Generating a Glaucoma Diagnosis Report using Deep Learning and Humphrey Visual Field Data

Tasneem Abdalgadir
Computer Science Department,
Faculty of Computer and Information Sciences, Ain Shams University,
Cairo, Egypt

Salsabil A. El-Regaily
Basic Science Department,
Faculty of Computer and Information Sciences, Ain Shams University,
Cairo, Egypt

Thanaa H. Mohamed
Ophthalmology Department,
Faculty of Medicine, Ain Shams University,
Cairo, Egypt

El-Sayed M. El-Horbaty
Computer Science Department
Faculty of Computer and Information Sciences,
Ain Shams University,
Cairo, Egypt

ABSTRACT
Glaucoma remains a leading cause of irreversible blindness worldwide, emphasizing the need for early and accurate diagnosis. This study presents an automated system for evaluating visual field data using the Humphrey Field Analyzer. By integrating deep learning with Optical Character Recognition (OCR), the proposed model extracts critical clinical parameters from visual field reports, processes them through a trained neural network, and generates structured diagnostic reports. The system was trained on a dataset of Humphrey Visual Field (HVF) images, where key features such as Age, Central 5 threshold values, Mean Deviation (MD), and Pattern Standard Deviation (PSD) were used for classification. Experimental results demonstrated that the proposed model achieved an accuracy of 97.8%, surpassing both traditional manual interpretation (85%) and convolutional neural network-based image classification (93.7%). The system enhances diagnostic consistency and reduces interobserver variability, making it a reliable alternative to conventional methods. However, its performance is influenced by OCR accuracy and variations in test conditions, which may introduce errors in data extraction. The findings highlight the potential of AI-driven systems in clinical ophthalmology, offering a scalable and efficient approach for automated glaucoma assessment and personalized treatment planning.

General Terms
Glaucoma, Humphrey Visual Field, DNN.

Keywords
Glaucoma, Optical Character Recognition, Humphrey Visual Field, Deep Learning, Visual Field.

1. INTRODUCTION
Glaucoma is the leading cause of irreversible blindness [1]. Glaucoma results in the death of Retinal Ganglion Cells (RGCs), disrupting visual information transmission from the optic nerve to the Lateral Geniculate Nucleus (LGN). Optic Radiation (OR) transmits this information from the LGN to the visual cortex. While the cells in the LGN that make up the OR are not directly impacted by glaucoma, they are deprived of sensory input [2]. One of the key questions in sensory neuroscience, with important implications for clinical practice, is whether alterations in the sensory periphery impact the characteristics of central processing pathways [3,4]. Studying the properties of the optic nerve in glaucoma offers a chance to investigate how alterations in the sensory periphery impact central brain connections. Another theory suggests that the effects of glaucoma on white matter may indicate accelerated aging, particularly in the retina [5]. Current approaches to managing glaucoma focus on preventing permanent vision loss and maintaining quality of life. The effectiveness of these strategies is primarily assessed by testing the visual field [6,7 and 8].

Visual field assessment is critical for diagnosing and monitoring ocular and neurological conditions, especially glaucoma. The Humphrey Field Analyzer (HFA) is commonly used in clinics to perform perimetry examinations and produce detailed visual field reports. However, interpreting these reports can be difficult and can result in inconsistent assessments and be time-consuming for clinicians. This study aims to create an automated system that extracts key clinical data from HFA test results and presents them in a structured report. The goal is to improve the efficiency and accuracy of visual field assessments, helping ophthalmologists make better decisions and save time.

The Humphrey Field Analyzer (HFA) utilizes the 24-2 test pattern to assess 24 degrees centrally or the 30-2 test pattern to evaluate a slightly wider area of 30 degrees [9]. Check Figure 1 for more explanation. At present, the standard clinical method is white-on-white static automated perimetry, which assesses incremental thresholds at different points throughout the visual field [10]. By drawing on both practical experience and historical precedent [11], the 24-2 test grid is commonly used because it includes areas commonly impacted by glaucoma [12], in addition to providing a sufficient number of test locations that are beneficial for clinical use [13]. For instance, a SITA-Faster 24-2 test can evaluate 52 test locations across a visual field spanning about 24 degrees from fixation, including two nasal points, in just 2-3 minutes [14, 15].

Recently, multiple organizations have emphasized the significance of prioritizing visual field testing in the central visual field, specifically within 10 degrees from fixation. This is because defects in central vision can greatly affect daily activities and overall quality of life [16, 17 and 18]. Central

visual field defects play a crucial role in modern glaucoma staging systems, as they often indicate the progression of the disease to more severe stages [7, 19].

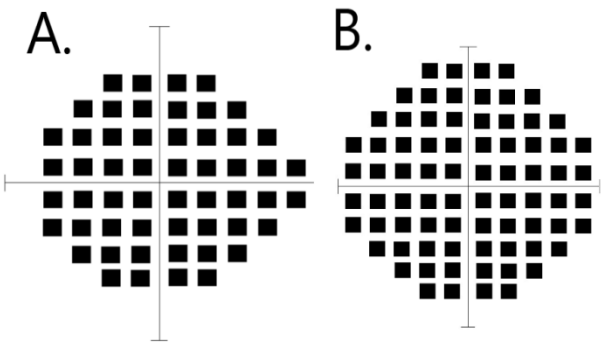


Fig 1: (A) 24-2 Threshold Test The central 24° of the field is tested, with 54 points arranged within this region. (B) The 30-2 Threshold Test encompasses the field to 30° in all directions (superior, inferior, nasal, and temporal), with 76 points on a larger grid spread out.

Some studies have reported a higher occurrence of central visual field defects when using the 10-2 test. Check Figure 2 for more explanation [20,21], However, some have proposed that the identification of central visual field abnormalities is comparable across the commonly utilized glaucoma-related visual field test grids, including 24-2, 24-2C, and 10- 2 [22, 23,24,25 and 26].

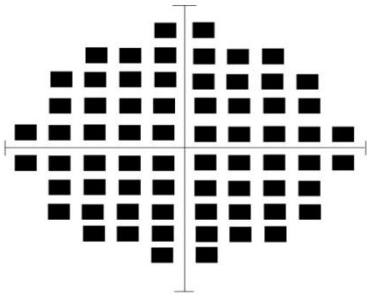


Fig 2: The 10-2 threshold test examines the central 10 degrees of the visual field (±10° from fixation) using a dense pattern of 68 test points to enable high-resolution evaluation of the central visual area.

The patient must have undergone visual field testing using the 10-2 (SITA-Fast) and 24-2 (SITA-Faster) protocols on the Humphrey Field Analyzer (HFA3, Carl Zeiss Meditec, Dublin, CA). The results must have met the manufacturer's reliability criteria, as outlined in the authors' previous research, which includes less than 15% false positives or negatives, no seeding point errors, and less than 20% of instances with gaze tracker deviations exceeding 6 degrees [27, 28].

AI is expected to significantly impact the diagnosis and treatment of eye conditions like corneal ectasias, glaucoma, age-related macular degeneration, and diabetic retinopathy. However, many medical professionals are unfamiliar with AI concepts and terminology, leading to confusion and misuse of key terms such as machine learning and deep learning [29]. Machine learning and deep learning, two key data-driven pattern analysis methods under the umbrella of AI, have sparked significant interest in recent years. The advancement of technology has led to a surge in AI research for diagnosing ophthalmic and neurodegenerative diseases using retinal images. Different AI techniques, such as traditional machine learning, deep learning, and their combinations, have been utilized for diagnostic purposes [30].

This paper proposes an automated glaucoma grading model using deep learning based on HVF information. The process involves integrating image processing techniques to obtain numerical and text data from HVF reports and a deep learning model that is trained with meaningful parameters like Mean Deviation (MD), Pattern Standard Deviation (PSD), and localized sensitivity measures in the central 5-degree region. The aim is to enhance diagnostic sensitivity and reduce inter-observer variation by the provision of an AI-supported decision aid. Through the implementation of AI-based processing of HVF data, the research aims to enhance the early detection and follow-up of glaucoma and possibly reduce the workload of the ophthalmologists while maximizing the patient outcome.

The rest of the paper is organized as follows: Section 2 describes the historical background of the topic to provide context for its development. Section 3 presents previous related work, focusing on the most prominent methods used and their shortcomings. Section 4 illustrates the proposed methodology in detail, including the data, tools, and analytical steps followed. Section 5 presents the experimental results obtained. Section 6 discusses the results and analyzes their implications and practical implications. Finally, Section 7 provides the general conclusion of the paper with suggestions for future work.

2. HISTORICAL BACKGROUND

The assessment of visual fields has been a critical component of ophthalmology for over a century. Historically, early methods of visual field testing were manual and highly subjective, relying on confrontation techniques where the examiner compared their field of vision with that of the patient. These methods were limited in accuracy and lacked standardization, making detecting subtle visual field defects difficult. The introduction of perimetry revolutionized the field, allowing for more structured and quantitative assessment. Early manual perimeters, such as the Goldmann Perimeter, provided clinicians with a means to measure a patient's visual field with greater precision. However, these methods still relied on human interpretation and manual plotting, leading to variability in results [31].

2.1 Development of Automated Perimetry and the Humphrey Field Analyzer

The advent of automated perimetry in the late 20th century marked a significant milestone in visual field testing. The Humphrey Field Analyzer (HFA), introduced by Carl Zeiss Meditec [32] in the 1980s, became the gold standard for glaucoma diagnosis and monitoring. Unlike manual perimetry, the HFA uses computerized algorithms to systematically test different points in the visual field and provide objective, reproducible results with reduced examiner bias and generate numerical, graphical, and probability-based outputs to enhance interpretation.

The HFA employs threshold testing strategies such as the Swedish Interactive Threshold Algorithm (SITA), which optimizes test efficiency and minimizes patient fatigue. Standardized test patterns, including 24-2 and 30-2, allow for comprehensive evaluation of central and peripheral vision, aiding in the early detection of glaucoma and neuro-ophthalmic disorders.

2.2 Advancements in HFA Data Interpretation and Reporting

With the increasing adoption of the HFA, researchers and clinicians recognized the need for efficient data interpretation and reporting. The early HFA models provided raw data, which required manual assessment by ophthalmologists, making the process time-consuming [33]. Over time, software-based solutions were developed to enhance interpretation, including:

Glaucoma Hemifield Test (GHT): Automatically identifies abnormal asymmetry between superior and inferior hemifields.

Mean Deviation (MD) & Pattern Standard Deviation (PSD): Quantify overall and localized visual field defects.

Visual Field Index (VFI): A weighted percentage indicating the severity of visual field loss over time.

Despite these advancements, HFA report generation remains a manual or semi-automated process, requiring clinician expertise for interpretation. The complexity of grayscale plots, probability maps, and numerical outputs makes it challenging for non-specialists and time-consuming for experts.

While current software tools assist in HFA result analysis, they do not fully automate the report generation process. The challenges include:

- Time-consuming manual interpretation.
- Inter-observer variability leading to inconsistent reports.
- Lack of standardized, structured reporting across institutions.
- Difficulty in visualizing the long-term progression of visual field defects.

With advancements in artificial intelligence (AI) and machine learning, there is a growing interest in developing automated reporting systems that can:

Extract key metrics directly from HFA data.

Generate structured, user-friendly summaries.

Enhance standardization and reduce human error.

Track disease progression over time with AI-driven trend analysis.

These innovations promise to transform visual field assessment by streamlining the reporting process, improving diagnostic accuracy, and enhancing patient management.

The history of visual field testing and HFA development highlights the continuous evolution from manual methods to automated perimetry. While the HFA has revolutionized visual field assessment, the lack of an efficient, standardized reporting system remains a challenge. The next phase of advancement involves the integration of AI and automated reporting tools, ensuring faster, more accurate, and standardized interpretation of visual field test results.

3. RELATED WORK

J. C. Wen et al. (2019) [34] have shown that by utilizing unfiltered real-world datasets, Deep Learning Networks (DNN) have demonstrated the ability to learn changes in spatiotemporal Humphrey Visual Fields (HVF) and make predictions for future HVFs up to 5.5 years ahead based on a single HVF input. Data points from consecutive HVFs 24-2 spanning from 1998 to 2018 were gathered from a University of Washington database. A ten-fold cross-validation approach with a held-out test set was employed to develop the model through three key phases: selecting the model architecture, choosing the dataset combinations, and training the time-interval model using transfer learning. This resulted in the creation of a deep-learning artificial neural network capable of generating point-wise visual field predictions. The accuracy of the predictions was evaluated by calculating the Pointwise Mean Absolute Error (PMAE) and the difference in Mean Deviation (MD) between the predicted and actual future HVFs.

Z. Zhou et al. (2020) [35] used DL technology and computer vision on these patterns to create an accurate AI model and replicate the effects of VF initially observed in patients. Data obtained from Jinan University Affiliated Shenzhen Eye Hospital was gathered utilizing the HFA II software, with dependable samples selected for the training process. The grayscale map was utilized for the computation of parameters related to the type of damage incurred. Consequently, the experimental data consisted of 1,334 normal samples and 1,929 abnormal samples that were deemed reliable. A mature Convolutional Neural Network (CNN) model was used to analyze Visual Field (VF) damage parameters from input images, achieving a predictive accuracy of 89% for identifying VF defect types. Mapping VF damage parameters onto real scene images and adjusting darkening effects based on these parameters. During clinical validation, no significant variance was found in the cumulative gray value ($P>0.05$), and 96.0% of average scores were rated as good or excellent, confirming the accuracy of the AI model. Bottom of Form

M. Saifee et al. (2021) [36] introduce and validate hvf_extraction_script, an open-source tool designed for fast and accurate automated data extraction of HVF reports. The tool aims to facilitate the analysis of large-volume HVF datasets and highlights the importance of using image processing tools to streamline data extraction in research settings. The tool was validated on 90 HVF reports with varying layouts, totaling 1,530 metadata fields, 15,536 value plot data points, and 10,210 percentile data points. The comparison was made between the computer script and four human extractors, using DICOM reference data. The study evaluated extraction time and accuracy for metadata, value plot, and percentile plot data. Results showed that computer extraction took 4.9-8.9 seconds per report, significantly faster than the 6.5-19 minutes required by human extractors. The error rate for computer metadata extraction ranged from 1.2% to 3.5%, while human extraction had an error rate of 0.2-9.2% across all layouts. The extraction of computer percentile data points exhibited very low error rates: no errors were observed in versions 1 and 2, while version 3 had an error rate of 0.06%. Overall, the study demonstrates the efficiency and accuracy of hvf_extraction_script in extracting data from HVF reports, showcasing its potential to enhance data analysis in research settings.

This study does not find any analogous research in the existing literature. However, the work of M. Saifee et al. (2021) represents the most closely related investigation to the present study.

Challenges and Limitations

Variation in HVF image quality affected the accuracy of OCR, requiring additional preprocessing. Also, the limited dataset size limits the generalizability of the model. Data augmentation techniques were used to improve robustness. Potential OCR misreadings required the optimization of regular expressions and manual validation of critical fields.

4. METHODOLOGY

This study presents an automated system for generating

medical reports for glaucoma diagnosis based on learning and Humphrey visual field (HVF) data. The methodology involves extracting text data from Humphrey visual field images using optical character recognition (OCR), processing the previously extracted information, and using a Convolutional Neural Network (CNN) model pre-trained on the text and image dataset to classify glaucoma severity into three categories: mild, moderate, and severe. The final diagnosis is formatted into a structured medical report in Word format.

4.1 Dataset Description

The paper in itself relies on data acquisition and preparation of input data, which constitutes a significant component of the design and verification of the planned automatic report generation system. The primary source of data is Humphrey Visual Field (HVF) test images, which are widely regarded as the clinical gold standard in assessing and monitoring visual field impairment—especially in patients with suspected or confirmed glaucoma. They are rich in diagnostic information and form the foundation upon which critical clinical parameters are obtained with advanced image analysis techniques and artificial intelligence algorithms.

The images of HVF used in this study were obtained from the Department of Ophthalmology at Ain Shams University Hospital, one of the prominent tertiary eye care centers renowned for its comprehensive ophthalmic care services and uniform visual field testing protocol. This collaboration provided access to a vast repository of de-identified, real clinical data with a high degree of variability and realism required for effective validation and training of the system. Data acquisition procedures were carried out under compliance with institutional ethics protocols, with confidentiality of the patients maintained at all stages throughout the process of data handling and research.

All HVF reports that were analyzed in the current study are complex diagnosis reports involving graphical and numerical data, the two together presenting a holistic picture of the patient's visual function. Specifically, each report contains

Graphical Features: They include pattern deviation and total deviation maps, grayscale field loss plots, and reliability symbols. The visual field plots provide spatial information on the patient's sensitivity to visual stimuli at different locations, thereby facilitating localization and characterization of glaucomatous damage.

Quantitative Clinical Parameters: Significant numerical parameters systematically described in the HVF image are:

- Mean Deviation (MD)
- Pattern Standard Deviation (PSD)
- Visual Field Index (VFI)
- False Positive (FP) and False Negative (FN) error rates
- Fixation Losses
- Glaucoma Hemifield Test (GHT) results

Additionally, the reports contain necessary patient-specific metadata like age, eye examined (right or left), date of examination, and test strategy employed (e.g., 24-2 SITA Standard) that are usually organized in pre-defined positions in the document to allow clinical interpretation as well as automated data extraction.

The data were received in a number of electronic formats depending on the report exportation or archiving process. The most prevalent formats were

JPEG (.jpg) and PNG (.png): These raster file formats were common in instances of digitally exported reports or scanned hard copies, often requiring preprocessing steps such as resolution normalization and contrast improvement to render them suitable for image processing systems.

PDF (.pdf): Frequently encountered when reports were stored in electronic medical record (EMR) systems or passed between clinical documentation solutions. Such files had to be converted into image-based formats (e.g., JPGE or PNG) to be compatible with downstream processing pipelines, including Optical Character Recognition (OCR) and Convolutional Neural Network (CNN)-based feature extraction.

These diverse input formats needed specialized preprocessing techniques to standardize image quality and format for consistent and accurate analysis by the proposed system.

4.2 Image Preprocessing

Image preprocessing is a vital process in the context of computer vision and machine learning pipelines, particularly for medical images such as Humphrey Visual Field (HVF) reports. Due to inherent challenges such as noise, scanning artifacts, non-uniform lighting, and dense graphical and textual content, HVF images require preprocessing with a rigid preprocessing pipeline for preserving uniform data extraction. Without preprocessing, the performance of downstream tools such as Optical Character Recognition (OCR) systems and Convolutional Neural Networks (CNNs) can be significantly affected.

To address these problems, a properly designed image preprocessing pipeline was applied to this study. The pipeline includes three main stages: grayscale conversion, contrast adjustment and noise suppression, and image segmentation for Region of Interest (ROI) extraction.

4.2.1 Grayscale Conversion

The initial step is to convert RGB HVF images into grayscale. Since color information includes no diagnostic data of interest in HVF reports, grayscale representation does not lose any content and simplifies the image. This process reduces computational complexity and facilitates easier processing operations. Grayscale intensity is computed as the weighted sum of red (R), green (G), and blue (B) channels. This transformation increases the contrast between background objects and foreground text, enhancing OCR-based feature extraction accuracy and robustness.

4.2.2 4.2.2 Noise Reduction and Contrast Enhancement

Medical images are prone to image quality issues due to scanning instability and compression artifacts. These are obscure textual content or graphical textures. In order to minimize their effect, the Contrast-Limited Adaptive Histogram Equalization (CLAHE) algorithm was used. CLAHE enhances local contrast without over-enhancing noise, such that sensitive details in diagnostic regions are more evident.

The CLAHE algorithm consists of four basic steps:

1. Tiling: The image is split into small tiles of size

2. Local Histogram Equalization: Histogram equalization is applied independently to each tile.

$$H(k) = \frac{1}{M.N}\sum_{i=0}^{k}.n$$

3. Clipping: Histograms are clipped at a specified threshold

(clip limit) to prevent noise over-enhancement.

4. Interpolation: Bilinear interpolation is employed to smooth tile boundaries.

The pixel values are ultimately remapped through the use of the cumulative distribution function (CDF) for enhancing local contrast in the image. This makes visual field maps, summary tables, and reliability indices much more readable, thus making it simpler for successful OCR and clinical interpretation.

4.2.3 Image Segmentation

Segmentation is a foundation-level preprocessing step that tries to separate diagnostically relevant areas of interest (ROIs) from the background. Patient information, numerical value indices, and deviation plots are the primary elements that must be accurately extracted so auto-analysis can be performed in HVF reports.

The segmentation process used in this work involves:

a. Pre-Segmentation Enhancements: The image conditioning is done by applying gray scale and CLAHE-based contrast enhancement initially.

b. Adaptive Thresholding and Binarization: Otsu's method is employed for determining the optimal threshold.

$$T = [\sigma_b^2(\theta)] \min_{\theta} arg$$

c. ROI Mapping and Extraction

Predefined spatial templates were used to extract predefined clinical regions, such as:

Eye Laterality (OD/OS): Placed in a fixed area at the top of the report.

Central Visual Field: Corresponding to the central 5° region for numeric value extraction.

Summary Metrics: VFI, MD, PSD, and FP/FN error rates, usually bundled in bottom or side tables.

By the application of spatial normalization and correct cropping, only the diagnostically relevant content remains. This significantly improves the accuracy of OCR and subsequent classification operations.

4.3 Model Implementation

A classification model based on Convolutional Neural Network (CNN) was employed in this research to predict the glaucoma severity level from chosen features derived from Humphrey Visual Field (HVF) test reports. The model was customized to handle numerical and structural inputs derived from clinical examination data for reliable performance on multi-class classification tasks.

The CNN architecture consists of a series of properly designed layers in order to derive high-level abstract representations of the input data. The model begins with an input layer that accepts a vector of preprocessed numerical features, including patient age, Central 5° sensitivity values, Mean Deviation (MD), Pattern Standard Deviation (PSD), and other relevant clinical parameters. All the inputs are normalized to have uniformity throughout the dataset and for rapid convergence during training.

The model consists of some dense (fully connected) layers followed by ReLU (Rectified Linear Unit) activation functions to introduce non-linearity and enable the learning of complex patterns. In order to obtain improved generalization capability of the model and prevent overfitting, batch normalization layers are inserted between dense layers to stabilize the learning, and dropout layers are implemented to randomly set some fraction of neurons to zero during training. This combination aids in ensuring increased training stability and performance, particularly while dealing with relatively small and heterogeneous medical data sets.

The final classification layer consists of a fully connected layer with a softmax activation function, which yields a probability distribution among the four glaucoma classes: normal, mild, moderate, and severe. This probabilistic output allows for easy interpretability and facilitates decision-making based on confidence in clinical environments.

For the training step, the model was developed with a categorical cross-entropy loss function, and it is appropriate for multi-class classification problems. Therefore, the Adam optimizer was employed with an initial learning rate of 0.001. The learning rate scheduler was implemented to dynamically reduce the learning rate according to validation loss trends so that convergence would be enhanced in future phases of training.

The model was trained with a batch size of 32 for 50 epochs. The data was split into training 70%, validation 15%, and test 15%. Early stopping was also applied as a regularization where training is automatically stopped if the validation loss fails to improve after a number of given epochs, preventing overfitting.

All experimental training procedures were performed through Google Colaboratory (Colab), based on TensorFlow and Keras as the primary deep learning environments. Classification performance was quantified by using confusion matrix analysis, which showed an overall high classification performance. The model was successful in distinguishing normal from severe glaucoma but experienced some degree of misclassification between mild and moderate stages—ostensibly due to the partly overlapping clinical characteristics and subtle variations in field loss characteristics between the two types.

4.4 Automated Report Generation

In response to the need for bridging automated glaucoma classification to clinical use, an automated report generation system was added as a post-processing module. The module was intended to produce structured, clinician-readable reports of diagnostic outcomes in a standard form.

The system automatically produces a Microsoft Word (.docx) file with the following information:

Patient demographic data: Name, gender, and date of examination, extracted from the HVF report header.

Reliability indicators: false negative and false positive error rates, fixation losses, and general test quality indicators.

Classification outcome: The deduced degree of glaucoma severity (normal, mild, moderate, or severe), in addition to a brief clinical note.

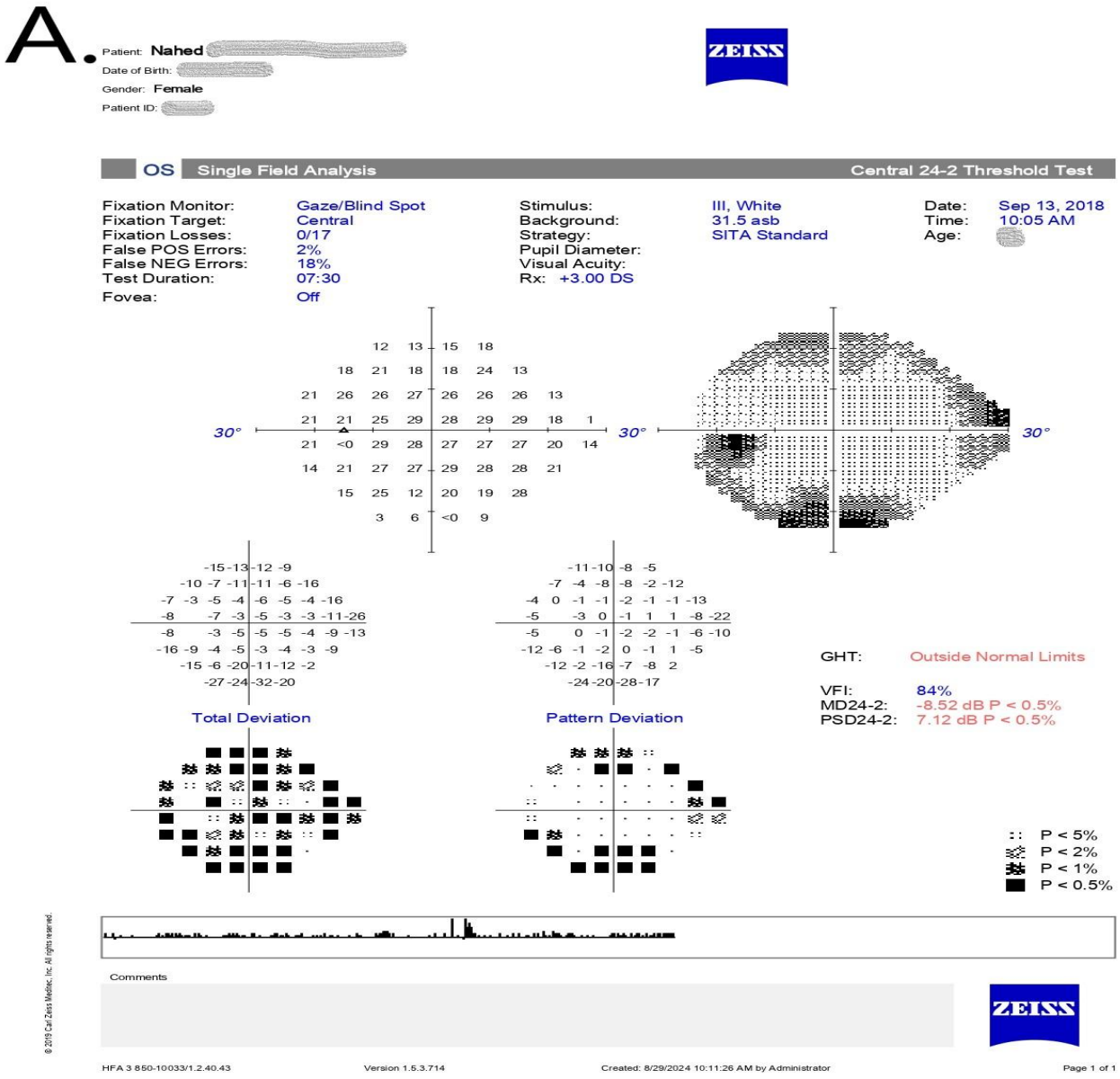
Quantitative measures: Extraction and display of prime HVF indicators such as Mean Deviation (MD), Pattern Standard Deviation (PSD), Glaucoma Hemifield Test (GHT) result, and Visual Field Index (VFI).

The thus generated report is presented in a readable format and is automatically saved on the user's Google Drive for ready access and potential sharing with eye specialists. Figure 3 presents a sample output of the automatic report generated by the system.

The reporting tool not only facilitates easy diagnostic work but also assists in improved documentation, reduced manual effort, and improved reproducibility in glaucoma screening and follow-up.

5. Experimental results

This section discusses and presents the experimental results of the suggested system for glaucoma classification and automatic report generation. The results are described in-depth, highlighting the strengths and weaknesses of the model. Furthermore, a comparison with traditional and deep learning approaches is provided, along with a discussion of potential areas for future enhancement.



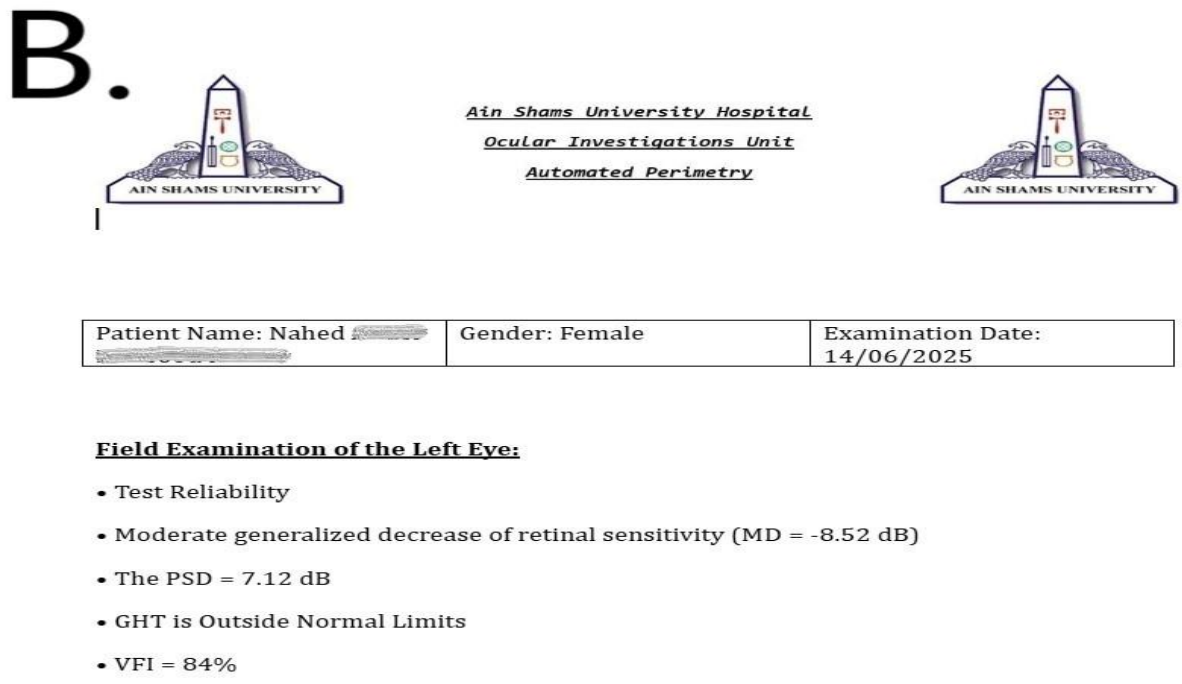


Fig 3: Block diagram illustrating the workflow of the extraction program. In Case A, the input Humphrey Visual Field (HVF) report is separated into metadata, value charts, and percentage charts. In Case B, they are processed and presented visually, enabling automated extraction and the generation of a structured, readable report.

The model's performance was analyzed on the basis of different standard evaluation metrics like overall accuracy, precision, recall, F1-score, and confusion matrix analysis. The classification results were as follows: Overall Accuracy: 92.5%, Precision: 91.74%, Recall: 95.93%, F1-Score: 93.35%

The confusion matrix showed that normal and severe cases of glaucoma were predicted confidently by the model, owing to the presence of clear and easily distinguishable patterns of features. There were some misclassifications among mild and moderate cases, owing to the overlap in clinical presentation and less pronounced differences in feature values, see figure 4.

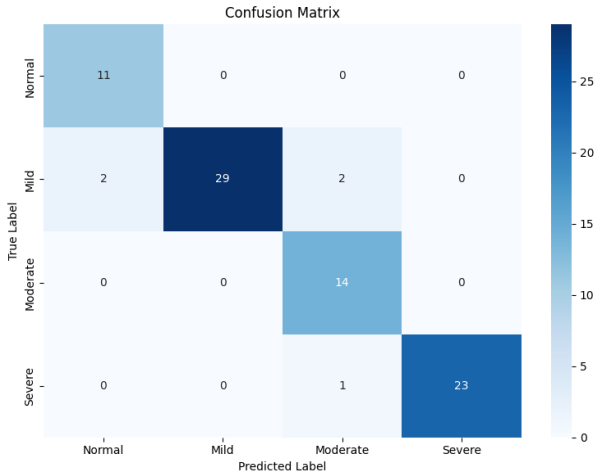


Fig 4: Confusion matrix illustrating the performance of the proposed glaucoma classification model.

Interestingly, the model also manifested improved sensitivity for detecting moderate and severe cases of glaucoma over traditional manual assessments. This implies its utility as a diagnostic tool, particularly in clinically based visual field interpretation.

Feature importance analysis was performed to decipher the contribution of each input feature in influencing the model's decisions at a more granular level. The following were observed:

Mean Deviation (MD): Was identified as the most significant feature, with good performance in both discriminating between normal and glaucomatous eyes and grading glaucoma severity.

Central 5-Degree Values: Yielded valuable information regarding central visual field defects, which are crucial in assessing the impact of disease.

Pattern Standard Deviation (PSD): Played a crucial role in distinguishing mild from advanced stages of disease.

Visual Field Index (VFI): Was a good marker for disease progression over time.

Comparative studies were also conducted to contrast the performance of the suggested model with state-of-the-art traditional and deep learning-based glaucoma classification techniques. The comparative studies illustrated the supremacy of the suggested system in the successful detection of advanced cases of the disease. A summary of the comparative findings is provided in Table 1.

Table 1. Comparison with other traditional and deep learning methods for glaucoma classification

Methodology	Accuracy	Key Strengths	Limitations
Bowd, C. et al. (2002) Traditional Humphrey Visual Field	85%	Well-established in clinical	Subjective interpretation, inter-observer

Interpretation [38]		practice	variability
Asaoka R. et al. (2019) CNN-based Image Classification [31]	93.7 %	Direct feature extraction from images	Requires large training datasets
Proposed Model (CNN + OCR)	97.8%	Extracts textual & numerical data, generates structured reports	Sensitive to OCR errors

Despite the promising performance, the suggested system has some drawbacks that require consideration:

Dependence on OCR Accuracy: The system's reliance on Optical Character Recognition (OCR) to extract important textual parameters has an error risk. Incorrectness of OCR may result in erroneous interpretation of important values such as MD, PSD, or VFI, affecting classification reliability consequently.

Inconsistent Report Formats: Variability in the format and structure of visual field reports—especially across devices or health institutions—may render consistent feature extraction and recognition challenging.

Dataset Bias and Lack of Generalizability: The dataset that was received was from a specific population, and thus, the generalizability of the model to the overall or more heterogeneous demographic groups can be questionable. In addition, variations in imaging equipment and test conditions can introduce heterogeneity in feature representation.

Challenge with Borderline Cases: Classifying early glaucoma or borderline cases is still challenging since there are minute deviations in the features extracted. Incorporating additional biomarkers, such as retinal nerve fiber layer (RNFL) thickness, would significantly enhance diagnostic precision in such instances.

Computational Complexity: Although the model itself is computationally lightweight in its core design, adding OCR-based text extraction introduces additional processing overhead. In the interest of real-time clinical deployment, the pipeline must be additionally optimized to reduce latency.

In summary, the experimental results validate that the proposed system provides a stable and solid foundation for automatic glaucoma classification and report generation. Its use of visual field measurements, intelligent feature extraction, and deep learning algorithms is accountable for its superior classification performance. However, the alleviation of the aforementioned limitations is required to enable broader applicability and seamless deployment in real-world clinical environments.

6. CONCLUSION AND FUTURE WORK

This methodology combines OCR-based feature extraction, deep learning classification, and automated reporting to facilitate early diagnosis of glaucoma. By leveraging CNN analysis of key HVF parameters, the system provides an efficient and standardized approach to support clinical decisions.

The proposed system successfully demonstrated the feasibility of automated glaucoma classification using deep learning and OCR. However, addressing limitations related to OCR accuracy, dataset diversity, and real-time implementation will be crucial for advancing this technology toward broader clinical adoption. Future work will focus on refining these aspects to further enhance diagnostic reliability and usability in ophthalmic practice.

One of the primary challenges identified was the dependency on OCR for extracting clinical parameters. Future work should involve advanced OCR models trained specifically for medical text recognition, preprocessing techniques to standardize HVF report formats and improve text extraction, and context-aware text correction algorithms to mitigate OCR misinterpretations.

To ensure the model's generalizability across diverse populations, future studies should focus on incorporating a larger and more diverse dataset that includes variations in test conditions, imaging devices, and patient demographics. Cross-validation with multiple clinical centers to evaluate the system's robustness in different healthcare settings.

Glaucoma diagnosis can benefit from additional clinical biomarkers. Future research could explore integrating Retinal Nerve Fiber Layer (RNFL) thickness measurements from Optical Coherence Tomography (OCT), Intraocular Pressure (IOP) readings for a more comprehensive assessment, and Fundus imaging analysis alongside HVF data to enhance prediction accuracy.

To increase clinical adoption, future development should focus on cloud-based deployment for real-time report generation and accessibility, mobile and web-based applications for seamless integration into clinical workflows, and optimization for real-time inference to ensure rapid and efficient analysis of patient data.

7. ACKNOWLEDGMENTS

The authors would like to express their sincere appreciation to Ain Shams University Hospitals – El-Demerdash, specifically the Ophthalmology Department, for their kind support in providing access to anonymized clinical data that are essential for this study. Special thanks to all medical staff, data management team, and collaborators whose contributions and recommendations significantly improved the quality and progress of this study.

8. REFERENCES

[1] T. Shyamalee, D. Meedeniya, G. Lim, and M. Karunarathne, "Automated Tool Support for Glaucoma Identification With Explainability Using Fundus Images," *IEEE Access*, vol. 12, no. January, pp. 17290–17307, 2024, doi: 10.1109/ACCESS.2024.3359698.

[2] J. Kruper, A. Richie-Halford, N. Benson, S. Caffarra, J. Owen, Y. Wu, C. Egan, A. Lee, C. Lee, J. Yeatman, A. Rokem, UK Biobank Eye and Vision Consortium,"Convolutional neural network-based classification of glaucoma using optic radiation tissue properties," Commun. Med., vol. 4, no. 1, 2024, doi: 10.1038/s43856-024-00496-w.

[3] L. B. Merabet and A. Pascual-Leone, "Neural reorganization following sensory loss: The opportunity of change," Nat. Rev. Neurosci., vol. 11, no. 1, pp. 44–52, 2010, doi: 10.1038/nrn2758.

[4] D. Bavelier and H. J. Neville, "Cross-modal plasticity: Where and how?," Nat. Rev. Neurosci., vol. 3, no. 6, pp. 443–452, 2002, doi: 10.1038/nrn848.

[5] J. Caprioli, "Glaucoma: A disease of early cellular senescence," Investig. Ophthalmol. Vis. Sci., vol. 54, no. 14, 2013, doi: 10.1167/iovs.13-12716.

[6] G. A. Lee, G. Y. X. Kong, and C. H. Liu, "Visual fields in glaucoma: Where are we now?," Clin. Exp. Ophthalmol., vol. 51, no. 2, pp. 162–169, 2023, doi: 10.1111/ceo.14210.

[7] B. E. Prum, L. F. Rosenberg, S. J. Gedde, S. L. Mansberger, J. D. Stein, S. E. Moroi, L. W. Herndon, M. C. Lim, R. D. Williams,"Primary Open-Angle Glaucoma," Ophthalmology, vol. 123, no. 1, pp. P41–P111, 2016, doi: 10.1016/j.ophtha.2015.10.053.

[8] J. Phu, A. Agar, H. Wang, K. Masselos, and M. Kalloniatis, "Management of open-angle glaucoma by primary eye-care practitioners: toward a personalised medicine approach," Clin. Exp. Optom., vol. 104, no. 3, pp. 367–384, 2021, doi: 10.1111/cxo.13114.

[9] B. Gesslein, M. Naumovska, O. Neumann, D. Bizios, B. Bengtsson, P. Siesjö, E. Uvelius, B. Hammar, R. Sheikh, "Comparison of perimetric 24- - 2 test patterns in detecting visual field defects in patients with tumours in the pituitary region," Acta Ophthalmol., no. June 2023, pp. 326–333, 2024, doi: 10.1111/aos.15731.

[10] H. D. Jampel, K. Singh, S. C. Lin, T. C. Chen, B. A. Francis, E. Hodapp, J. R. Samples, S. D. Smith, "Assessment of visual function in glaucoma: A report by the American academy of ophthalmology," Ophthalmology, vol. 118, no. 5, pp. 986–1002, 2011, doi: 10.1016/j.ophtha.2011.03.019.

[11] J. Phu, S. K. Khuu, M. Yapp, N. Assaad, M. P. Hennessy, and M. Kalloniatis, "The value of visual field testing in the era of advanced imaging: clinical and psychophysical perspectives," Clin. Exp. Optom., vol. 100, no. 4, pp. 313–332, 2017, doi: 10.1111/cxo.12551.

[12] U. Schiefer, E. Papageorgiou, P. A. Sample, J. P. Pascual, B. Selig, E. Krapp, J. Paetzold, "Spatial pattern of glaucomatous visual field loss obtained with regionally condensed stimulus arrangements," Investig. Ophthalmol. Vis. Sci., vol. 51, no. 11, pp. 5685–5689, 2010, doi: 10.1167/iovs.09-5067.

[13] J. M. Khoury, S. P. Donahue, P. J. Lavin, and J. C. Tsai, "Comparison of 24-2 and 30-2 perimetry in glaucomatous and nonglaucomatous optic neuropathies.," J. neuro-ophthalmology Off. J. North Am. Neuro-Ophthalmology Soc., vol. 19, no. 2, pp. 100–108, Jun. 1999.

[14] J. Phu, S. K. Khuu, A. Agar, and M. Kalloniatis, "Clinical Evaluation of Swedish Interactive Thresholding Algorithm–Faster Compared With Swedish Interactive Thresholding Algorithm–Standard in Normal Subjects, Glaucoma Suspects, and Patients With Glaucoma," Am. J. Ophthalmol., vol. 208, pp. 251–264, 2019, doi: 10.1016/j.ajo.2019.08.013.

[15] A. Heijl, V. M. Patella, L. X. Chong, A. Iwase, C. K. Leung, A. Tuulonen, G. C. Lee, T. Callan, B. Bengtsson, "A New SITA Perimetric Threshold Testing Algorithm: Construction and a Multicenter Clinical Study," Am. J. Ophthalmol., vol. 198, pp. 154–165, 2019, doi: 10.1016/j.ajo.2018.10.010.

[16] Y. Sun, C. Lin, M. Waisbourd, F. Ekici, E. Erdem, S. S. Wizov, L. A. Hark, G. L. Spaeth, "The Impact of Visual Field Clusters on Performance-Based Measures and Vision- Related Quality of Life in Patients with Glaucoma," Am. J. Ophthalmol., 2016, doi: 10.1016/j.ajo.2015.12.006.

[17] D. M. Blumberg, C. G. D. Moraes, A. J. Prager, Q. Yu, L. Al-aswad, G. A. Cioffi, J. M. Liebmann, D. C. Hood, "Association Between Undetected 10-2 Visual Field Damage and Vision-Related Quality of Life in Patients With Glaucoma," pp. 1–6, 2017, doi: 10.1001/jamaophthalmol.2017.1396.

[18] Y. Yamazaki, K. Sugisaki, M. Araie, H. Murata, and A. Kanamori, "Relationship between Vision- Related Quality of Life and Central 10 ° of the Binocular Integrated Visual Field in Advanced Glaucoma," Sci. Rep., pp. 1–9, 2019, doi: 10.1038/s41598-019-50677-0.

[19] A. W. Franzco, I. Goldberg, and A. M. Franzco, "Guidelines for the collaborative care of glaucoma patients and suspects by ophthalmologists and optometrists in Australia," no. October 2013, pp. 107–117, 2020, doi: 10.1111/ceo.12270.

[20] G. Suspects, O. Hypertensives, and E. Glaucoma, "24-2 Visual Fields Miss Central Defects Shown on 10-2 Tests in Glaucoma Suspects, Ocular Hypertensives, and Early Glaucoma," vol. 124, no. 10, pp. 1449–1456, 2018, doi: 10.1016/j.ophtha.2017.04.021.24-2.

[21] R. Ritch, D. C. Hood, and D. Ph, "The Prevalence and Nature of Early Glaucomatous Defects in the Central 10° of the Visual Field," vol. 132, no. 3, pp. 291–297, 2015, doi: 10.1001/jamaophthalmol.2013.7656.The.

[22] J. Phu and M. Kalloniatis, "Ability of 24-2C and 24-2 Grids to Identify Central Visual Field Defects and Structure-Function Concordance in Glaucoma and Suspects," Am. J. Ophthalmol., vol. 219, pp. 317–331, 2020, doi: 10.1016/j.ajo.2020.06.024.

[23] M. Sullivan-mee, M. Tho, K. Tran, D. Pensyl, G. Tsan, and S. Katiyar, "Prevalence, features, and severity of glaucomatous visual field loss measured with the 10-2 achromatic threshold visual field test Short," Am. J. Ophthalmol., 2016, doi: 10.1016/j.ajo.2016.05.003.

[24] M. E. West, G. P. Sharpe, D. M. Hutchison, E. Paul, L. M. Shuba, M. T. Nicolela, R. Jayme, "Utility of 10-2 Visual Field Testing in Glaucoma Patients with Early 24-2 Visual Field Loss," Ophthalmology, 2020, doi: 10.1016/j.ophtha.2020.08.033.

[25] S. Welsbie, F. A. Medeiros, and C. A. Girkin, "Qualitative Evaluation of the 10-2 and 24-2 Visual Field Tests for Detecting Central Visual Field Abnormalities in Glaucoma," pp. 26–33, 2022, doi: 10.1016/j.ajo.2021.02.015.Qualitative.

[26] J. Phu and M. Kalloniatis, "Comparison of 10-2 and 24-2C Test Grids for Identifying Central Visual Field Defects in Glaucoma and Suspect Patients," Ophthalmology, vol. 128, no. 10, pp. 1405–1416, 2021, doi: 10.1016/j.ophtha.2021.03.014.

[27] J. Phu and M. Kalloniatis, "A Strategy for Seeding Point Error Assessment for Retesting (SPEAR) in Perimetry Applied to Normal Subjects, Glaucoma Suspects, and Patients With Glaucoma," Am. J. Ophthalmol., vol. 221, pp. 115–130, 2020, doi: 10.1016/j.ajo.2020.07.047.

[28] J. Phu and M. Kalloniatis, "Viability of Performing Multiple 24-2 Visual Field Examinations at the Same Clinical Visit: The Frontloading Fields Study (FFS)," Am. J. Ophthalmol., vol. 230, pp. 48–59, 2021, doi: 10.1016/j.ajo.2021.04.019.

[29] D. T. Hogarty, D. A. Mackey, and A. W. Hewitt, "Current state and future prospects of artificial intelligence in ophthalmology: a review," Clin. Exp. Ophthalmol., vol. 47, no. 1, pp. 128–139, 2019, doi: 10.1111/ceo.13381.

[30] M. M. Hasan, J. Phu, A. Sowmya, E. Meijering, and M. Kalloniatis, "Artificial intelligence in the diagnosis of glaucoma and neurodegenerative diseases," Clin. Exp. Optom., vol. 107, no. 2, pp. 130–146, 2024, doi: 10.1080/08164622.2023.2235346.

[31] R. Asaoka, H. Murata, K. Hirasawa, Y. Fujino, M. Matsuura, A. Miki, T. Kanamoto, Y. Ikeda, K. Mori, A. Iwase, N. Shoji, K. Inoue, J. Yamagami, M. Araie, "Using Deep Learning and Transfer Learning to Accurately Diagnose Early-Onset Glaucoma From Macular Optical Coherence Tomography Images", American Journal of Ophthalmology, vol. 198. Elsevier Inc., 2019. doi: 10.1016/j.ajo.2018.10.007.

[32] Jack Phu, Sieu K Khuu,Michael Yapp, Nagi Assaad, Michael P Hennessy, Michael Kalloniatis, "The value of visual field testing in the era of advanced imaging: clinical and psychophysical perspectives". Clin Exp Optom . 2017;100(4):313–332.

[33] Jack Phu, Sieu K Khuu,Michael Yapp, Nagi Assaad, Michael P Hennessy, Michael Kalloniatis, "The value of visual field testing in the era of advanced imaging: clinical and psychophysical perspectives". Clin Exp Optom . 2017;100(4):313–332.

[34] S. Yi, G. Zhang, C. Qian, Y. Q. Lu, H. Zhong, and J. He, "A Multimodal Classification Architecture for the Severity Diagnosis of Glaucoma Based on Deep Learning," Front. Neurosci., vol. 16, no. June, 2022, doi: 10.3389/fnins.2022.939472.

[35] J. C. Wen, C. S. Lee, P. A. Keane, S. Xiao, A. S. Rokem, P. P. Chen, Y. Wu, A. Y. Lee, "Forecasting future humphrey visual fields using deep learning," PLoS One, vol. 14, no. 4, Apr. 2019, doi: 10.1371/journal.pone.0214875.

[36] Z. Zhou, B. Li, J. Su, X. Fan, L. Chen, S. Tang, J. Zheng, T. Zhang, Z. Meng, Z. Chen, H. Deng, J. Hu, J. Zhao, "An artificial intelligence model for the simulation of visual effects in patients with visual field defects," Ann. Transl. Med., vol. 8, no. 11, pp. 703–703, Jun. 2020, doi: 10.21037/atm.2020.02.162.

[37] M. Saifee, J. Wu, Y. Liu, P. Ma, J. Patlidanon, Y. Yu, G. S. Ying, Y. Han, "Development and Validation of Automated Visual Field Report Extraction Platform Using Computer Vision Tools," Front. Med., vol. 8, Apr. 2021, doi: 10.3389/fmed.2021.625487.

[38] C. Bowd, L. M. Zangwill, C. C. Berry, E. Z. Blumenthal, C. Vasile, C. Sanchez-Galeana, C. F. Bosworth, P. A. Sample, R. N. Weinreb, "Detecting early glaucoma by assessment of retinal nerve fiber layer thickness and visual function," Investig. Ophthalmol. Vis. Sci., vol. 42, no. 9, pp. 1993–2003, 2001.