Comparative Analysis of Automated Frame Selection Methods for Glaucoma Detection in Smartphone-Based Low-Quality Fundus Video

Solomon Gebru Abay[1,2][0000−0003−4590−6538], Melkamu Hunegnaw Asmare[1,3][0000−0003−0990−3348], Julie Jacob[4][0000−0002−4716−6820], and Lucca Geurts[1][0000−0002−9608−9147]

[1] e-Media Research Lab, Department of Computer Science, KU Leuven, 3000 Leuven, Belgium
solomongebru.abay@kuleuven.be
[2] School of Biomedical Engineering, Jimma University, Jimma 378, Ethiopia
[3] Addis Ababa Institute of Technology, Addis Ababa University, Addis Ababa 1176, Ethiopia
[4] Ophthalmology Research Group, Ophthalmology Department, UZ Leuven, 3000 Leuven, Belgium

Abstract. Ultra-low-cost smartphone-based fundus imaging offers a scalable glaucoma screening solution in resource-limited settings, yet deployment is challenged by severe acquisition noise. While short video is preferred over single-image capture in such settings, selecting diagnostically relevant frames from degraded recording remains an open problem, as existing quality assessment methods often fail under these extreme noise conditions of low-cost imaging. This work presents a domain-informed frame selection framework tailored to ultra-low-cost fundus video. The method integrates optic disc localization as a spatial quality gate with complementary spectral and structural quality metrics (retinal saturation, red-channel dominance, vessel density, and sharpness). These metrics are combined via a multiplicative fusion strategy, penalizing frames that fail in any single modality. Evaluation on a multi-center dataset of 284 smartphone-acquired videos using a MobileNetV3-based classifier shows this holistic fusion (mainly S_{total}) significantly stabilizes sensitivity and F1-Score compared to single-metric approaches. Fusion-based frame selection achieves an area under the receiver operating characteristic curve of 0.95 ± 0.04, closely approaching a semi-automated expert baseline (0.98 ± 0.01). These findings demonstrate that robust frame selection bridges the gap between automated and expert intervention, enabling reliable glaucoma screening on ultra-low-cost imaging platforms.

Keywords: Frame selection · Glaucoma screening · Smartphone fundus imaging · Video quality assessment.

1 Introduction

Glaucoma is one of the leading cause of irreversible blindness worldwide, characterized by progressive structural damage to the optic nerve head (ONH) [6,

18]. While early detection via ONH assessment is crucial, mass screening is constrained by the cost and complexity of standard imaging infrastructure [16]. Smartphone-based fundus imaging offers a scalable, low-cost alternative, but its deployment is hindered by the stochastic nature of handheld acquisition. Continuous motion, unstable focus, and limited field-of-view introduce severe artifacts, such as glare, bluriness, and limited retinal visibility, resulting in unreliable single-image capture [13].

To mitigate these failures, short video acquisition is increasingly adopted to capture at least some diagnostically valid content. However, this shifts the burden to automated frame selection, a critical computational bottleneck. In the context of ultra-low-cost adapters (e.g., GSoP [1]), existing quality assessment methods often fail. Classical metrics (e.g., Laplacian variance [4]) are frequently fooled by high-frequency artifacts like eyelashes, while deep learning models trained on clinical-grade data [5, 12, 15] struggle with the extreme domain shift and frequent absence of visible retina. Consequently, selecting valid frames is not simply a preprocessing step but a primary determinant of diagnostic accuracy.

To address this, domain-informed frame selection framework is proposed tailored to the noise characteristics of ultra-low-cost video. Unlike prior work that assumes stable image quality, our method enforces an optic disc (OD)-centric approach. By localizing the OD to gate quality analysis and combining complementary spectral and structural metrics through multiplicative fusion, the framework effectively distinguishes diagnostic regions from noise. Rather than relying on intrinsic quality scores, this approach is evaluated based on its impact on downstream glaucoma classification performance. The main contributions of this work are:

1. Establish automated frame selection as a governing factor of clinical utility in ultra-low-cost glaucoma screening, demonstrating that poor selection strategies directly degrade diagnostic sensitivity.
2. Propose a novel OD localization-guided spectral-spatial fusion framework that robustly identifies diagnostic frames despite severe degradation and intermittent retinal visibility.
3. Provide a systematic benchmark comparing single-metric and fusion-based strategies against a semi-automated baseline, analyzing performance stability and scalability.

2 Materials and Methods

2.1 Dataset and acquisition protocol

A private dataset of smartphone-based fundus videos was collected using the GSoP adapter and a smartphone camera, without the need for additional lens. These data were collected at Jimma University Medical center (Ethiopia) and UZ Leuven (Belgium). Ethical approvals for data collection were obtained from the Institutional Review Board of Jimma University (Ref. JUIH/IRB/283/24

and JUIH/IRB/512/25) and the Ethics Committee Research of UZ/KU Leuven (Study No. S66810). Written informed consent form was signed and collected from all participants prior to data collection. It comprises 284 videos (114 glaucoma, 170 healthy) acquired at two clinical centers using a Redmi Note 11 smartphone. Videos (10 - 17 s) were recorded under mydriasis by both trained staff and minimally trained operators. Glaucoma cases were clinically confirmed, while healthy subjects were validated through comprehensive ophthalmic examination (visual field testing, intraocular pressure, slit-lamp exam) and approved by senior ophthalmologists.

2.2 Preprocessing videos

Each video contains $300 - 510$ frames at 1080×1920 resolution, exhibiting large variability in illumination, focus, and retinal coverage. Since only a small subset of frames contain diagnostic OD views, a two-stage preprocessing pipeline was applied. Videos were first decomposed into frames and centrally cropped to 810×810 pixels to reduce irrelevant background. OD localization was then performed using a pre-trained YOLOv8n detector from [2]. Frames without detected OD (confidence < 0.5) were flagged as non-diagnostic. For evaluating the classification model, detected OD bounding boxes were padded by 25 pixels to preserve the peripapillary context and used to generate OD-centered region of interest (ROI).

2.3 Artifact characterization and frame selection metrics

Ultra-low-cost fundus videos exhibit characteristic artifacts that differ spectrally and structurally from retinal tissue. For example, glare appears as bright white spots, and areas outside the retina lack the expected red-channel dominance. Vascular patterns provide clear structural features, while blurriness causes the edges of these vessels to lose their sharpness. Rather than treating these effects as noise, these are explicitly modeled to distinguish diagnostically useful frames from those with severe artifacts.

Each frame I_t was evaluated using four quality metrics computed on the retinal ROI R_t. If OD localization failed, these metrics were set to zero to discard the frame. To serve as a performance upper bound, a semi-automated baseline was also implemented. Since manually reviewing all 117,185 frames was impractical, two ophthalmologists first reviewed a subset of 694 frames to establish strict visual quality criteria. The first author then used these to select the final frames. To make this manual review manageable, the algorithm pre-filtered the top 20 candidates per video, skipping the massive volume of entirely blank or blurry images. From these 20 candidates, a single frame with the best OD visibility was manually selected.

Single metrics: For each ROI $R_t \subset I_t$, four complementary quality metrics were computed, each designed to capture clinically relevant properties of retinal

visibility while maintaining robustness against common artifacts like glare and motion blur. Such metrics have previously been used for retinal image quality assessment, typically in combination with other approaches [3, 7].

1. Retinal Saturation (SA) quantifies chromatic content using the HSV color space, which penalizes glare and whiteout. A retinal mask M_t was obtained via hue and brightness thresholding, and SA was computed as:

$$\mathrm{SA}_t = \frac{1}{|M_t|} \sum_{(x,y)\in M_t} S_t(x,y), \tag{1}$$

   where $S_t(x,y)$ is the saturation channel and $|M_t|$ is mask's pixel count. This descriptor effectively captures the chromatic content of the tissue [10].
2. Red-Channel Intensity Score (RSC) serves as a proxy for illumination and retinal presence. To quantify retinal tissue presence, a red dominance criterion was applied in RGB space. Pixels satisfying Eq.2 were considered retinal.

$$R_t(x,y) > G_t(x,y) + \delta \quad \textit{and} \quad R_t(x,y) > B_t(x,y) + \delta, \tag{2}$$

   where $\delta = 30$ is a fixed margin, and R_t, G_t, and B_t denote the red, green, & blue channels, respectively. The red-channel intensity score was defined as:

$$\mathrm{RSC}_t = \frac{1}{H \times W} \sum_{(x,y)\in \Omega_t} R_t(x,y), \tag{3}$$

   where Ω_t denotes red dominant pixels, and $H \times W$ is the spatial resolution of the ROI.
3. Vessel Density (VE) captures anatomical content by enhancing vascular structures using a morphological black-hat transform on the green channel [17, 9]. After thresholding the response B_t, VE was computed as:

$$\mathrm{VE}_t = \frac{1}{|M_t|} \sum_{(x,y)\in M_t} \mathbf{1}[B_t(x,y) > \tau], \tag{4}$$

   where $B_t(x,y)$ denotes the black-hat response at pixel (x,y), τ is a fixed threshold, and $\mathbf{1}[\cdot]$ is the indicator function.
4. Sharpness (SHP) measures optical focus via the variance of the Laplacian on the green channel, restricted to the retinal mask [8]. To avoid confounding non-retinal edges, the computation was restricted to the retinal mask on the green channel:

$$\mathrm{SHP}_t = \mathrm{Var}\big(\nabla^2 G_t(x,y) \;\big|\; (x,y) \in M_t\big), \tag{5}$$

   where ∇^2 denotes the Laplacian operator and $G_t(x,y)$ is the green channel intensity.

All metrics were min–max normalized per video to mitigate inter-video exposure variation. VE and SHP signals were further temporally smoothed using a centered moving average filter.

Metrics fusion strategies: In this study, a multiplicative fusion framework based on a conjunctive quality hypothesis is proposed: diagnostically valid frames must simultaneously satisfy spectral and spatial criteria. Three fusion strategies were evaluated:

- S_{vis} (visibility): Assesses visibility and anatomical content, excluding high-frequency focus measures. It tests whether good frames can be selected purely based on illumination and anatomy.

$$S_{vis}(t) = \tilde{SA}_t \cdot \tilde{RSC}_t \cdot \tilde{VE}_t \tag{6}$$

- S_{struct} (structural): Assesses structural clarity and illumination, excluding color saturation. This tests if saturation is redundant given the correlation between red intensity and saturation.

$$S_{struct}(t) = \tilde{SHP}_t \cdot \tilde{RSC}_t \cdot \tilde{VE}_t \tag{7}$$

- S_{total} (holistic): The proposed holistic metric integrating all four quality indicators. This enforces the simultaneous presence of retinal tissue, vascular structure, adequate illumination, and spatial sharpness.

$$S_{total}(t) = \tilde{SA}_t \cdot \tilde{RSC}_t \cdot \tilde{VE}_t \cdot \tilde{SHP}_t \tag{8}$$

Clinical impact of frame selection with these approaches was assessed via downstream glaucoma classification using MobileNetV3-Small[11], selected for its high efficiency in edge deployment scenarios. The model was fine-tuned on 224×224 inputs for 30 epochs (Adam, $lr = 10^{-5}$), with class imbalance. Data augmentation included with horizontal and vertical flipping. Evaluation followed 5-fold stratified group cross-validation with subject-level split. Hyperparameters (dropout=0.2, batch size=8, patience=10) were tuned based on validation area under the receiver operating characteristic curve (AUC). Performance was reported using AUC, accuracy, F1-score, sensitivity, and specificity, with 95% confidence interval (CI).

3 Experimental results

3.1 Qualitative assessment

The impact of the ROI strategy on frame quality assessment is illustrated in Fig. 1. The baseline approach (middle), based on a fixed center crop, analyzes both the OD region and irrelevant peripheral content such as eyelids or sclera. In contrast, OD localization (right) constrains analysis to the anatomical ROI, acting as a spatial filter that prevents quality metrics from being inflated by high-contrast in non-diagnostic structures.

Quantitatively, OD localization served as an effective primary automated quality gate. Across the 284 videos, approximately 10% of frames were rejected due to failed OD detection. A subsequent inspection confirmed that all rejected

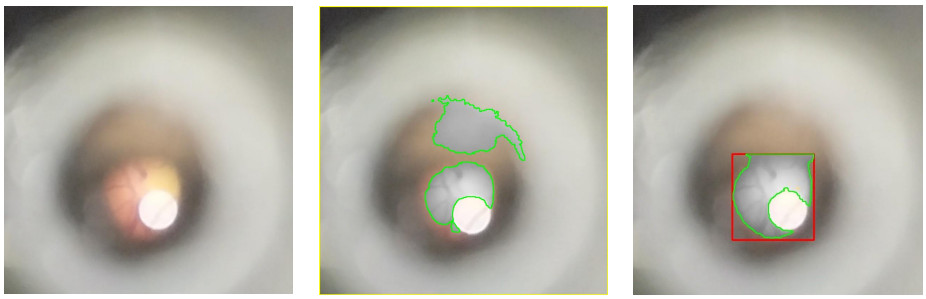

Fig. 1. Impact of ROI strategy. (Left) Original cropped frame. (Middle) Evaluation without OD localization includes irrelevant peripheral noise (gray). (Right) OD localization restricts analysis to the diagnostic region, improving metric reliability.

frames were non-diagnostic, typically affected by severe blur, poor illumination, or complete loss of OD visibility due to blinking or misalignment.

Fig. 2 illustrates the temporal behavior of the quality metrics for a representative video. Single metrics such as SA exhibited substantial spikes during transient artifacts (e.g., a smooth reddish at frame number-219), leading to incorrect frame selection. In contrast, the fusion score S_{total} suppressed these artifacts by jointly enforcing spectral and structural consistency, correctly identifying a stable diagnostic window around frame number-241. Single-metric approaches frequently selected frames dominated by glare or eyelashes, whereas fusion-based strategies consistently favored frames with visible vasculature and stable color characteristics.

3.2 Metric validation

To assess complementarity among quality indicators, Pearson correlation coefficients were computed between all pairs of base metrics on the entire dataset. Fig. 3 shows the resulting correlation matrix across the dataset. Spectral and spatial metrics exhibited low mutual dependence, indicating that they capture distinct aspects of the image. In particular, red intensity and SHP were nearly uncorrelated ($r = -0.11$), suggesting that illumination alone is a poor predictor of optical focus.

Moderate correlation was observed between SA and red intensity ($r = 0.43$), reflecting shared dependence on retinal illumination, and between VE and SHP ($r = 0.48$), consistent with the requirement of focus for resolving fine vascular structures. The overall inter-metric correlation ($|r| < 0.6$ for all pairs) [14] confirms that each feature captures a non-redundant aspect of image quality, supporting the integration of these methods via multiplicative fusion.

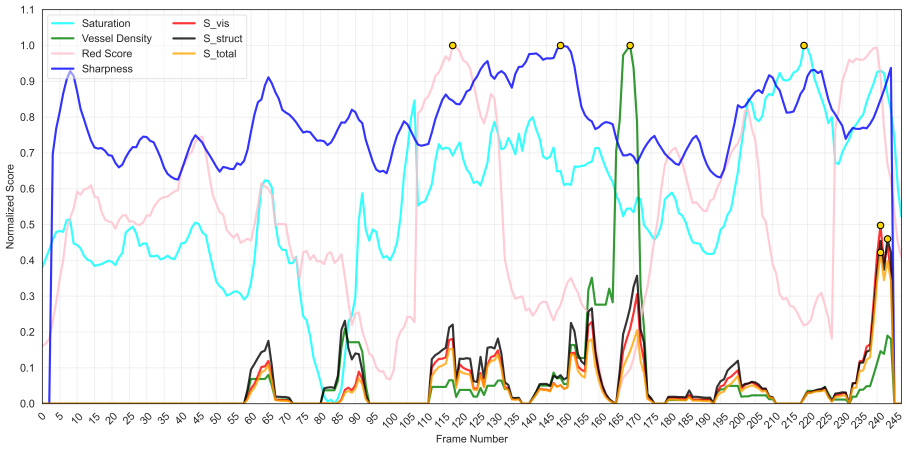

Fig. 2. Temporal evolution of quality metrics for a representative video. The fusion metric such as S_{total} (orange) penalizes artifact-driven spikes in single metrics (e.g., saturation, cyan), enabling robust frame selection.

3.3 Glaucoma classification performance

A lightweight MobileNetV3-Small-based model was used to evaluate the clinical impact of frame selection, trained on frames selected by each strategy under 5-fold cross-validation. Table 1 details the comparative performance and statistical significance across these selection strategies. Single-metric approaches, particularly SHP, achieved reasonable mean specificity but exhibited instability in sensitivity and F1-Score, with 95% CI spanning from 0.52 to 1.0. This high variability confirms a significant susceptibility to high-frequency noise and inconsistent frame selection when using individual metrics. In contrast, the holistic fusion approach, specifically S_{total}, significantly stabilized performance, outperforming other automated strategies in mean AUC, Accuracy, and F1-Score.

Moreover, paired t-tests revealed that S_{total} achieved performance levels that were statistically indistinguishable from the semi-manual approach across all

Table 1. Performance comparison (Mean ± 95% CI) across frame selection methods using a single frame per video. Values indicate stability across 5-fold cross-validation for S_{total} and the semi-manual.

Metric	RSC	SA	SHP	VE	S_{total}	Semi-Manual
AUC	0.94±0.03	0.93±0.04*	0.92±0.06*	0.93±0.04*	0.95±0.04	**0.98±0.01**
Accuracy	0.87±0.02*	0.85±0.06*	0.82±0.09*	0.84±0.05*	0.89±0.07	**0.93±0.02**
Sensitivity	0.84±0.08*	0.77±0.10*	0.78±0.24	0.81±0.02*	0.87±0.09	**0.94±0.03**
Specificity	0.88±0.09	0.90±0.09	0.83±0.12	0.86±0.08*	0.90±0.09	**0.93±0.04**
F1-Score	0.84±0.02*	0.81±0.09*	0.76±0.18	0.81±0.04*	0.86±0.09	**0.92±0.03**

* indicates statistically significant difference ($p < 0.05$) compared to the semi-manual approach.

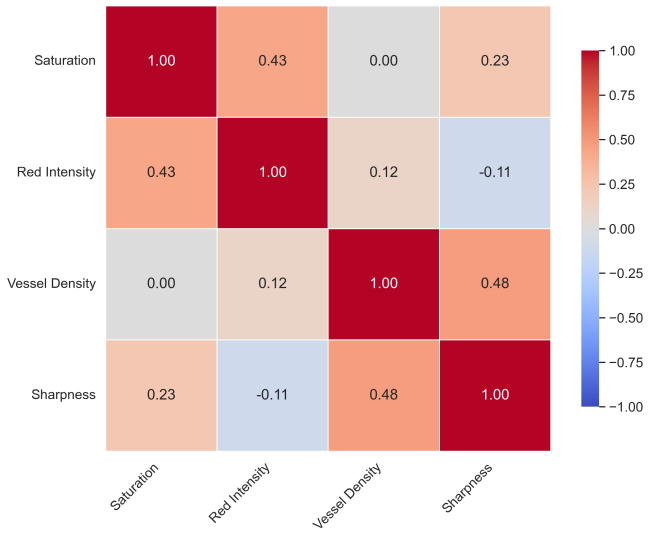

Fig. 3. The correlation heatmap of the four base quality metrics, demonstrating low redundancy between spectral and spatial features.

reported metrics, including AUC ($p = 0.1795$), accuracy ($p = 0.1594$), and F1-Score ($p = 0.1301$). While the semi-manual approach retained a numerical advantage in mean values, the lack of significant difference confirms that S_{total} effectively bridges the gap between automated selection and expert intervention.

4 Conclusion

This work showed automated frame selection is a primary determinant of screening performance in ultra-low-cost smartphone-based fundus videos. Unlike traditional clinical imaging, this modality is dominated by severe artifacts. Consequently, reliably extracting diagnostic frames is a critical methodological challenge for affordable screening platforms, rather than a simple preprocessing step.

Analyses of individual quality metrics revealed low inter-metric correlations ($|r| < 0.6$), confirming they capture complementary properties with minimal redundancy. This is further enhanced by integrating OD localization as an automated quality gate. By rejecting approximately 10% of frames lacking detectable OD content before metric evaluation, the system prevents quality scores from being artificially inflated by peripheral noise.

Downstream classification confirmed that individual quality metrics are insufficiently stable for clinical deployment. For example, SHP-based selection frequently responded to high-frequency artifacts rather than meaningful structures (Table 1). In contrast, the proposed multiplicative fusion strategy (S_{total}) mitigated these failures by enforcing a conjunctive quality criterion. While single-metric methods showed significant performance gaps relative to expert selection,

S_{total} achieved performance statistically indistinguishable from the semi-manual baseline across all clinical metrics ($p > 0.05$).

Finally, this approach aligns well with deployment constraints in low-resource settings. The use of lightweight image processing operations and a compact object detector enables efficient on-device execution, reducing operator burden and supporting scalable screening workflows on widely available hardware. While these results are promising, validating this framework on a larger, external dataset remains a primary direction for future work. Additionally, investigating the selection and aggregation of multiple frames per video, rather than a single representative frame, may further improve robustness and determine whether automated methods can consistently match expert manual selection in complex clinical settings.

Acknowledgments. This study was funded by the 5E-Group T fund (3E220426). The authors gratefully acknowledge Marie Adriaenssens and Ralph Van Ginkel for their assistance with data collection at UZ Leuven.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Abay, S.G., Asmare, M.H., Geurts, L.: An affordable smartphone-based fundus imaging device for retinal examination. In: 2025 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). pp. 1–6 (2025). https://doi.org/10.1109/EMBC58623.2025.11254224
2. Abay, S.G., Geurts, L.: Automated optic disc localization from smartphone-captured low quality fundus images using yolov8n-based model. In: 2024 IEEE International Symposium on Medical Measurements and Applications (MeMeA). pp. 1–6 (2024). https://doi.org/10.1109/MeMeA60663.2024.10596783
3. Abdel-Hamid, L., El-Rafei, A., El-Ramly, S., Michelson, G., Hornegger, J.: Retinal image quality assessment based on image clarity and content. Journal of biomedical optics **21**(9), 096007–096007 (2016)
4. Bansal, R., Raj, G., Choudhury, T.: Blur image detection using laplacian operator and open-cv. 2016 International Conference System Modeling & Advancement in Research Trends (SMART) pp. 63–67 (2016)
5. Coyner, A., Swan, R., Campbell, J., Ostmo, S., Brown, J., Kalpathy-Cramer, J., Kim, S., Jonas, K., Chan, R., Chiang, M., Sönmez, M., Chan, R., Jonas, K., Horowitz, J., Coki, O., Eccles, C.A., Sarna, L., Orlin, A., Berrocal, A., Montero, C.: Automated fundus image quality assessment in retinopathy of prematurity using deep convolutional neural networks. Ophthalmology Retina **3** (05 2019). https://doi.org/10.1016/j.oret.2019.01.015
6. Davuluru, S.S., Jess, A.T., Kim, J.S.B., Yoo, K., Nguyen, V., Xu, B.Y.: Identifying, understanding, and addressing disparities in glaucoma care in the united states. Translational Vision Science & Technology **12**(10), 18–18 (2023)
7. Fasih, M., Langlois, J.P., Tahar, H.B., Cheriet, F.: Retinal image quality assessment using generic features. In: Medical Imaging 2014: Computer-Aided Diagnosis. vol. 9035, pp. 772–778. SPIE (2014)

8. Ferzli, R., Karam, L.J.: A no-reference objective image sharpness metric based on the notion of just noticeable blur (jnb). IEEE Transactions on Image Processing **18**(4), 717–728 (2009). https://doi.org/10.1109/TIP.2008.2011760
9. Fraz, M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A., Owen, C., Barman, S.: Blood vessel segmentation methodologies in retinal images – a survey. Computer Methods and Programs in Biomedicine **108**(1), 407–433 (2012). https://doi.org/https://doi.org/10.1016/j.cmpb.2012.03.009
10. Gonzalez, R.C., Woods, R.E.: Digital Image Processing. Pearson Education Limited, rth edn. (2009)
11. Howard, A., Sandler, M., Chu, G., Chen, L.C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., et al.: Searching for mobilenetv3. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 1314–1324 (2019)
12. Karlsson, R.A., Jonsson, B.A., Hardarson, S.H., Olafsdottir, O.B., Halldorsson, G.H., Stefansson, E.: Automatic fundus image quality assessment on a continuous scale. Computers in Biology and Medicine **129**, 104114 (2021). https://doi.org/https://doi.org/10.1016/j.compbiomed.2020.104114
13. Mittal, A., Sanchez, V., Azad, N.S., Zuyev, Y., Robles, R., Sherwood, M.: The utility of a smartphone-based retinal imaging device as a screening tool in an outpatient clinic setting: Protocol for an observational study. JMIR Res Protoc **14**, e52650 (Mar 2025). https://doi.org/10.2196/52650
14. Mukaka, M.M.: A guide to appropriate use of correlation coefficient in medical research. Malawi Medical Journal **24**(3), 69–71 (2012)
15. Scarpa, F., Berto, A., Tsiknakis, N., Manikis, G., Fotiadis, D.I., Marias, K., Scarpa, A.: Automated analysis for glaucoma screening of retinal videos acquired with smartphone-based ophthalmoscope. Heliyon **10**(14), e34308 (2024). https://doi.org/https://doi.org/10.1016/j.heliyon.2024.e34308
16. Stein, J.D., Khawaja, A.P., Weizer, J.S.: Glaucoma in adults—screening, diagnosis, and management: A review. JAMA **325**(2), 164–174 (01 2021). https://doi.org/10.1001/jama.2020.21899
17. Tamim, N., Elshrkawey, M., Abdel Azim, G., Nassar, H.: Retinal blood vessel segmentation using hybrid features and multi-layer perceptron neural networks. Symmetry **12**(6) (2020). https://doi.org/10.3390/sym12060894
18. Tham, Y.C., Li, X., Wong, T.Y., Quigley, H.A., Aung, T., Cheng, C.Y.: Global prevalence of glaucoma and projections of glaucoma burden through 2040: A systematic review and meta-analysis. Ophthalmology **121**(11), 2081–2090 (2014). https://doi.org/https://doi.org/10.1016/j.ophtha.2014.05.013