ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

Interpretable Glaucoma Classification and Severity Assessment Using Deep Learning Models

Adla Ananya,
PG Scholar, Department of Computer Science and Engineering, Jawaharlal Nehru Technological University Hyderabad University College of Engineering Jagtial(Autonomous), Nachupally(Kondagattu), Telangana, India - 505501. Email: edlaananya222@gmail.com

Dr. B. Sateesh Kumar
Professor, Department of Computer Science and Engineering, Jawaharlal Nehru Technological University Hyderabad University College of Engineering Jagtial(Autonomous), Nachupally(kondagattu), Telangana, India - 505501. Email: sateeshbkumar@jntuh.ac.in

ABSTRACT

Globally, glaucoma is a major cause of permanent blindness and early and accurate diagnosis is essential to saving sight and enhancing patient outcomes. Retinal fundus imaging can be used for non-invasive glaucoma screening. However, regular manual evaluation is time consuming and requires a lot of clinical expertise and is prone to inter-observer variability. The categorisation of glaucoma was done using a publicly available dataset of retinal fundus images from multiple ophthalmic sources. The photos were acquired during the past few years containing normal and glaucomatous retinal samples. The designed preparation pipeline aims to increase the resilience and generalisation of the models by image scaling, normalisation, contrast enhancement, data augmentation, and dataset balancing . Various DL models such as ResNet50, DenseNet201, ViT, ViT-CapsNet, Xception and ConvNeXtTiny were compared to classify glaucoma. The task of segmentation of optic disc and cup was done by EfficientUNet to estimate CDR and the task of glaucoma localisation was done by the YOLO based models. Visual Explanation for Interpretability Improvement: Grad-CAM and Grad-CAM++. The accuracy, precision, recall, F1-score, Dice Score, IoU and mAP served as performance metrics. DenseNet201 attained the highest classification accuracy of 89.60%, while EfficientUNet performed best at 95/82.40% Dice Scores for optic disc and optic cup segmentation, respectively, and YOLOv5s6 had mAP@0.5 of 97.50%. To improve computer-aided ocular screening, the integrated framework offers the following: Accurate diagnosis, interpretable prediction, anatomical analysis and clinically useful glaucoma assessment.

Index Terms: *Glaucoma classification, Vision Transformer, Capsule Network, retinal fundus images, explainable artificial intelligence, deep learning, medical image analysis.*

INTRODUCTION

Glaucoma is a slowly evolving optic neuropathy and one of the leading causes of permanent vision loss worldwide. The disease gradually eats away the optic nerve, and may have no signs or symptoms early in its course, which is why early diagnosis is important to prevent permanent loss of sight. The clinical evaluation of glaucoma is typically accomplished by checking the retinal fundus, optic disc, optic cup and CDR. But manual diagnosis requires high clinical expertise, and is prone to inter observer variation particularly for large population screening. To this end, there has been an increasing interest in developing computer-aided diagnostic tools to help ophthalmologists via accurate and efficient retinal image processing [1]. In recent years, automated glaucoma detection has made great progress thanks to major developments in AI and DL. CNNs like ResNet, DenseNet, Xception and ConvNeXt have demonstrated good performance in automatically learning the discriminative features from retinal fundus images [2]. Moreover, segmentation models like U-Net and its variants have been used for accurately extracting features related to the optic disc and optic cup and evaluating the glaucoma severity by CDR calculation [3]. In addition, object identification models have been applied in localising clinically-important retinal regions [4] and XAI approaches have been suggested to make DL predictions more transparent [5]. However, there are issues with the literature available. Traditional CNN based approaches only focus on local spatial information and are not sufficient to capture long-range contextual interactions within the retinal image. Furthermore, there are many existing research that focus on only one diagnostic task (e.g. classification, segmentation or localisation) and do not present an integrated and interpretable diagnostic framework. Additionally, explainability techniques are not widely employed, which impacts trust in automated decision-making systems, highlighting the need for more transparent AI systems in medical picture analysis [6]. To tackle these limitations, current works have been focusing on ViT designs which utilise self-attention processes to capture the global contextual dependencies of the retinal fundus images [7]. CapsNet have also shown a lot of interest, with their ability to preserve hierarchical spatial relations between retinal anatomical properties [8]. The combination of Vision Transformers global feature learning and Capsule Networks spatial feature learning can be a potential avenue for improving the glaucoma classification. However, by combining retinal structure segmentation, glaucoma region localisation and XAI, more comprehensive and clinically interpretable diagnostic tools are developed [9] [10]. Inspired by these progresses, this survey gives a complete evaluation of interpretable DL methods for glaucoma classification from retinal fundus pictures. The current research on Vision Transformers, Capsule Networks, retinal image segmentation, glaucoma localisation, and XAI are methodologically classified in this survey. It also contrasts good practice studies, summarises datasets and evaluation criteria that are widely used, analyses current issues in the field and offers some possible future directions. This survey aims to show the current advancements and provide a well-defined evaluation of the state of the art and emerging trends in interpretable AI for automated glaucoma diagnosis for academics and practitioners.

SURVEY METHODOLOGY

In this survey, we use a systematic literature review methodology and provide a comprehensive and unbiased analysis of the literature of interpretable DL approaches for glaucoma classification using retinal fundus images [11]. The methodology is designed to ensure maximum coverage of recent advances in Vision Transformers, Capsule Networks, retinal structure segmentation, glaucoma localisation, and XAI while minimising the selection bias. The entire review process involves identifying literature, selecting the studies, categorising them, comparing and synthesising research results [12].

A. Data Sources

The literature assessed was retrieved from trusted scientific databases that publish good quality peer-reviewed research in fields of AI, medical image analysis, ophthalmology, and computer vision. The major sources are:

- IEEE Xplore
- Direct Science
- Springer link
- Scopus
- Google scholarly

The digital libraries include a wide range of academic articles, conference proceedings and review papers on automated glaucoma diagnosis and interpretable DL.

B. Search Strategy

Keywords were used to find relevant papers using a search method. Multiple search terms were created using the following keywords:

- Glaucoma Classification

- ViT
- CapsNet
- Fundus Retinal Image
- Medical Image Computing
- XAI
- Grad-CAM
- Segmentation of optic disc
- Segmentation of Optic Cup
- DL for Glaucoma diagnosis is a tool that is creating a significant impact in the diagnosis of glaucoma.

Boolean operators (AND, OR, NOT) were used to narrow and enhance the relevance of the papers retrieved.

C. Inclusion and Exclusion Criteria

The literature selection procedure used was inclusion-exclusion criteria to ensure the quality and relevance of literature reviewed.

Inclusion Criteria

- Peers reviewed papers and conference papers.
- Publications in the English language.
- Studies published between 2018-2025.
- Research interest in DL and Vision Transformers, Capsule Networks, retinal image segmentation, XAI, and glaucoma diagnosis.
- Experimental evaluation studies on publicly available retinal imaging dataset.

Exclusion Criteria

- Duplicate publications.
- Articles, editorials and opinion papers that are not peer reviewed.
- Non-glaucoma studies, non-image analysis of retina.
- Publications with insufficient technical or experimental information.

D. Study Selection and Analysis

A first phase screening was done based on the title and abstract of the collected publications to decide whether or not the publication was relevant to the survey question. Afterwards, a full text review was performed to validate the technical contributions, methodology, datasets, assessment metrics and experimental findings of the shortlisted publications. We carefully examined each of the selected papers and classified them as per their primary research focus which includes glaucoma classification, Vision Transformer based approaches, Capsule Network models, retinal structure segmentation, glaucoma localisation, XAI, and integrated diagnostic frameworks.

Finally, a comparative analysis of the selected studies was done to identify the research trends, representative datasets, assessment methodology, strengths and limitations of the studies, and existing research gaps. The synthesized results provide the organized understanding of current evolutions and hint to future research directions regarding interpretable DL for automated glaucoma classification.

TAXONOMY AND CLASSIFICATION OF EXISTING APPROACHES

The previous works are divided into different categories according to the main diagnostic task and DL approach [13] to comprehensively review the literature on automated glaucoma diagnosis. We have proposed a taxonomy that classifies the studies investigated in this paper into five major categories: glaucoma classification methods, Vision Transformer-based methods, Capsule Network-based methods, retinal structure segmentation methods, and interpretable DL frameworks. This classification helps in systematically understanding the research trends and indicates the evolution of intelligent glaucoma diagnosis systems [14][15].

A. CNN-Based Glaucoma Classification Approaches

Most of the early models used for glaucoma detection were built upon CNNs and the task of the classifiers was to classify an image of the fundus of the eye as normal or glaucomatous. Handmade feature extraction is no longer required as popular architectures such as ResNet, DenseNet, VGGNet, Xception, and ConvNeXt automatically learn the visual features hierarchically. These algorithms have shown good classification accuracy on publicly available retinal imaging datasets and have greatly improved automated screening of glaucoma. However, typical CNNs mainly learn local spatial information, and often miss the long-range contextual linkages in retinal images, which limits their potential to exploit the global structural information.

B. Vision Transformer-Based Approaches

In recent years, ViT architectures have gained popularity for glaucoma classification because of their ability to model the self-attention mechanism which captures the global contextual dependency. Vision Transformers eliminate the need for the convolution procedures and divide the retinal fundus images into fixed-size patches and use transformer encoders to process these patches. By doing this, effective learning of the relation between the different portions of the retina can be done and therefore, more feature representations are obtained when dealing with complex medical image analysis problems. The models using Vision Transformers have demonstrated competitive performance in glaucoma diagnosis, especially when provided with large and diverse retinal image datasets.

C. Capsule Network-Based Approaches

CapsNet have been suggested as a novel DL model for medical image classification which preserves the spatial and hierarchical relationships between image components. Unlike conventional convolutional networks, Capsule Networks learn features in capsules, which represent both the presence and direction of a feature. The dynamic routing method can be used to model the structural relationship, and thus CapsNet is particularly suitable for retinal image processing, which appears to be of great importance in the diagnosis. Recently the performance of glaucoma classification has been enhanced by studying the individual Capsule Networks and also in combination with transformer-based models.

D. Retinal Structure Segmentation and Localization Approaches

Several works expand the work on glaucoma categorisation to the diagnosis of glaucoma using retinal structure segmentation and glaucoma region localisation. Automatic segmentation of OD and OC has been widely performed by the Encoder-Decoder approaches like U-Net, U-Net++ [30, 31], ResUNet [32] and EfficientUNet [33] to aid in computing CDR for assessing glaucoma severity. Further, YOLO family of object detection models have been applied to localise clinically significant retinal regions, which offers complementary visual information to assist automated diagnosis and enhance the clinical interpretation.

E. Explainable and Integrated Diagnostic Frameworks

As AI becomes a key component of healthcare, model interpretability and clinical transparency are more critical than ever. Thus, recent studies have resorted to XAI techniques like Grad-CAM and Grad-CAM++ to highlight areas of the retina that are relevant to glaucoma prediction. Besides, a few studies have also introduced multimodal diagnostic pipelines that can classify and segment the structures in the retina, localise and visualise the glaucoma and explain the results in the same pipeline. These complete frameworks aim to break the barrier among the different diagnostic methods and to design more consistent, interpretive and useful computer aided glaucoma diagnosis systems.

COMPARATIVE ANALYSIS OF SURVEYED STUDIES

This section includes a comparative study of literature evaluated for automatic glaucoma diagnosis utilising retinal fundus images. Two aspects of comparison are the main methodological approach and the diagnostic goal, which determines the datasets, the evaluation measures, and the stated limitations [16]. This survey provides a detailed review of the ongoing research trends by studying the strengths and weaknesses of some DL techniques, and suggesting directions for future research [17].

Comparison of CNN-Based Classification Approaches

To provide automated glaucoma classification, CNN based models have been widely adopted, using strong feature extraction capabilities. To achieve good classification performance on retinal fundus pictures, ResNet, DenseNet, Xception, ConvNeXt and other architectures have autonomously learned hierarchical visual representations. These models however, mainly model the local spatial elements and are not good at modelling the long-range contextual dependencies well inside the retinal pictures. Besides, they also require large annotated datasets for training and accurate parameter adjusting.

Comparison of Vision Transformer and Capsule Network Approaches

ViT based approaches improve the performance of glaucoma classification by introducing the self-attention mechanism to acquire the global contextual information of the retinal fundus images. Capsule Networks also enhance the representation of the features by capturing the hierarchical spatial relationships among the different anatomical systems of the retina. The recent studies reveal that the fusion of Vision Transformer and Capsule Network will have the benefits of both networks, yielding better classification, better feature representation and higher robustness as compared to traditional CNN based approach.

Comparison of Segmentation and Localization Approaches

Some investigations further extend the diagnosis of glaucoma to the segmentation of the retinal anatomy and the localisation of the glaucoma region. The precise segmentation of optic disc and optic cup using encoder-decoder designs like U-Net, U-Net++ ResUNet and EfficientUNet helps us in calculating the CDR and hence determining the severity of glaucoma. In the same way, the use of YOLO-based object detection algorithms, effective localisation of the clinically relevant regions of the retina is provided. These techniques provide valuable diagnostic information but most of the existing literature studies segmentation and localisation separately and do not consider them for categorisation of illness.

Comparison of Explainable AI Frameworks

To overcome the opacity and clinical acceptance of automated glaucoma detection systems, XAI is becoming increasingly relevant. Visual heatmaps are obtained by methods like Grad-CAM and Grad-CAM++ which highlight specific retinal regions that drive the classification. These strategies make the model easier to understand and validate the automated predictions for physicians. Current explainability approaches, however, are mostly post-hoc and provide visual interpretation, and none of them enables a quantitative assessment of the quality of explanations or consistency in different clinical situations.

Comparative Summary

Table I summarises with respect to the main methodological focus, important methodologies, application goals and declared limitations representative works included in this review.

Table I. Comparative Summary of Surveyed Studies

Focus Area	Techniques	Application	Reported Limitations
CNN-Based Classification	ResNet, DenseNet, Xception, ConvNeXt	Automated glaucoma classification	Limited global contextual learning
Vision Transformer Approaches	Vision Transformer (ViT), Self-Attention	Global retinal feature extraction	Requires large-scale training data

Capsule Network Approaches	Capsule Network, Dynamic Routing	Spatial feature representation	Higher computational complexity
Retinal Structure Segmentation	U-Net, U-Net++, ResUNet, EfficientUNet	Optic disc and optic cup segmentation	Segmentation often evaluated independently
Glaucoma Localization	YOLO-based object detection	Localization of retinal abnormalities	Limited integration with complete diagnostic workflow
Explainable AI Frameworks	Grad-CAM, Grad-CAM++	Model interpretability and visualization	Limited quantitative evaluation of explanations
Integrated Diagnostic Frameworks	Classification + Segmentation + XAI	Comprehensive glaucoma diagnosis	Limited availability of unified end-to-end systems

From the comparative analysis, it can be concluded that the development of DL and medical image analysis has significantly contributed to the advancement of automated glaucoma diagnosis. Yet, the existing literature still focuses on how to improve the interpretability of models, combine multiple diagnostic tasks into one framework, and develop clinically confident AI systems that can make accurate predictions and be transparent in decision-making.

DATASETS AND EVALUATION METRICS

For reliable evaluation of automated glaucoma diagnosis algorithms, standard retinal imaging data sets and well-defined parameters of performance are needed. Previously released datasets help researchers train, develop, and benchmark their DL models in a fair experimental environment, and standardised evaluation metrics provide objective comparisons of different approaches for classification, segmentation, localisation and XAI. This review does not put forward new experiments but only summarizes the data and evaluation indexes commonly used in the existing literature.

A. Commonly Used Datasets

Glaucoma Classification Datasets

For the study of automated glaucoma classification, several publicly available retinal fundus image databases have been popularly used.

- **Multichannel Glaucoma Benchmark Dataset:** A big benchmark database of retinal fundus images and glaucoma labels along with retinal structural annotations for classification, segmentation and localisation.
- **REFUGE Dataset:** A publicly available dataset of retinal fundus images, annotated with glaucoma status and provided with optic disc/cup segmentation masks, which are frequently used in glaucoma screening studies.
- **RIM-ONE DL Dataset:** Offers retinal fundus photos taken from healthy and glaucoma patients for automated glaucoma classification and investigation of retinal anatomy.
- **ACRIMA Dataset:** A retinal imaging collection for the purpose of glaucoma classification using DL.
- **ORIGA Dataset:** Retinal fundus photos with annotations of optic disc and optic cup for glaucoma diagnosis and Cup-to-Disc Ratio study.

Retinal Structure Segmentation Datasets

Several benchmark datasets are available for optic disc and optic cup segmentation with pixel-wise ground-truth.

- SANCTUARY -
- ORIGA
- DRISHTI-GS
- RIM-ONE D/L

These datasets are frequently employed to test and assess the performances of models for retinal structure segmentation and glaucoma severity evaluation algorithms.

Localization Datasets

The studies of the localisation of Glaucoma generally rely on database of retinal images where the retinal structures or lesion location are manually identified. In some studies, the object detection annotations have been obtained from datasets in the retina and used for training and testing models based on YOLO for the localisation of clinically relevant areas of the retina.

B. Evaluation Metrics

The efficiency of the glaucoma diagnosis systems for different diagnostic tasks is assessed based on standardised assessment metrics that are adopted from the surveyed literature.

Glaucoma Classification

There are several metrics to assess glaucoma classification models:

- Precise
- Accurate
- Remember
- F1-Score
- ROC-AUC - Receiver Operating Characteristic - Area Under Curve
- Confusion Matrix

These measures offer a full evaluation of prediction accuracy, reliability and class discrimination ability in regard to the categorisation of disease.

Retinal Structure Segmentation

There are a number of common ways to evaluate segmentation models, including:

- Dice Coefficient
- IoU (Intersection over Union)
- Accurate
- Remember

The following metrics are employed for estimating the similarity between predicted optic disc and optic cup segmentation masks, and manually annotated ground truth masks to assess the accuracy of the optic disc and optic cup extraction.

Glaucoma Localization

Models for object detection for retinal area localisation are usually evaluated by:

- Accuracy
- Memoria

- Mean Average Precision (mAP@.5)

These measures investigate the accuracy of localisation of glaucoma-related areas of the retina, and the performance of object detection systems.

Explainable Artificial Intelligence

Typically, explainability methods are assessed through a qualitative visual interpretation where the heatmaps generated are compared to the actual clinical anatomical structures that are known to be associated with glaucoma. Some studies also evaluate the consistency of the explanation and its clinical usefulness by relying on expert judgement.

The evaluated datasets and evaluation parameters serve as a foundation for the evaluation of systems for automated glaucoma diagnosis. Given their wide use, these tools can also be used to objectively compare state of the art methods and lay a foundation towards the creation of more accurate, interpretable and clinically reliable DL frameworks for the categorisation of glaucoma.

CHALLENGES AND OPEN RESEARCH ISSUES

Although DL and medical image analysis have achieved great success, there are still some technical and clinical challenges [18] for automatic glaucoma diagnosis. Despite encouraging results reported in glaucoma classification, retinal structure segmentation, localisation and XAI, there are still many research difficulties to be tackled. To develop reliable, interpretable and practically useful glaucoma diagnosis methods, these issues need to be addressed.

A. Limited Model Interpretability

Many DL models have high diagnostic accuracy, but act as a black box; they give little insight as to why they make a diagnosis. A lack of transparency in the decision-making process hampers the trust of physicians and the adoption of AI in routine ophthalmological practice. The visual interpretation has been enhanced with the use of XAI (e.g. Grad-CAM), however, there is still a need for more powerful and clinically relevant models of explanation.

B. Preservation of Global and Spatial Features

Most of the traditional CNNs are mainly based on local feature extraction, and they are not enough to model the long-range contextual relationships in the retinal fundus images. But, while models based on transformer can learn the global contextual information, they may not retain the fine-grained spatial correlations among the retinal anatomical structures. It is still an important research topic to design an architecture that can take advantage of the global and geographical information simultaneously.

C. Accurate Retinal Structure Analysis

Reliable segmentation of optic disc and optic cup is crucial for accurate CDR computation and glaucoma severity evaluation. Segmentation of retinal structures, however, is a challenging task because of the variation in retinal anatomy, image quality and illumination conditions and disease characteristics. Research into the segmentation of robust in clinical data sets is ongoing.

D. Limited Integration of Diagnostic Tasks

In the majority of the previous studies glaucoma categorisation, retinal structure segmentation, localisation and XAI have been investigated as distinct research problems. Typically they are stand-alone systems that do not offer full diagnostic support. Research efforts that lead to the development of integrated frameworks which can incorporate a multitude of diagnostic tasks into one intelligent system are still a major avenue for future research.

E. Dataset Availability and Generalization

Visualization of the vast, diversified and well annotated retinal images is essential for the performance of DL models. Current benchmark datasets have different image acquisition protocols, differ in annotation quality, and are not necessarily representative of the demography, potentially reducing model generalisation to other clinical settings. Enabling the creation of larger multi-center datasets and generalisation across datasets is still key.

F. Clinical Adoption and Real-World Deployment

While a number of DL models have shown good results in experiments, few studies have explored the practical application in a clinical environment. In practice, applicability requires good performance, computational efficiency, user-friendliness, intelligible predictions and compatibility with current healthcare systems. The challenge of connecting research prototypes and clinical applications is an important open research question.

Based on the literature studied, it is generally recommended that future glaucoma diagnosis systems should be designed to be more interpretable, have better representation of features, combine multiple diagnostic tasks, have better generalisation of the model and for ease of deployment in the clinical setting. This will help overcome these challenges and make the development of more accurate, more precise and more clinically accepted computer-aided glaucoma diagnosis frameworks possible.

FUTURE RESEARCH DIRECTIONS

The field of AI, computer vision and medical image analysis is rapidly evolving and offers new opportunities for improved automated diagnosis of glaucoma. Even though there are some potential research fields to be explored, the current studies have made great progress in the field of disease classification, retinal structure analysis and XAI. Future research should be directed towards developing more interpretable, accurate, scalable and clinically applicable glaucoma diagnosis frameworks that will help clinicians in real healthcare environment.

A. Advanced Vision Transformer Architectures

In the future, more efficient topologies of transformer can be explored to reduce the computational complexity and enhance the representation of features, thus to train and deploy the model in a resource-limited clinical setting.

B. Hybrid Deep Learning Models

The combination of several DL architectures has great potential in improving the diagnosis of glaucoma. Future research could explore hybrid architectures such as the combination of a transformer network with Capsule Networks, convolutional neural networks and attention mechanisms, which aim to jointly learn both global contextual information and preserve fine-grained spatial connections in the retinal images.

C. Multi-Task Learning Frameworks

Most existing works classify, segment structure of retina, estimate and localise the severity of the glaucoma separately. In the future, a multi-task learning framework that can accomplish several diagnostic tasks simultaneously should be developed, which can reduce the computing time and enhance the comprehensive clinical decision support.

D. Explainable and Trustworthy Artificial Intelligence

With the increasing integration of AI into clinical practice, research into improving the models' transparency and trustworthiness will remain a critical aspect. Future research should focus on more advanced explainability approaches that supply more pertinent visual and feature-level explanations, enabling physicians to assess and trust automated diagnostic results.

E. Large-Scale and Multi-Center Clinical Validation

The publicly available benchmark datasets are used to evaluate most of the existing works. Future research is needed to confirm glaucoma diagnostic models in larger multi-center clinical datasets from diverse populations, imaging modalities, and health care centers to enhance the generalisation, robustness, and applicability of the model.

F. Intelligent Clinical Decision Support Systems

Future glaucoma diagnosis frameworks can be improved other than predicting the disease with the inclusion of intelligent clinical decision-support functionality. Automated diagnosis and severity evaluation, patient monitoring, disease progression analysis and comprehensive diagnostic report can lead to better support for ophthalmologists and patient management.

Future research should be in the direction of development of integrated, interpretable and clinically reliable systems for diagnosis of glaucoma by combining advanced DL architectures, XAI, comprehensive retinal image analysis and large-scale clinical validation. These advances would result in improved, trusted and feasible computer-aided glaucoma screening for contemporary eye care.

DISCUSSION AND CONCLUSION

To conclude, the proposed methodology meets the important demand for reliable, interpretable and automated glaucoma diagnosis by using retinal fundus imaging for prompt clinical decision making. It is a publicly available retinal fundus image dataset and compares the classification of glaucoma using various state-of-the-art DL architectures including ResNet50, DenseNet201, ViT-Base, ViT-CapsNet, Xception and ConvNeXtTiny. Also, EfficientUNet is used to segment the optic disc and optic cup to compute the CDR to estimate the severity of glaucoma. YOLO based object identification models are used to detect the glaucoma related retinal areas. XAI techniques, namely Grad-CAM and Grad-CAM++, are used to highlight the salient parts of the image, thus improving the transparency of the prediction. The experimental evaluation showed the efficiency of the framework with DenseNet201 achieving the maximum classification accuracy of 89.60%, indicating its competence to reliably identify glaucoma. The framework is expanded to cover explainable visualisation, retinal structural analysis, glaucoma localisation, and a Flask based web application, which goes beyond mere categorisation to provide full support for diagnosis and automated generation of a clinical report. To summarise, the created approach offers a promising and interpretable computer-aided diagnostic solution that promotes screening efficiency, diagnostic consistency, and helps ophthalmologists make educated clinical decisions for glaucoma therapy.

A. Trends and Key Insights

There is a clear shift in the literature from using conventional CNN towards exploring more complex models like ViT and hybrid DL (HYT) models. As researchers continue to integrate retinal structure segmentation, CDR analysis and XAI to enhance the performance of diagnosis and clinical interpretability. It has also become possible to fairly compare alternative techniques due to the increasing use of publicly available benchmark datasets and standardised evaluation measures. Further, integrated diagnostic system where many retinal image analysis tasks are included, is becoming a potential pathway towards developing more complete glaucoma screening systems.

B. Implications and Future Perspectives

The reviewed research show that the future of glaucoma diagnosis systems must be high in classification accuracy as well as interpretability, resilience and clinical application. Future research should be directed towards developing integrated solutions that include glaucoma categorisation, retinal structure analysis, severity estimation, localisation and XAI. In clinical practice, this combination of transparent AI techniques and full retinal image analysis can boost the confidence of clinicians, aid early glaucoma detection, and allow efficient large-scale screening campaigns.

C. Summary

The report offers an extensive survey of the latest advances in glaucoma classification using retinal fundus images using interpretable DL techniques. The existing studies were categorised based on their methodological approaches and classification techniques, retinal structure segmentation techniques, localisation models, XAI frameworks, commonly used datasets and evaluation metrics were comparatively analysed. The study also revealed important research questions: interpretability, incompleteness of diagnostic procedures, generalisation of the data sets and clinical application. This survey summarises the existing research trends and identifies future directions for academics and practitioners in the goal of building an accurate, interpretable and clinically reliable computer-aided glaucoma diagnosis system.

REFERENCES

[1] V. Vigneshwaran, S. Kumar E., and R. Shoba, "Automated Glaucoma Detection Using Vision Transformers and Capsule Networks: A Deep Learning Approach for Early Diagnosis from Retinal Fundus Images," in *Proceedings of the 2025 International Conference on Data Science, Agents and Artificial Intelligence (ICDSAAI)*, Chennai, India, 2025.

[2] H. Wang, H. Sun, Y. Fang, S. Li, M. Feng, and R. Wang, "A Workflow for Computer-Aided Diagnosis of Glaucoma," in *Proceedings of the 2022 IEEE International Symposium on Biomedical Imaging Challenges (ISBIC)*, pp. 1–4, 2022.

[3] Dr.B.Sateesh Kumar, Detection and Classification of Objects in Satellite Images using Custom CNN, ZKG International, ISBN No.2366-1313, Vol No.VIII, Issue No.II, pp.78-86, Web of Science Group, JULY, 2023.

[4] R. Hemelings, B. Elen, A. K. Schuster, M. B. Blaschko, N. Beare, A. Birhanu, M. Bogunović, C. Bowd, J. De Fauw, K. Drexler, *et al.*, "A Generalizable Deep Learning Regression Model for Automated Glaucoma Screening from Fundus Images," *npj Digital Medicine*, vol. 6, no. 1, p. 112, 2023.

[5] Y. Madadi, H. Abu-Serhan, and S. Yousefi, "Domain Adaptation-Based Deep Learning Model for Forecasting and Diagnosis of Glaucoma Disease," *Biomedical Signal Processing and Control*, vol. 92, p. 106061, 2024.

[6] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An Image Is Worth 16×16 Words: Transformers for Image Recognition at Scale," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2021.

[7] Dr.B.Sateesh Kumar, A Benchmark Survey on High Resolution Remote Sensing Images Learning Techniques, International Journal of Techno Engineering, IJTE, ISBN No.2057-5688, Vol No.XV, Issue No.III, pp.23-32, IJTE, JULY, 2023.

[8] R. Hussain and H. Basak, "UT-Net: Combining U-Net and Transformer for Joint Optic Disc and Cup Segmentation and Glaucoma Detection," *arXiv preprint arXiv:2303.04939*, 2023.

[9] S. Virbukaitė, "Glaucoma Identification Using Convolutional Neural Networks and Cup-to-Disc Ratio Analysis," *IEEE Access*, vol. 12, pp. 35841–35855, 2024.

[10] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 618–626, 2017.

[11] A. Chattopadhay, A. Sarkar, P. Howlader, and V. N. Balasubramanian, "Grad-CAM++: Improved Visual Explanations for Deep Convolutional Networks," in *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 839–847, 2018.

[12] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLO," Ultralytics, 2023.

[13] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7263–7271, 2017.

[14] H. G. Lemij, C. de Vente, C. I. Sánchez, and K. A. Vermeer, "Characteristics of a Large, Labeled Dataset for the Training of Artificial Intelligence for Glaucoma Screening with Fundus Photographs," *Ophthalmology Science*, vol. 3, no. 3, p. 100300, 2023.

[15] A. Septiarini and A. Harjoko, "Automatic Glaucoma Detection Based on the Type of Features Used: A Review," *Journal of Theoretical and Applied Information Technology*, vol. 72, no. 3, pp. 314–328, 2015.

[16] E. Arrieta-Rodriguez, D. Morales, J. M. Molina, and J. Hernández, "A Comprehensive Review on Fundus Imaging Approaches for Glaucoma Detection," *IEEE Reviews in Biomedical Engineering*, vol. 18, pp. 154–178, 2025.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[18] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4700–4708, 2017.

[19] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11976–11986, 2022.

[20] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1251–1258, 2017.