

GlauCUTU-WebXR: Virtual Reality Visual Field Perimetry System Embedded with Enhanced Humphrey Field Analyzer Prediction Models for Glaucoma Screening in Resource-limited Settings

P. Boonyarattanasoontorn, P. Tieworn, S. Puangarom, S. Sangchocanonta, P. Phienphanich, P. Kunumpol, A. Munthuli, R. Itthipanichpong, K. Ratanawongphaibul, S. Chansangpetch, A. Manassakorn, V. Tantisevi, P. Rojanapongpun, and C. Tantibundhit*

Abstract—This study presents GlauCUTU-WebXR, an advanced glaucoma screening system leveraging commercial virtual reality (VR) devices with WebXR for broader accessibility. Transitioning from custom 3D-printed hardware, GlauCUTU-WebXR integrates an improved prediction model that enhances sensitivity predictions using positional features (position index: PID) and severity classification metadata (SeverityScore). These additions help the model capture positional and severity-specific patterns, which, in turn, significantly improve prediction accuracy in converting GlauCUTU sensitivity to Humphrey Field Analyzer (HFA) sensitivity. The model trained and validated on 97 eyes (47 patients) and tested on 24 eyes (12 patients) with varying glaucoma severity, it achieved a mean absolute error (MAE) of 4.3398, root mean square error (RMSE) of 6.6614, and a Pearson correlation of 0.8033. The system significantly reduced testing time compared to HFA ($p < 0.001$). GlauCUTU-WebXR shows no significant differences in results compared to HFA across severity levels: none ($p = 0.8750$), mild ($p = 0.5888$), moderate ($p = 0.9153$), and severe ($p = 0.9958$). By utilizing commercial VR hardware and enhanced prediction models, GlauCUTU-WebXR provides a cost-effective, efficient alternative for glaucoma screening, particularly in resource-limited settings, demonstrating its potential to improve accessibility and early detection efforts.

Keyword—glaucoma, visual field testing, virtual reality, WebXR, machine learning

I. INTRODUCTION

Glaucoma, a chronic and progressive ocular disorder characterized by optic neuropathy and visual field loss, is the leading cause of irreversible blindness worldwide. In 2020, approximately 64.3 million people were affected, a figure projected to rise to 111.8 million by 2040 [1]. Despite its incurable nature, early detection and regular ophthalmologic care are essential to preserving vision, as glaucoma typically remains asymptomatic until advanced stages [2]. In Thailand, glaucoma is the second leading cause of blindness after cataracts, with a prevalence of 2.5% to 3.8%. Alarming, the incidence of glaucoma in Thailand is expected to rise threefold for males and fourfold for

*C. Tantibundhit (E-mail: tchartur@engr.tu.ac.th), P. Boonyarattanasoontorn, S. Puangarom, S. Sangchocanonta, P. Phienphanich, and A. Munthuli are with Center of Excellence in Nexus for Advanced Intelligence in Law, Engineering, and Medicine (Nail'Em), Faculty of Engineering, Thammasat School of Engineering, Thammasat University, Thailand. P. Tieworn is with Nail'Em and Choate Rosemary Hall, USA. R. Itthipanichpong, K. Ratanawongphaibul, S. Chansangpetch, A. Manassakorn, V. Tantisevi, and P. Rojanapongpun are with Center of Excellence in Glaucoma, Department of Ophthalmology, Chulalongkorn University, King Chulalongkorn Memorial Hospital, Thai Red Cross Society, Thailand.

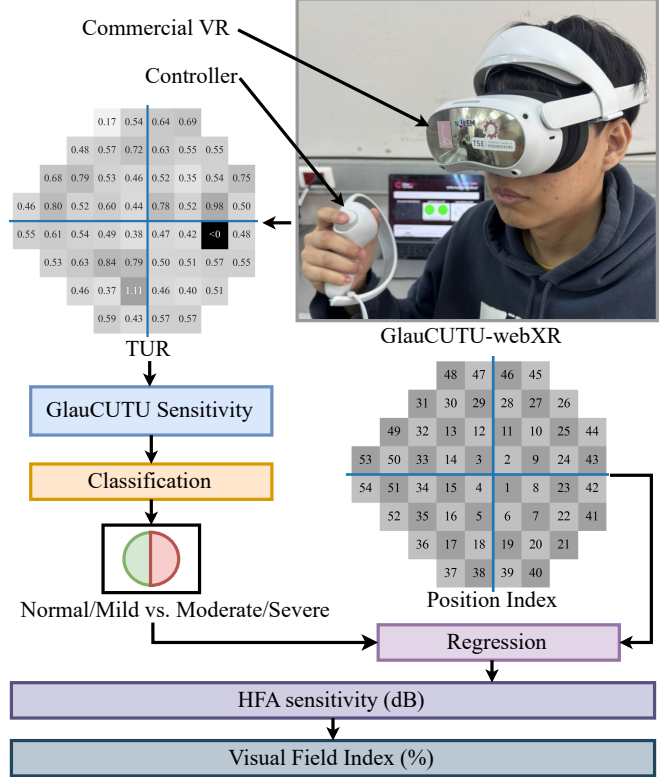


Fig. 1. GlauCUTU-WebXR an enhanced virtual reality visual field perimetry system for HFA sensitivity prediction.

females over the next 50 years [3]. While Thailand meets the World Health Organization's recommended ratio of one ophthalmologist per 100,000 people nationally, rural areas face significant disparities, with only 26% of provinces and 14% of districts meeting this standard [4]. This urban-rural gap, coupled with limited access to essential screening equipment, exacerbates preventable cases of blindness in underserved populations. Furthermore, the economic burden of glaucoma is substantial, with treatment costs in the US exceeding \$1 billion annually, particularly for patients with advanced disease stages [3]. Addressing these disparities and ensuring equitable access to care is critical to mitigating the growing impact of glaucoma globally and within Thailand.

Visual field (VF) testing plays a vital role in diagnosing and monitoring glaucoma progression, with the Humphrey Field Analyzer (HFA) being the most widely used device [5], [6]. Despite its widespread use, the HFA's limitations,

which include immobility, high cost, and the need for patients to maintain a fixed head position restrict its accessibility in low-resource settings [6], [7]. To overcome these limitations, virtual reality (VR) based perimetry systems have emerged as a promising alternative, providing portability, improved patient comfort, and potential for home-based testing [6], [8]. The development of reliable VR perimetry devices holds significant promise for improving glaucoma screening and management in underprivileged areas [7].

Virtual reality (VR) technologies are increasingly explored for their capabilities in visual field assessment. Early work by Tsapakis *et al.* demonstrated an implementation of VR by using Trust EXOS 3D VR glasses with a smartphone, achieving a high correlation coefficient ($r=0.8080$, $p<0.0001$) with HFA [9]. However, the small sample size of 10 patients highlights the need for further validation to confirm its efficacy. Building upon this foundation, Razeghinejad *et al.* evaluated the VisuALL head-mounted perimeter, which demonstrated a strong correlation with HFA for mild and moderate glaucoma. Nonetheless, its reduced accuracy in severe cases and the longer testing durations compared to traditional HFA testing illustrated the challenges in optimizing VR-based approaches for glaucoma screening [10].

Recent advances focus on improving HFA sensitivity prediction accuracy and testing efficiency. For example, the NOVA trial by Bradley *et al.* compared estimated sensitivities between their RATA-Standard algorithm and SITA-Standard, finding no significant differences for right eye measurements, but noting discrepancies in left eye results and reduced accuracy near its 15 dB measurement floor [11]. Additionally, our previous work, GlauCUTU introduced a novel Time-Until-Perceived (TUP) approach and achieved a significant reduction in test time compared to HFA (290 seconds vs. 600 seconds) while show no significant differences with HFA Visual Field Index (VFI) predictions ($p=0.996$ for HFA vs. GlauCUTU-ML) with high-reliability agreement with clinicians' interpretations of HFA results [8].

Despite these advances, existing VR perimetry systems still face challenges in balancing test duration with prediction accuracy. GlauCUTU demonstrated the potential for rapid testing using TUP measurements. Nevertheless, it showed limitations in sensitivity prediction accuracy, particularly for moderate and severe glaucoma cases [8]. Additionally, like other systems, GlauCUTU relied on custom hardware solutions, in turn, limiting scalability and accessibility in low-resource settings [8], [11]. These challenges motivated the development of an enhanced system, GlauCUTU-WebXR. This enhanced system sought to transform GlauCUTU into a more accessible and affordable VR perimetry system while maintaining accurate HFA sensitivity estimation. It maintains GlauCUTU's efficient test duration while also enhancing both its HFA sensitivity prediction accuracy and scalability. The key contributions of our work are as follows:

- 1) Transitioning to a commercial-grade VR headset for GlauCUTU-WebXR that improves affordability and scalability and facilitates broader adoption for glaucoma screening in low-resource settings.

- 2) Creating an enhanced algorithm in GlauCUTU-WebXR that predicts HFA sensitivity with higher accuracy than the original GlauCUTU version [8], as demonstrated by the improved performance metrics.
- 3) The statistical results show no significant differences in VFI prediction between GlauCUTU-WebXR and HFA across all glaucoma severity levels, indicating that GlauCUTU-WebXR is a viable alternative to HFA for accessible glaucoma screening.

The remainder of this paper is organized as follows: Section II details the data sets used in this study. Section III describes the experimental design and setup, while Section IV presents the results. Section V provides the discussion, conclusion, and future directions.

II. DATA SET

A. Study group

The study recruited outpatients from King Chulalongkorn Memorial Hospital's ophthalmology clinic. Participants were selected based on their medical records, provided with comprehensive information about the study, and asked to sign an informed consent statement. The study protocol was approved by the Chulalongkorn Institutional Review Board (No. 715/61). A total of 70 patients were enrolled, ranging in age from 38 to 89 years (average age with standard deviation of 64.43 ± 9.65 years). In total, 140 eyes were examined and categorized by glaucoma severity: 46 eyes as normal, 40 as mild, 21 as moderate, and 33 as severe.

The participants were divided into two groups: a control group and a glaucoma study group. Inclusion criteria for both groups were as follows: age ≥ 18 years, fixation loss, false-positive, and false-negative rates below 30% on Standard Automated Perimetry (SAP), visual acuity (VA) of 20/70 or better, and no use of alcohol or nervous system medications within the past 14 days [8].

Participants in the glaucoma group required a confirmed diagnosis of primary glaucoma based on Hodapp-Parrish-Anderson (HPA) criteria and SAP-detected visual field defects within the past 3 months [12]. Control group participants needed normal intraocular pressure and no glaucomatous optic neuropathy per SAP results and HPA criteria. Exclusion criteria for both groups included non-glaucomatous visual field conditions and medications affecting the nervous system.

B. Data preparation

The data cleaning was performed to exclude the data that did not pass the requirements or was defective. A total of 11 subjects were dropped due to various issues including standard testing protocols, medication effects, fixation issues, excessive focus on central stimuli, inability to maintain central fixation, and high false positive rates. These data cleaning and subject drops were discussed and unanimously approved by all the expert ophthalmologists in our team.

After data cleansing, the dataset was stratified by glaucoma severity and split 80:20 for training and testing. Five-fold cross-validation was applied to the training set, ensuring

TABLE I
COMPARISON OF MAE, RMSE, AND PEARSON'S R ACROSS DIFFERENT MODELS ON TEST SET.

Metrics	Models								
	LWL	AddReg	Voter	LWL+PID	AddReg+PID	Voter+PID	LWL+PID +SeverityScore	AddReg+PID +SeverityScore	Voter+PID +SeverityScore
MAE	5.0446	5.0921	5.0731	4.8078	4.8080	4.8022	4.3377	4.3930	4.3398
RMSE	7.2118	7.2279	7.2173	6.8419	6.7696	6.7636	6.7123	6.6762	6.6614
Pearson's r	0.7552	0.7545	0.7555	0.7835	0.7920	0.7906	0.8007	0.8035	0.8033

no data leakage by keeping eyes from the same patient within the same fold. The final dataset included 59 patients (118 eyes): 94 eyes (47 patients) for training and validation (27 normal, 27 mild, 17 moderate, and 23 severe) and 24 eyes (12 patients) for testing (7 normal, 7 mild, 4 moderate, and 6 severe). Additionally, the GlauCUTU dataset used in our previous study [8], with 31 patients (62 eyes: 46 normal, 7 mild, 3 moderate, and 6 severe), was used to enhance training and improve model performance.

III. EXPERIMENTAL DESIGN & SETUP

A. Transition to commercial-grade VR

In our previous work [8], the GlauCUTU perimetry system was developed, and the TUR (Time Until Response) of the patient was collected using the in-house 3D-printed VR headset. Then, Time Until Perceive (TUP) was derived from TUR and the Best Possible Response Time (BPRT) using Equation (1) of [8]. Additionally, the GlauCUTU sensitivity was calculated using Equation (2) of [8]. Machine learning and deep learning models were then applied to map GlauCUTU sensitivity to HFA sensitivity, with the best regression model, Locally Weighted Learning (LWL), achieving an MAE of 3.50 ± 1.66 , RMSE of 4.92 ± 2.18 , and Pearson correlation coefficient (Pearson's r) of 0.69 ± 0.18 based on leave one out cross validation. Furthermore, the GlauCUTU-ML showed no significant difference in VFI compared to HFA across all glaucoma severity levels: Normal ($p=0.999$), Mild ($p=0.999$), Moderate ($p=0.997$), and Severe ($p=0.573$).

In addition of what stated earlier, although the custom 3D-printed VR headset validated the system's feasibility, it faced significant limitations in scalability and production costs, rendering it impractical for widespread use in rural and low-resource settings. To overcome these limitations, we transitioned to a commercial VR headset. Commercial devices offer consistent quality, reduced production costs, and improved durability. The system was modernized using WebXR [13], a web standard that enables cross-platform VR/AR applications through web browsers. This implementation allows GlauCUTU-WebXR to run on web browsers with compatible VR devices. Moreover, the web-driven architecture enables an easy software update pipeline and improves accessibility for data collection and clinical use.

B. Regression model with severity score

The GlauCUTU-WebXR software, coupled with a commercial VR headset, was utilized to collect data from the study group, adhering to the methodology established in our previous work [8]. In this work, we extended the HFA sensitivity prediction model by adding two new features:

positional features (position index: PID) and metadata from a classification model (SeverityScore). These additions were motivated by research showing that glaucomatous visual field defects demonstrate patterns of damage and progression [14], [15].

First, each of the test points was labeled according to the methodology illustrated in Fig. 8 of [8]. Positional features were introduced to pinpoint specific locations on the 24-2 test grid where sensitivity predictions were focused. This spatial information was crucial, as glaucomatous visual field defects exhibit distinct regional patterns [14]. By incorporating these features, the model could account for location-specific sensitivity losses more effectively.

Second, classification metadata were added from a confidence score generated by a classification model. This model was trained to differentiate between normal/mild and moderate/severe glaucoma using GlauCUTU sensitivity. This classification boundary was selected based on visual field progression patterns. As studies show, damage typically evolves from limited defects in mild cases to more extensive involvement in moderate and severe stages [14], [15]. Using this classification, the confidence score provides additional high-level information to the regression model for predicting HFA sensitivity. Together with neighboring point features, 11 features were used to predict HFA sensitivity at specific points on the visual field test grid. Figure. 1 shows the proposed HFA sensitivity prediction model.

IV. EXPERIMENTAL RESULTS

Mean absolute error (MAE), Root mean square error (RMSE), and Pearson's r were used to evaluate the model's performance. The best model was selected after performing 5-fold cross validation on the validation set, and the final performance was evaluated and compared on the test set as detailed in Table I.

The models from previous work showed the following results: the LWL model achieved the best overall performance with an MAE of 5.0446, an RMSE of 7.2118, and a Pearson's r of 0.7552. The Additive Regression (AddReg) model had the highest MAE of 5.0921 and RMSE of 7.2279, with a Pearson's r of 0.7545. The Voter model demonstrated better performance, with an MAE of 5.0731, an RMSE of 7.2173, and a Pearson's r of 0.7555, slightly surpassing LWL. Among the baseline models, the LWL model demonstrated the best performance.

Introducing the PID feature improved performance across all models. The LWL+PID model achieved an MAE of 4.8078, RMSE of 6.8419, and Pearson's r of 0.7835. The AddReg+PID model had an MAE of 4.8080, RMSE of

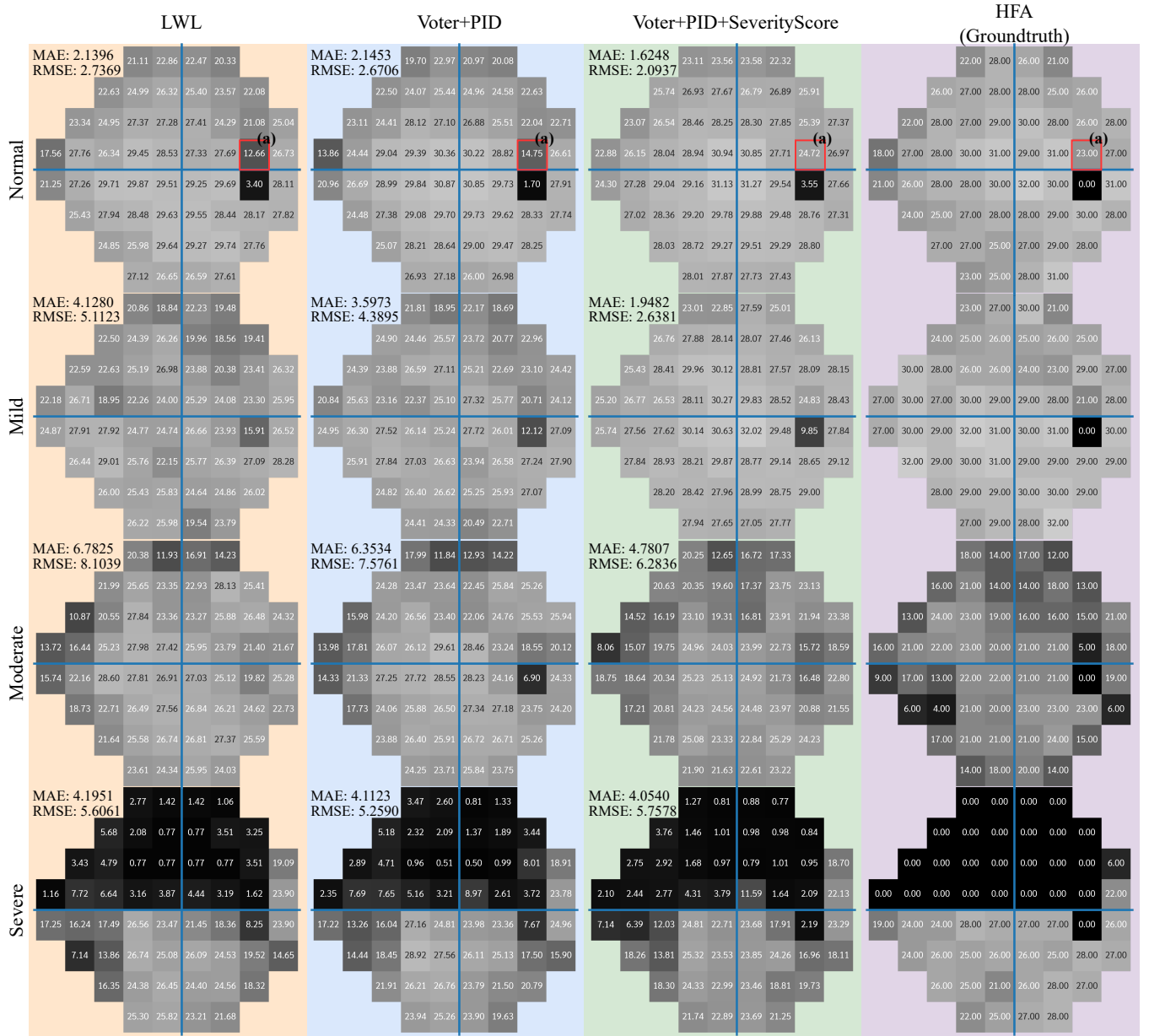


Fig. 2. HFA sensitivity comparison across different severities and models. Each row (from top to bottom) represents a severity category (Normal, Mild, Moderate, and Severe), while each column (from left to right) corresponds to a model (LWL, Voter+PID, Voter+PID+SeverityScore, and Ground Truth). The heatmaps display VF sensitivities at each test point, with darker shades indicating lower HFA sensitivities. The MAE and RMSE values in the top-left corner of each prediction show the average error compared to the ground truth. Blue lines highlight the horizontal and vertical midlines of the visual field.

6.7696, and Pearson's r of 0.7920. The Voter+PID model achieved the lowest MAE and RMSE, at 4.8022 and 6.7636, respectively, along with a Pearson's r of 0.7906. Given its enhanced performance, the Voter+PID model emerged as the best overall performer in this category.

Introducing both PID and SeverityScore features further enhanced the performance of the models. The LWL+PID+SeverityScore model achieved the lowest MAE of 4.3377, RMSE of 6.7123, and Pearson's r of 0.8007. The AdReg+PID+SeverityScore model improved with an MAE of 4.3930, RMSE of 6.6762, and Pearson's r of 0.8035. The Voter+PID+SeverityScore model achieving an MAE of 4.3398, RMSE of 6.6614, and Pearson's r of 0.8033, overall, making it the best performer for the category with both PID and SeverityScore features.

When evaluating eyes with varying glaucoma severity, the improved model exhibited better alignment with ground truth visual field patterns as depicted by Fig. 2. For the normal eye category (first row), all models nicely predicted the blind spot, but only Voter+PID+SeverityScore correctly predicted the VF region above the blind spot (area (a)). For the mild eye category (second row), none of the models effectively predicted the VF at the blind spot. However, Voter+PID+SeverityScore estimated the lowest sensitivity at 9.85 dB while achieving the most accurate pattern with an MAE of 1.9482 and RMSE of 2.6381. In the moderate severity category (third row), while the LWL + PID model captured the blind spot precisely, the Voter+PID+SeverityScore generated a VF pattern closer to the ground truth, with an MAE of 4.7807 and RMSE of 6.2836. For the severe

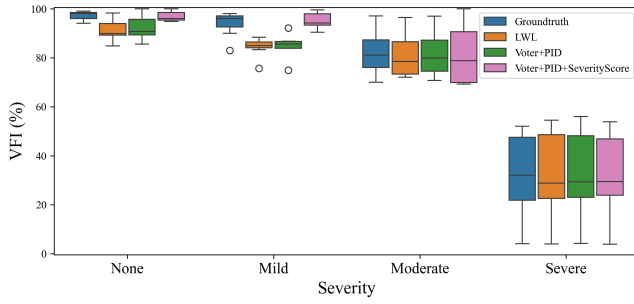


Fig. 3. Comparison of box plots of VFI calculated from predicted HFA sensitivity of models: LWL, Voter+PID, and Voter+PID+SeverityScore, across different levels of glaucoma severity (None, Mild, Moderate, and Severe). The predictions are compared to the ground truth from the HFA report.

category (fourth row), all models produced similar VF predictions. However, the Voter+PID+SeverityScore model delivered the closest overall sensitivity to the ground truth with an MAE of 4.0540 and displayed the darkest shading of VF, demonstrating superior overall performance compared to other models.

The predicted HFA sensitivity from each model was used to calculate the VFI and the box plots of the results are shown in Fig. 3, along with the ground truth. A two way ANOVA analysis was conducted to evaluate the effects of disease severity and model type on prediction performance. The analysis revealed no significant difference for severity [$F(3, 60)=2.3854, p=0.0780$], no significant difference for model type [$F(2, 60)=2.0976, p=0.1317$], and no significant interaction between severity and model [$F(6, 60)=0.9671, p=0.4553$]. While the ANOVA results indicated no significant differences between groups, they only provided information about overall differences between means.

To evaluate the clinical validity of our models, we conducted paired t -tests to compare the predicted VFI against the ground truth across different severity levels. The results for the Voter+PID+SeverityScore model are summarized in Table II. The analysis revealed no significant differences across all severity levels for this model.

For the overall comparison, the Voter+PID+SeverityScore model achieved a mean of differences ($\bar{d}$) of -0.2663 with a standard deviation of differences (S_D) of 5.9022 , which was not statistically significant difference compared to the ground truth [$t(23)=-0.2210, p=0.8270$]. In contrast, both the LWL model and Voter+PID model showed significant differences from the ground truth ($\bar{d} \pm S_D = 4.5907 \pm 8.1858$, [$t(23)=2.7474, p=0.0115$] and $\bar{d} \pm S_D = 3.9649 \pm 8.6802$, [$t(23)=2.2377, p=0.0352$], respectively).

When examining specific severity levels, the Voter+PID+SeverityScore model demonstrated minimal deviation for normal eyes, with an $\bar{d} \pm S_D$ of 0.2481 ± 3.9993 [$t(6)=0.1641, p=0.8750$]. In contrast, the LWL model and Voter+PID model showed significant deviations from the ground truth ($\bar{d} \pm S_D = 5.8422 \pm 5.1332$, [$t(6)=3.0075, p=0.0238$], and $\bar{d} \pm S_D = 4.8339 \pm 5.9711$, [$t(6)=2.1418, p=0.0760$], respectively).

For mild cases, the Voter+PID+SeverityScore model

TABLE II
PAIRED t -TEST ANALYSIS COMPARING PREDICTED AND GROUND TRUTH VFI SCORES OF VOTER+PID +SEVERITYSCORE MODEL ACROSS DIFFERENT SEVERITY LEVELS. RESULTS SHOW MEAN OF DIFFERENCES ($\bar{d}$), STANDARD DEVIATION OF DIFFERENCES (S_D), t -STATISTICS, AND p -VALUES.

Model	Severity	$\bar{d}$	S_D	t -statistic	p -value
Voter+PID +SeverityScore	Overall	-0.2663	5.9022	$t(23)=-0.2210$	0.8270
	None	0.2481	3.9993	$t(6)=0.1641$	0.8750
	Mild	-1.4929	6.9195	$t(6)=-0.5708$	0.5888
	Moderate	0.5949	10.2914	$t(3)=0.1156$	0.9153
	Severe	-0.0092	4.1108	$t(5)=-0.0055$	0.9958

had an $\bar{d} \pm S_D$ of -1.4929 ± 6.9195 [$t(6)=-0.5708, p=0.5888$], which was not statistically significant difference. However, both the LWL model and Voter+PID model exhibited significant deviations ($\bar{d} \pm S_D = 9.4385 \pm 7.0668$, [$t(6)=3.5331, p=0.0123$], and $\bar{d} \pm S_D = 8.9408 \pm 8.0940$, [$t(6)=2.9193, p=0.0267$], respectively).

For moderate and severe cases, none of the models showed significant differences from the ground truth. In moderate cases, the LWL model achieved an $\bar{d} \pm S_D$ of 0.9227 ± 15.1108 [$t(3)=0.1221, p=0.9105$], the Voter+PID model achieved an $\bar{d} \pm S_D$ of 0.4151 ± 15.6960 [$t(3)=0.0529, p=0.9611$], and the Voter+PID+SeverityScore model achieved an $\bar{d} \pm S_D$ of 0.5949 ± 10.2914 [$t(3)=0.1156, p=0.9153$]. In severe cases, the LWL model had an $\bar{d} \pm S_D$ of -0.0796 ± 2.7486 [$t(3)=-0.0710, p=0.9462$], the Voter+PID model had an $\bar{d} \pm S_D$ of -0.4875 ± 2.7586 [$t(3)=-0.4328, p=0.6832$], and the Voter+PID+SeverityScore model had an $\bar{d} \pm S_D$ of -0.0092 ± 4.1108 [$t(5)=-0.0055, p=0.9958$].

The total testing time analysis revealed significantly shorter duration when comparing GlauCUTU-WebXR and HFA, with mean $\pm$ standard deviation testing times for GlauCUTU-WebXR and HFA being 255 ± 28.7555 seconds and 721.0339 ± 114.2914 seconds, respectively [$t(58)=36.5714, p<0.001$]. The GlauCUTU-WebXR showed more consistent performance with testing times ranging from 202 to 308 seconds compared to HFA's range of 377 to 1066 seconds.

V. DISCUSSION AND CONCLUSION

This study focuses on enhancing the GlauCUTU perimetry system by improving and evaluating the HFA sensitivity prediction model using data collected from commercial VR devices. Building on previous work, the inclusion of PID and SeverityScore significantly improved the system's performance. These enhancements are illustrated in Fig. 2 and Fig. 4, which compare the performance of LWL, Voter+PID, and Voter+PID+SeverityScore models.

The incorporation of PID features addressed spatial accuracy limitations, particularly in critical areas like the blind spot. By accounting for location-specific sensitivity patterns, the model achieved more consistent performance across the visual field. For example, Figure. 4 shows that the PID feature smoothed sharp error spikes, reducing MAE at key positions, such as position 23 (the blind spot) in the LWL model. Adding SeverityScores further enhanced

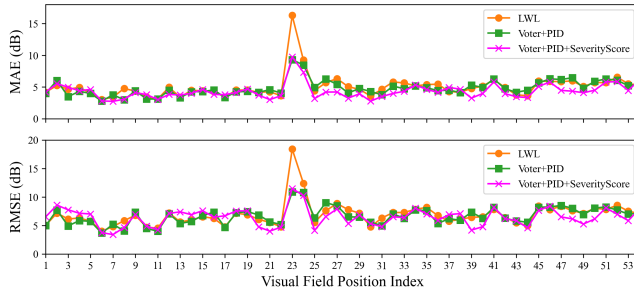


Fig. 4. Comparison of the average predicted HFA sensitivity across all 54 visual field test points in the test set for LWL, Voter+PID, and Voter+PID+SeverityScore. MAE (top) and RMSE (bottom) were calculated based on the error between the predicted and its corresponding ground truth at each VF position. Lines and markers represent the models: LWL (orange line with circle), Voter+PID (green line with square), and Voter+PID+SeverityScore (magenta line with cross).

performance, improving accuracy across the visual field. Although blind spot errors increased slightly, overall MAE and RMSE improved, demonstrating the value of integrating glaucoma severity into the regression model to capture high-level patterns.

Statistical analysis demonstrated the advantages of incorporating PID and SeverityScore features in over all cases, with only the Voter+PID+SeverityScore model showing no significant difference from the ground truth ($p=0.8270$). This finding was most pronounced in normal and mild cases, where the Voter+PID+SeverityScore model exhibited no significant deviation ($p=0.8750$ and $p=0.5888$, respectively) and achieved smaller $\bar{d}$ values compared to the LWL and Voter+PID models.

For mild cases, the Voter+PID+SeverityScore model achieved a $\bar{d} \pm S_D$ of only -1.4929 ± 6.9195 compared to larger S_D in the LWL (9.4385 ± 7.0668) and Voter+PID (8.9408 ± 8.0940) models, underscoring the value of incorporating spatial and severity information into predictions. In moderate and severe cases, all models performed similarly, showing no significant differences from the ground truth. This finding suggests that while our enhanced model improves accuracy in early-stage detection, there may be additional factors to consider for advanced disease states. The comparable performance across models in these cases might be attributed to the more pronounced and consistent patterns of visual field loss in advanced glaucoma.

The analysis of testing duration showed significant differences between the two methods. GlauCUTU-WebXR substantially reduced examination times (255 ± 28.7556 seconds) when compared to traditional HFA (721.0339 ± 114.2914 seconds, $p < 0.001$). Because GlauCUTU-WebXR can test both eyes at the same time, patients experienced less fatigue and greater comfort during the examination due to the shorter testing period.

Despite advancements, the models showed limitations in predicting moderate and severe glaucoma cases, with higher S_D ($S_D = 10.2914$ for moderate cases), indicating a need for more data targeting these severity levels. Hardware variability among commercial VR devices, such as differences in resolution and brightness, also requires attention [9], [11].

Further point-by-point VF analysis is needed to validate the model and assess the impact of additional features. Future efforts will address these challenges by expanding datasets, evaluating hardware variations, developing calibration protocols, and conducting clinical validations to improve reliability of the system.

The GlauCUTU-WebXR system builds upon our previous VR-based glaucoma screening work [8]. By leveraging commercial VR headsets and enhanced prediction models, it offers a portable, affordable, and scalable alternative to traditional tools. Features like PID and SeverityScore significantly improve the accuracy of HFA sensitivity predictions. Broader implementation could help improve early detection and treatment to reduce preventable vision loss, especially in settings without traditional perimetry.

VI. ACKNOWLEDGMENTS

The support of the Health Systems Research Institute (HSRI) of Thailand through grant number HSRI 68-042 greatly aided this study. Human subjects were involved as part of the research, and their involvement was conducted following the ethical guidelines and approved experimental protocols by the Chulalongkorn Institutional Review Board (Application No. 715/61).

REFERENCES

- [1] Y.-C. Tham *et al.*, "Global prevalence of glaucoma and projections of glaucoma burden through 2040: A systematic review and meta-analysis," *Ophthalmol.*, vol. 121, no. 11, pp. 2081–2090, 2014.
- [2] "Global estimates on the number of people blind or visually impaired by glaucoma: A meta-analysis from 2000 to 2020," *Eye*, pp. 1–11, 2024.
- [3] S. Asawaphureekorn *et al.*, "Health related quality of life instruments for glaucoma: A comprehensive review," *J Med Assoc Thai*, vol. 88, no. 9, pp. S155–62, 2005.
- [4] C. B. Estopinal *et al.*, "Access to ophthalmologic care in Thailand: A regional analysis," *Ophthalmic Epidemiol.*, vol. 20, no. 5, pp. 267–273, 2013.
- [5] Y. Wu *et al.*, "Measures of disease activity in glaucoma," *JBSBE*, vol. 196, p. 113700, 2022.
- [6] M. Montelongo *et al.*, "A virtual reality-based automated perimeter, device, and pilot study," *TVST*, vol. 10, no. 3, pp. 20–20, 2021.
- [7] H. Jayaram *et al.*, "Glaucoma: now and beyond," *Lancet*, vol. 402, no. 10414, pp. 1788–1801, 2023.
- [8] P. Kunumpol *et al.*, "GlauCUTU: Time until perceived virtual reality perimetry with humphrey field analyzer prediction-based artificial intelligence," *IEEE Access*, vol. 10, pp. 36 949–36 962, 2022.
- [9] S. Tsapakis *et al.*, "Visual field examination method using virtual reality glasses compared with the Humphrey perimeter," *Clin Ophthalmol*, pp. 1431–1443, 2017.
- [10] R. Razeghinejad *et al.*, "Preliminary report on a novel virtual reality perimeter compared with standard automated perimetry," *J. Glaucoma*, vol. 30, no. 1, pp. 17–23, 2021.
- [11] C. Bradley *et al.*, "Validation of a wearable virtual reality perimeter for glaucoma staging, the NOVA trial: Novel virtual reality field assessment," *TVST*, vol. 13, no. 3, pp. 10–10, 2024.
- [12] E. Hodapp *et al.*, *Clinical Decisions in Glaucoma*. Maryland Heights, MO, USA: Mosby, 1993.
- [13] Immersive Web Working Group, "WebXR device api specification," <https://github.com/immersive-web/webxr>, 2022, accessed: 2024-11-13.
- [14] E. Atalay *et al.*, "Pattern of visual field loss in primary angle-closure glaucoma across different severity levels," *Ophthalmol.*, vol. 123, no. 9, pp. 1957–1964, 2016.
- [15] L.-I. Lau *et al.*, "Patterns of visual field defects in chronic angle-closure glaucoma with different disease severity," *Ophthalmol.*, vol. 110, no. 10, pp. 1890–1894, 2003.