual performance to algorithm uniformity value (UV) scale of 0 to 100.

RESULTS

Alpha=0.75 achieves optimal metric-visual correlation

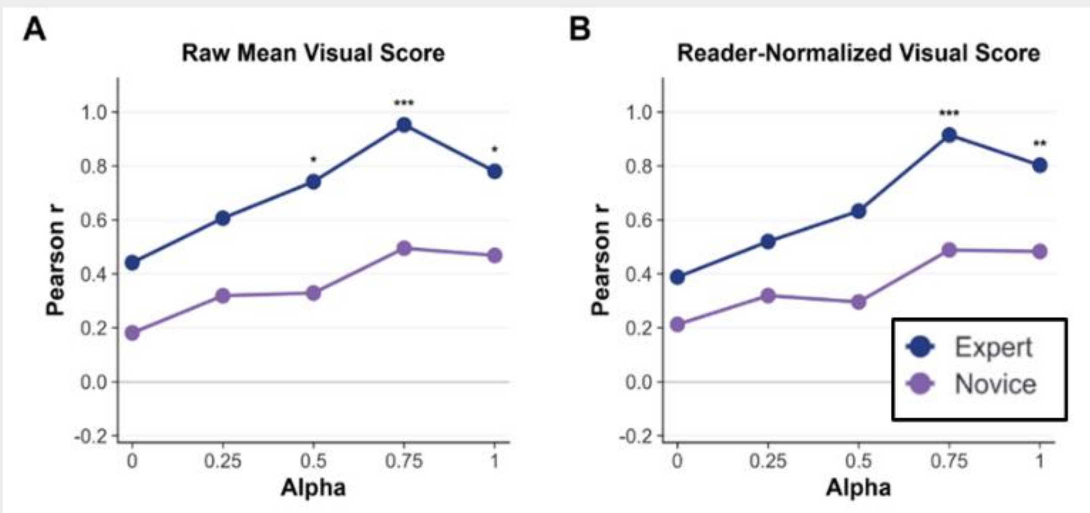

Figure 3. Correlation strength by algorithm parameter and reader expertise. (**A**) Raw visual scores showed strongest Pearson agreement at α = 0.75 overall (r = 0.841, R^2 = 0.707), with expert and novice correlations of r = 0.954 and r = 0.476, respectively. (**B**) Reader-normalized scores also peaked overall at α = 0.75 (r = 0.798, R^2 = 0.637), with expert and novice correlations of r = 0.921 and r = 0.463.

Visual assessment aligns with the automated metric

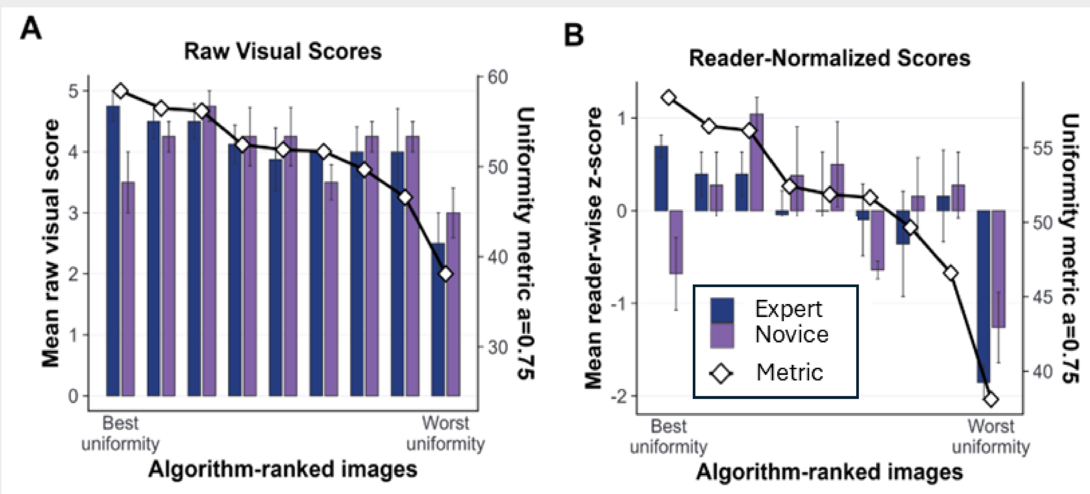

Figure 4. Correlation between visual scores and algorithm metric at α = 0.75. (**A**) Raw mean visual scores with metric overlay (r = 0.841, p = 0.0023, R^2 = 0.707). (**B**) Reader-normalized scores (r = 0.798, p = 0.0057, R^2 = 0.637).

Inter-reader agreements highlights subjectivity

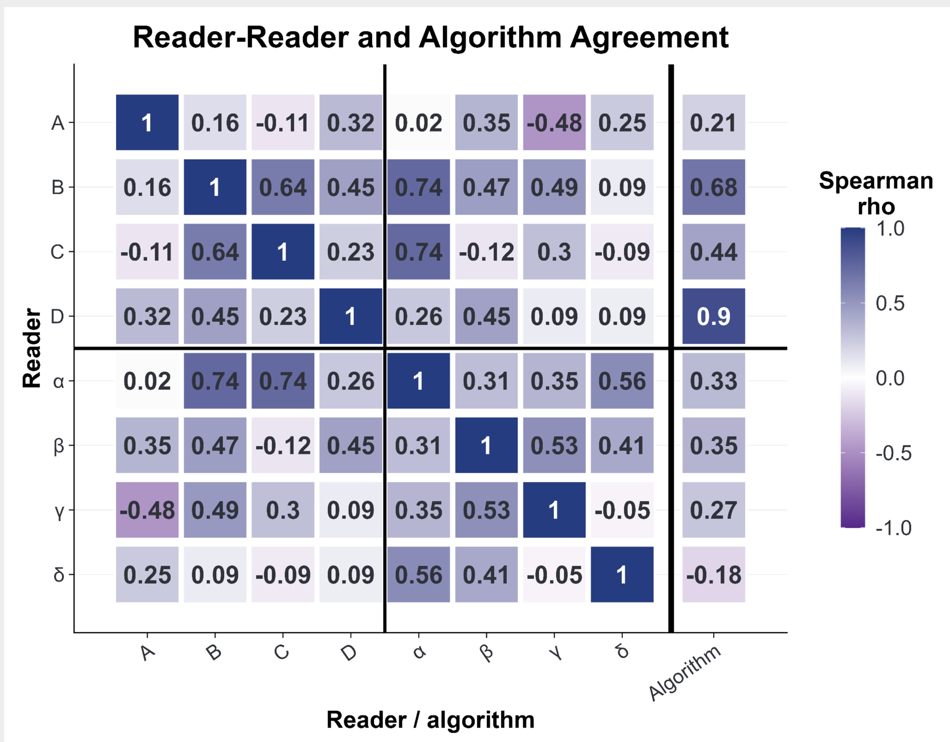

Figure 5. Spearman rho correlations comparing inter-reader agreement. A-D expert visual readers, α-δ novice visual readers.

Reader-algorithm agreement exceeds reader-reader agreement

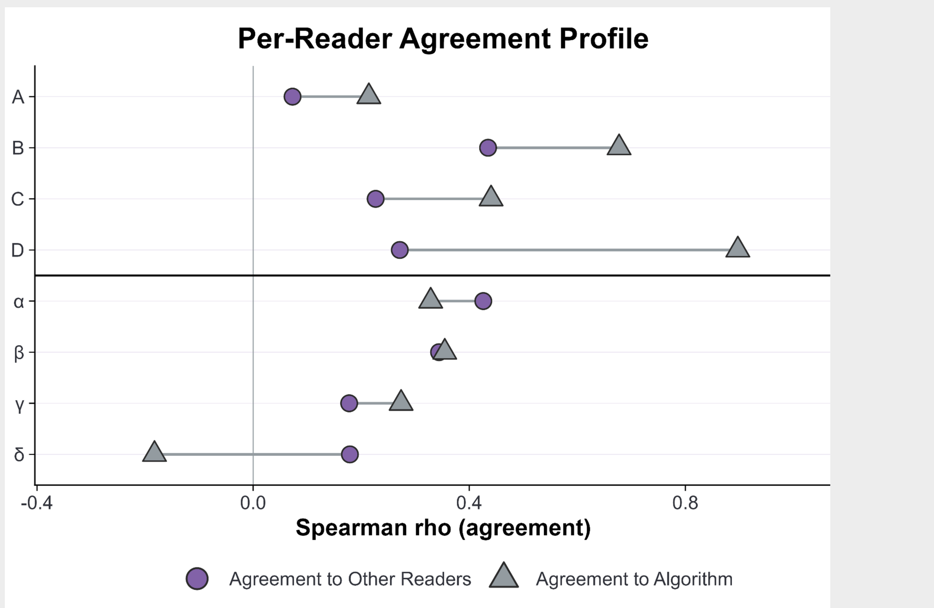

Figure 6. Spearman rho correlations comparing reader-algorithm agreement. Per-reader comparison: circles show mean reader-reader agreement; triangles show reader-algorithm agreement (a=0.75).

DISCUSSION

The automated metric showed **strong agreement with visual assessment**, supporting its ability to reflect clinically meaningful ultrasound uniformity differences.

- SWHM served as the primary driver because it emphasized darker regions associated with dropout, shadowing, and localized signal loss.
- LBP contributed complementary texture information not fully captured by intensity variation alone.

Readers agreeing more with the algorithm than other readers further **highlights the subjectivity of visual ultrasound uniformity assessment** and the efficacy of the algorithm.

By providing a continuous objective score, the metric may help **standardize routine QC, reduce subjectivity, and save medical physicists time** by rapidly identifying high-uniformity images while flagging cases requiring closer review.

These results suggest that the metric captures image quality changes in a way that is consistent with how readers visually perceive uniformity, while providing a **more standardized framework for longitudinal QC tracking.**

This study was limited by the small reader-scored image set, which restricted evaluation across severe artifacts and broader image quality variation. Future work should validate the metric on larger, more diverse datasets with additional artifact types, probes, and systems, while also expanding automated analysis to other metrics of ultrasound image quality, such as depth of penetration.

CONCLUSIONS

The hybrid metric showed strong agreement with visual assessment, supporting its potential as an objective ultrasound uniformity QC tool. Its continuous scale may help reduce subjectivity and improve detection of gradual probe or system degradation over time.

REFERENCES

1. Goodsitt MM, Carson PL, Witt S, Hykes DL, Kofler JM Jr. Real-time B-mode ultrasound quality control test procedures: Report of AAPM Ultrasound Task Group No. 1. *Med Phys*. 1998;25(8):1385-1406.
2. American Institute of Ultrasound in Medicine. *Routine Quality Assurance of Clinical Ultrasound Equipment*. Version 2.0. American Institute of Ultrasound in Medicine; 2021.
3. Gallet JM. Embracing ultrasound quality control. *Med Phys Int J*. 2021;9(2):199-204.
4. American College of Radiology. *Quality Control: Ultrasound and Breast Ultrasound*. Revised April 9, 2025.
5. Singh P, Mukundan R, de Ryke R. Texture based quality analysis of simulated synthetic ultrasound images using local binary patterns. *J Imaging*. 2018;4(1):3.
6. Scorza A, Conforto S, Schmid M, Bibbo D, Sciuto SA. A preliminary comparison of two different methods for objective uniformity evaluation in diagnostic ultrasound imaging. In: *XIV Mediterranean Conference on Medical and Biological Engineering and Computing 2016*. IFMBE Proceedings. Vol 57. Springer; 2016:476-481.

Acknowledgements

The authors thank the visual readers, Alexandra Christensen, Sandra Larson, Zheng Feng Lu, Damian Ream, and James Zagzebski, for their time, expertise, and independent assessment of ultrasound uniformity used to assess the performance of the automated metric. The authors would also like to thank Conor Greenberg for his contributions to this research effort.