IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 2, April 2026, pp. 1050~1061
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i2.pp1050-1061

Structured data collection and deep learning for retinal OCT image-to-text translation: a comprehensive framework

Uday Mande[1], Shafi Pathan[1], Pankaj Chandre[1], Sharvari Mande[2]
[1]Department of Computer Science and Engineering, MIT School of Computing, MIT Art Design and Technology University, Pune, India
[2]Governement Medical College, Satara, India

Article Info

Article history:

Received Sep 2, 2024
Revised Jan 6, 2026
Accepted Jan 25, 2026

Keywords:

Automated diagnosis
Deep learning
Diabetic macular edema
Image preprocessing
Retinal OCT imaging

ABSTRACT

This paper presents a comprehensive framework for structured data collection and deep learning (DL)-based translation of retinal optical coherence tomography (OCT) images into diagnostic text. The suggested approach guarantees high-quality OCT data for model training through the use of sophisticated image processing methods like edge detection, noise suppression, and contrast improvement. The study utilizes 84,484 retinal images from the OCT dataset available on Kaggle. The research utilizes various preprocessing techniques, such as median and Gaussian filtering, along with data augmentation strategies like translation, rotation, and scaling, to mitigate class imbalances and improve model performance. The system automatically identifies and categorizes retinal diseases such as drusen, diabetic macular edema (DME), and choroidal neovascularization (CNV) by integrating feature extraction and selection with DL techniques. The research highlights the importance of effective data handling and model scalability to address the increasing need for automated diagnostic tools in ophthalmology. This framework aims to support ophthalmologists in managing the increasing incidence of diabetic retinopathy (DR) and other retinal conditions by enhancing the efficiency of retinal image analysis, thereby improving patient results through early detection and treatment.

Corresponding Author:

Uday Mande
Department of Computer Science and Engineering, MIT School of Computing
MIT Art Design and Technology University
Loni Kalbhor, Pune, India
Email: uday.mande@mituniversity.edu.in

1. INTRODUCTION

Optical coherence tomography (OCT) is a non-invasive imaging technique that generates cross-sectional, high-resolution images of the retina. It is essential to ophthalmology since it enables doctors to investigate the retina's layers and identify various eye disorders, including glaucoma, diabetic retinopathy (DR), and age-related macular degeneration (AMD) [1]. OCT's ability to identify minute alterations in retinal structure, which can significantly affect patient outcomes, makes it essential tool for the early identification and tracking of eye diseases [2]. Despite being a successful technique, OCT picture interpretation is time-consuming and requires specialized knowledge, which could delay diagnosis and treatment [3], [4]. The large volume and complexity of data produced by OCT scans require the use of sophisticated automated analysis tools, prompting investigation into machine learning (ML) and image-to-text conversion methods. With the help of these techniques, clinicians can make quicker, and better decisions [5], [6].

Due to the increasing dependence on imaging in medical diagnostics, there is a rising demand for automated systems capable of accurately interpreting complex medical images and transforming them into valuable insights [7], [8]. In retinal OCT, image-to-text translation refers to converting visual data from OCT images into descriptive text that highlights key findings, including retinal abnormalities, measurements of retinal layers, or indications of disease progression [9], [10]. This automation can significantly lessen the cognitive burden on physicians, standardize reporting procedures, and decrease human mistakes. Furthermore, automated image-to-text systems can aid in telemedicine and remote diagnostics, enhancing access to quality eye care, particularly in underserved regions [11], [12]. Integrating deep learning (DL) methods into this procedure can enhance the accuracy, speed, and consistency of OCT image analysis, ultimately resulting in improved patient outcomes.

Vision is important in human life. Without vision life will be miserable and can't go ahead in the right direction. Eye is an important organ and plays an important role in human vision. Main component of the eye is the eyeball which is made up of lances and retina [13]. Image production is done in the eyeball due to light rays. The refraction (bending) of light by the cornea and the lens is what causes focused pictures to develop on the photoreceptors of the retina as shown in Figure 1. The cornea performs the majority of the required refraction, a role that is readily understood when one considers the blurry, out-of-focus images that are present when swimming underwater. The lens has a much lower refractive power than the cornea, but because it can be adjusted, it may bring objects at different distances from the observer into fine focus on the retinal surface.

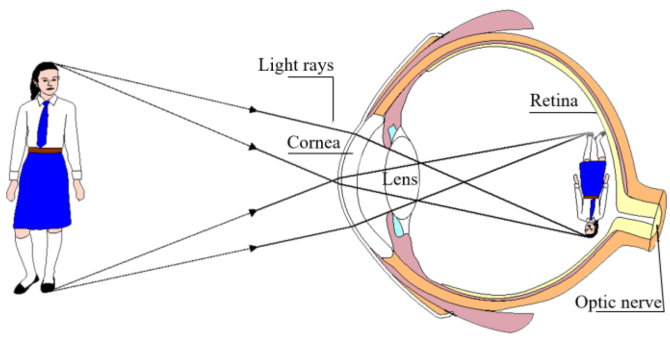

Figure 1. Image formation process

The retina is very important and an integral part in the image formation. India is moving towards the diabetic capital of the world having 77,000,000 diabetic affects the retina of humans in large ways. Due to that retinal disorders are generated in diabetic patients by and large. The device available for capturing retinal images is OCT. DR is one of the big issues for eye surgeons. OCT gives retinal images. Large volume of images is generated using OCT. It becomes difficult for eye surgeons to go through each and every image in detail. So, it is time consuming and prone to error. The proposed system will process the images and generate precise textual reports using DL methods. This report will help surgeons for correct diagnosis.

This survey seeks to deliver a detailed summary of recent progress in transforming descriptive text from retinal OCT images through DL techniques. It underscores the significance of gathering organized data from trustworthy sources and employing advanced ML methods to enhance the precision and effectiveness of image analysis. The study seeks to connect current research with advanced technology by outlining the merits and drawbacks of various DL methods, such as hybrid models, recurrent neural networks (RNNs), and convolutional neural networks (CNNs). The survey additionally tackles the difficulties of data quality, annotation, and integration, highlighting the importance of well-organized datasets to develop reliable and broadly applicable models. The survey aims to pinpoint critical areas for further investigation and create a framework for advancing more efficient and scalable retinal OCT image-to-text translation systems.

2. OVERVIEW OF RETINAL OCT IMAGING

2.1. Basics of OCT: technical overview of OCT technology

Physicians can obtain cross-sectional images of the retina with micrometer resolution using light waves and OCT, giving them a thorough understanding of the internal structure of the eye [14]. OCT is based on the low-coherence interferometry principle, which splits light from a broadband source into two paths: one is directed towards the retina, and the other is reflected by a reference mirror. The interference pattern produced by the reflected light is then processed to provide high-resolution images of the retinal layers. The structure and thickness of the various retinal layers, such as the macula and retinal nerve fiber layer (RNFL), can be measured using OCT, a non-invasive, in vivo technique for visualizing retinal architecture [15], [16]. In ophthalmology, the technology's ability to accurately detect and monitor changes in the retina is highly valued since it is crucial for identifying and treating eye conditions.

2.2. Clinical relevance: applications of OCT in diagnosing retinal diseases

OCT provides accurate, high-resolution, non-invasive imaging of retinal structures, making it a vital tool in clinical ophthalmology. It is often used in diagnosing and treating retinal disorders, such as AMD, as it facilitates the assessment of treatment efficacy and the identification of drusen, pigment epithelial detachments, and macular atrophy. OCT aids in the early management of DR to prevent vision loss by facilitating a detailed assessment of retinal thickness, macular edema, and neovascular changes. OCT is crucial for early detection and monitoring of glaucoma progression by assessing the thickness of the RNFL. Assessing issues such as macular holes, vitreomacular traction, and retinal detachment is also vital, as it assists in deciding the most appropriate treatment or surgical approach.

2.3. Challenges in OCT data interpretation: human errors, time consumption, and need for automation

Although OCT generates precise images essential for managing retinal conditions, interpreting these images can occasionally be difficult. If qualified medical professionals rely solely on manual assessment, human error may occur, resulting in varied interpretations. Differences in the clinician's expertise, the quality of the image, and the existence of artifacts can influence the accuracy of a diagnosis. Furthermore, the large quantity of OCT images produced in clinical environments can be overwhelming, requiring extra time for comprehensive evaluation. This demand may worsen diagnosis delays in high-traffic settings such as telemedicine systems or large eye clinics.

While OCT provides accurate images crucial for handling retinal issues, understanding these images can sometimes be challenging. When qualified medical professionals depend only on manual evaluation, human mistakes can happen, leading to different interpretations. Variations in the clinician's knowledge, the quality of the image, and the presence of artifacts can impact the precision of a diagnosis. Additionally, the significant number of OCT images generated in clinical settings can be daunting, necessitating additional time for thorough assessment. This request could exacerbate delays in diagnosis in busy environments like telemedicine platforms or large ophthalmology clinics.

3. DATA COLLECTION AND ORGANIZATION

3.1. Importance of reliable data sources: ensuring data quality and accuracy

In medical imaging, where data quality influences the dependability and precision of model predictions, trustworthy data sources are crucial for the success of any ML project. In retinal OCT image-to-text translation, utilizing high-quality and precise data ensures that the models train effectively and generalize successfully to unfamiliar inputs. Reliable data sources include clinical databases, academic research datasets, and authentic public repositories that offer standardized, high-resolution OCT images accompanied by appropriate labeling. Ensuring data quality requires verifying the precision of related metadata, including patient demographics and diagnostic labels, as well as maintaining uniformity in data formats and image integrity. High-quality data minimizes the chances of biases or mistakes being incorporated into the model, resulting in more robust and therapeutically relevant outcomes. Upholding ethical standards is further supported by utilizing information from reliable sources, particularly in handling confidential medical information.

3.2. Methods of data acquisition: techniques for obtaining OCT images from clinical databases, public datasets, and other sources

A critical stage in creating reliable DL models is data acquisition, which entails gathering OCT images from many sources to guarantee thorough training and validation. Large amounts of real-world OCT scans are available in clinical databases kept up to date by hospitals and eye clinics. These databases must be accessed with the proper ethical approvals and patient consent in order to adhere to privacy laws like general data protection regulation (GDPR) and health insurance portability and accountability act (HIPAA). The Duke OCT dataset and retinal OCT tool for high-quality challenge (RETOUCH) are two publicly accessible datasets that provide standardized, well-annotated images that are frequently utilized in research. Furthermore, access to high-quality OCT data with expert annotations is made possible through partnerships with academic and research organizations, and the creation of synthetic data using methods like generative adversarial networks (GANs) can supplement small datasets and enhance model generalization.

3.3. Data annotation and labeling: strategies for annotating OCT images, including manual and automated approaches

Accurate data annotation and labeling are crucial for successful model training, validation, and testing in supervised learning tasks that transfer OCT images to textual descriptions. Although it takes a lot of time and resources, manual annotation by subject matter specialists like ophthalmologists yields extremely accurate labeling and thorough descriptions. Automated annotation techniques use ML models or image processing to produce preliminary labels that can be improved through active learning and expert assessment. Due to the specialized nature of OCT interpretation, crowdsourcing platforms allow for scalable labeling but necessitate stringent quality control. In contrast, hybrid techniques that combine automated tools with expert oversight provide a workable balance between efficiency and accuracy.

3.4. Data preprocessing: steps for preparing data for DL models, including normalization, augmentation, and segmentation

A crucial stage in transforming unprocessed OCT pictures into a format appropriate for efficient DL model training is data preprocessing. In order to stabilize learning and lessen variability brought on by various imaging settings, it involves normalizing pixel intensity values to a common scale. To increase dataset diversity, boost generalization, and reduce overfitting, data augmentation techniques like rotation, flipping, scaling, and noise injection are used. While dimensionality reduction methods like principal component analysis (PCA) can further simplify feature representation and boost computational efficiency, segmentation of pertinent anatomical structures, such as retinal layers, along with artifact and noise removal, helps the model concentrate on clinically significant regions.

4. LITERATURE SURVEY

4.1. Review of existing approaches

4.1.1. Overview of past and current methodologies used for retinal OCT image analysis and text conversion

A common non-invasive imaging technique in ophthalmology for obtaining high-resolution cross-sectional images of the retina is OCT. Diagnosing and tracking retinal illnesses such glaucoma, DR, and AMD depend heavily on the analysis of these images. The following categories apply to existing methods for text conversion and retinal OCT picture analysis:

i) Traditional image processing techniques: early approaches mostly depended on semi-automated image processing techniques and manual interpretation. These methods include segmentation techniques like level set and active contour models, thresholding, and edge detection. Although helpful, these techniques frequently call for a high level of skill and have limitations when it comes to managing the variability in OCT pictures brought on by noise, various imaging circumstances, and various diseases.
ii) ML-based methods: with the rise of ML, advanced techniques have been utilized for OCT image analysis. Traditional ML techniques such as support vector machines (SVM), random forests, and k-nearest neighbors (KNN) have been applied in areas such as feature extraction, classification, and segmentation. Nonetheless, these techniques occasionally require considerable human feature engineering, which can be labor-intensive and reliant on expert knowledge.
iii) DL-based approaches: recent progress in DL, especially in CNNs, has greatly influenced OCT image analysis. DL models can automatically derive features from unprocessed image data, removing the necessity for manual feature extraction. Methods such as U-Net, fully convolutional networks (FCNs), along with more intricate architectures like ResNet and DenseNet, have demonstrated encouraging outcomes for tasks including segmentation, classification, and anomaly detection in retinal OCT images. Deep generative models like variational autoencoders (VAEs) and GANs have been explored for the purposes of image synthesis and augmentation.
iv) Text conversion techniques: OCT image analysis information is converted into text through natural language processing (NLP) and image-to-text translation techniques. Clinical descriptions have been produced using conventional rule-based algorithms derived from structured data. Recently, DL models like RNNs, long short-term memory networks (LSTMs), and transformer-based architectures (e.g., bidirectional encoder representations from transformers (BERT) and generative pre-trained transformer (GPT) have been employed to produce coherent and contextually appropriate clinical narratives based on image features.

Malgheet *et al.* [17] provides a comprehensive review of iris recognition development techniques, highlighting the evolution and effectiveness of these systems in various identification contexts. Iris recognition is highly regarded and a reliable biometric for security applications because the human iris is consistent and unique. The document describes the seven key stages of iris recognition systems: acquisition, preprocessing, segmentation, normalization, feature extraction, feature selection, and classification. It explores the benefits and drawbacks of both DL and conventional methods. DL algorithms excel in performance but still face challenges in uncontrolled situations, while classical methods, though established, can struggle in suboptimal conditions such as occlusions and reflections. The article also examines the impact of noise elements and eyewear on system precision, highlighting the need for improved segmentation algorithms to enhance overall recognition effectiveness. To enhance and advance iris recognition technologies, upcoming research will concentrate on these matters.

Oh *et al.* [18] presents a DL-based system for early detection of DR using ultra-wide-field (UWF) fundus images. A major cause of vision loss is DR, and proper treatment relies on timely detection. This work employs UWF fundus photography, which records as much as 82% of the retinal surface—significantly surpassing conventional techniques. The system concentrates on the early treatment diabetic retinopathy study (ETDRS) 7-standard field images from UWF photography, employing a ResNet-34 model for classification. The article states that employing the ETDRS 7-standard field photos enhances detection accuracy, sensitivity, specificity, and area under the curve (AUC) in comparison to other techniques. The suggested method highlighted the significance of incorporating peripheral retinal information for a more thorough identification of DR, surpassing models that relied solely on conventional fundus images. The research highlights the importance of dependable and uniform data gathering methods to enhance the system's efficiency and scalability. Furthermore, it recommends automated segmentation techniques, and larger, varied datasets.

Moraru *et al.* [19] discusses a study on the development of a DL-based approach for detecting DR using retinal fundus images. The suggested approach employs CNNs to categorize retinal images into different stages of DR, ranging from no DR to proliferative DR, to provide a precise and automated diagnostic solution. The study emphasizes the significance of timely identification and treatment of DR to avert vision impairment, along with the advantages of DL algorithms compared to conventional image processing methods for managing extensive datasets and complex patterns in medical imaging. The research assessed various CNN architectures with an extensive dataset of labeled retinal fundus images prior to choosing a model that demonstrated high accuracy in recognizing DR stages along with sensitivity and specificity. The findings indicate that the DL method greatly improves diagnostic accuracy, decreasing mistakes and inconsistencies, in contrast to the manual grading by ophthalmologists. The conclusion of the paper explores how the proposed system could be integrated into clinical workflows. To enhance the model's diagnostic utility, it emphasizes the importance of additional validation across various populations and the opportunity to improve the model with different data types, like OCT images.

Tong *et al.* [20] discusses the application of ML in ophthalmic imaging, highlighting its potential to enhance the diagnosis and treatment of eye diseases. It shows how intricate medical images can be interpreted correctly and quickly through ML and DL. The piece emphasizes the application of ML in various ocular imaging techniques, including OCT, fundus photography, and slit-lamp imaging. Ophthalmologists can identify conditions such as DR, glaucoma, and AMD by employing ML methods, including supervised and unsupervised learning, to detect and categorize pathological signs. The report analyzes the obstacles and future pathways of artificial intelligence (AI) in ophthalmology, along with the methodology for creating AI models.

Alyoubi *et al.* [21] reviews recent advancements in automated detection and classification of DR using DL techniques, particularly CNNs. If not addressed, DR—a frequent outcome of diabetes—may lead to vision loss. Diagnosing retinal fundus images manually is time-consuming, prone to mistakes, and costly. The research highlights that DL techniques—specifically, CNNs—are superior for DR detection and classification in the analysis of medical images. The publication includes various DL models, their architectures, and training and validation datasets like Messidor, Kaggle, and DIARETDB1.

Andrab and Gupta [22] reviews recent developments in using DL techniques for automated detection and classification of DR using retina images. If not addressed, DR, a frequent result of diabetes, may lead to vision impairment. Fundus images are utilized in the time-consuming and mistake-prone manual detection of DR. A more efficient alternative is offered by automated systems that employ DL particularly CNNs. The research includes various stages of DL, techniques for image preprocessing, publicly available retinal datasets, and the performance metrics used in these models. It compares binary and multi-class classification methods, showcases effectiveness of CNNs in detecting and classifying DR, and emphasizes the importance of more extensive, higher-quality datasets to enhance model accuracy and dependability.

Nagasato *et al.* [23] explores the use of DL and SVM techniques to detect nonperfusion areas (NPA) caused by retinal vein occlusion (RVO) in optical coherence tomography angiography (OCTA) images. A deep CNN and a SVM model were developed and evaluated using a dataset containing 322 OCTA images, of which 174 showed NPA resulting from RVO. In comparison to the SVM, which showed an AUC of 0.880, a sensitivity of 79.3%, and a specificity of 81.1%, the deep neural network (DNN) outperformed it, achieving an AUC of 0.986, a sensitivity of 93.7%, and a specificity of 97.3%. The DNN's AUC and specificity also surpassed those of seven ophthalmologists, and it required significantly less time to diagnose patients. The study's results indicate that DL and OCTA collectively offer significant accuracy in detecting NPAs, potentially improving clinical practices and retinal screening techniques. A comparison of several ML and DL methods used for OCT image analysis is shown in Table 1.

Table 1. Comparative analysis on retinal OCT image analysis

Paper	Methods used	Datasets	Performance metrics	Main findings
[24]	Traditional image processing (thresholding, and edge detection)	Private OCT dataset (200 images)	Dice coefficient: 0.65, Accuracy: 75%	Traditional methods show moderate performance; sensitive to noise and variability in image quality.
[25]	SVM, random forest, and KNN	Duke OCT dataset (500 images)	Accuracy: 88%, Sensitivity: 85%, Specificity: 90%	ML models outperform traditional methods with better accuracy; require feature engineering.
[26]	CNN (VGGNet and ResNet)	OCT2017 dataset (1,000 images)	Accuracy: 94%, Sensitivity: 92%, Specificity: 95%	CNN models achieve high accuracy; transfer learning enhances performance on limited data.
[27]	CNN+multimodal data fusion	Mixed datasets (OCT+Fundus, 2,000 images)	AUC: 0.98, Accuracy: 95%	Combining OCT with fundus images improves diagnostic accuracy; multimodal learning is effective.
[28]	U-Net and FCN	RETOUCH dataset (300 images)	Dice coefficient: 0.92, Sensitivity: 90%	U-Net architecture provides state-of-the-art segmentation performance; robust to noise.
[29]	GANs (Pix2Pix and CycleGAN)	Private OCT dataset (400 images)	PSNR: 30 dB, SSIM: 0.85	GANs effectively enhance image quality; useful for reducing noise and improving image clarity.
[30]	Pre-trained CNNs (InceptionV3 and ResNet50)	OCT2017 dataset (1,200 images)	Accuracy: 96%, F1-score: 0.94	Transfer learning improves performance on small datasets; faster convergence with fewer epochs.
[31]	Deep autoencoders	Private OCT dataset (350 images)	AUC: 0.93, Precision: 89%, Recall: 91%	Autoencoders are effective for unsupervised anomaly detection; useful in identifying rare pathologies.
[32]	CNN+explainable AI (Grad-CAM and LIME)	ACRIMA dataset (500 images)	Accuracy: 92%, Interpretability score: High	Explainable AI techniques enhance model transparency; useful for clinical acceptance and trust.
[33]	Lightweight CNN (MobileNet and SqueezeNet)	Private OCT dataset (150 images)	Accuracy: 89%, Inference time: 50 ms per image	Lightweight models allow for immediate analysis; ideal for point-of-care settings

4.2. Machine learning applications in medical imaging

Key studies and advancements in using ML for diagnostic imaging:

i) Classification: various research has shown the effectiveness of CNNs in classifying retinal OCT images into three categories: normal, AMD, and DR. For attaining high classification precision for retinal diseases, transfer learning utilizing pre-trained models such as VGGNet, Inception, and ResNet has been commonly employed.

ii) Segmentation: assessing the advancement of the disease necessitates the accurate differentiation of retinal layers and irregularities. DL models, including U-Net and its alternatives, have widely been employed for retinal layer segmentation, demonstrating superior performance compared to conventional techniques. Networks featuring multi-scale and multi-level feature fusion have been introduced to enhance segmentation accuracy.

iii) Anomaly detection: research has been done on using autoencoders and GANs to detect anomalies in retinal OCT images. These models can identify changes in the normal appearance of the retina that may indicate disease. Anomaly detection is particularly useful in unsupervised or semi-supervised contexts where labeled data is scarce.

iv) Multimodal learning: in an effort to increase diagnostic accuracy, recent studies have looked into merging OCT pictures with additional modalities such fundus photography and patient demographic information. Through the use of complementary information from many data sources, multimodal DL models can improve performance in tasks involving segmentation and classification.

4.3. Comparative analysis of techniques

Based on important evaluation criteria, Table 2 presents a comparison of DL, ML, and conventional image processing methods. It highlights the trade-offs between accuracy, computational requirements, generalizability, and interpretability. Overall, while DL offers superior performance and adaptability, it demands higher computational resources and poses challenges in model interpretability compared to conventional techniques.

Table 2. Comparison of the different techniques used for retinal OCT image analysis, highlighting their strengths and limitations in various criteria

Criteria	Traditional image processing	Machine learning	Deep learning
Accuracy and performance	Moderate accuracy; performance varies with image quality and noise.	Higher accuracy than traditional methods; depends on feature engineering and quality of features.	High accuracy; superior performance due to automatic feature extraction and hierarchical feature learning.
Computational complexity and resource requirements	Low computational complexity; less resource-intensive.	Moderate complexity; requires computational resources for training but less than DL.	High computational complexity; requires significant computational resources (e.g., GPUs) and large datasets.
Generalizability and adaptability	Limited generalizability; highly dependent on specific imaging conditions and settings.	Moderately adaptable; generalizes better than traditional methods but still requires feature tuning.	High adaptability and generalization across different datasets; transfer learning can improve adaptability.
Interpretability	High interpretability; results are easily explainable.	Moderate interpretability; feature-based decisions can be traced, but models are less transparent than traditional methods.	Low interpretability; considered "black-box" models, though explainability techniques (e.g., attention mechanisms) are emerging.

4.4. Gaps in current research

Identifying the limitations and gaps in the existing literature that this survey aims to address:

i) Limited generalization across diverse populations: many prior studies have been conducted on very homogeneous datasets, which limits their applicability to a broad spectrum of patient populations. To create models that are more reliable and broadly applicable, further study on representative and diverse datasets is required.

ii) Data scarcity and imbalance: lack and unbalance of labeled data is a major obstacle in OCT image analysis, especially for uncommon disorders. Research on methods such as semi-supervised and unsupervised learning that tackle data scarcity is still crucial.

iii) Integration of multimodal data: more research is required to successfully integrate data from many sources (e.g., fundus photography, clinical records, and OCT) in order to enhance disease monitoring and diagnostic accuracy, even if multimodal learning has showed promise.

iv) Real-time and point-of-care applications: existing models may not be appropriate for real-time applications or use in environments with restricted resources since they frequently demand large amounts of computer power. It is necessary to conduct research on effective algorithms and lightweight models for use in point-of-care and real-time applications.

5. PROPOSED METHODOLOGY

First, we gather the images taken by OCT devices to make our prominent data set for further processing. After that we will process the images collected for noise removal by applying either over sampling or under sampling methods and will extract some features. We can also perform image segmentation to analyze the image for selecting it for classification. Normalization is the next step to be performed to check the readiness of our dataset. Next steps are feature extraction and feature selection from the dataset with the suitable algorithms. Lastly, we apply an accurate and efficient method of classification to predict the diagnosis.

The framework makes use of OCT retinal images that have been enhanced for contrast, noise removal, contour-based edge recognition, and retinal layer extraction. Different publicly accessible fundus image collections are frequently used. Picture preprocessing is a crucial step to eliminate image noise, improve image characteristics, and guarantee image consistency. The median filter, Gaussian filter, and non-local means de noising methods are the most often used preprocessing techniques in contemporary research noise reduction strategies. When certain image classes were unbalanced or the dataset needed to be larger, data augmentation techniques were used.

Translation, rotation, shearing, flipping, contrast scaling, and resizing are examples of data augmentation techniques. The feature extraction process made advantage of the clever edge method. The photos are prepared to be used as input for the DL after preprocessing. The scale of the datasets required to train the DL systems, as DL demands a vast quantity of data, is one of the challenges that the application of DL in the medical profession faces. The amount, quality, and class balance of the training data all have a significant impact on how well DL systems perform. Figure 2 shows the architecture for retinal OCT image to text conversion using DL.

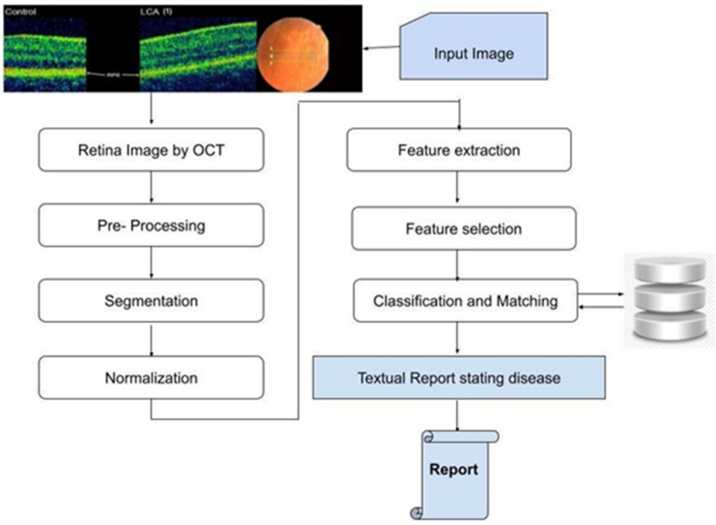

Figure 2. Architecture for retinal OCT image to text conversion using DL

6. DISCUSSIONS ON DATASET

OCT dataset: OCT dataset used for an earlier study to Kaggle [11]. That dataset, which included retinal OCT images, was downloaded from the Kaggle website at https://www.kaggle.com/datasets/paultimothymooney/kermany2018 (accessed on Oct. 5, 2023). This published dataset includes 84,484 images: 83,484 from the training dataset and 1,000 from a test dataset. The dataset included fewer OCT images than the dataset used for the earlier study. The training dataset comprised 37,205 images showing choroidal neovascularization (CNV), 11,348 showing diabetic macular edema (DME), 8,616 showing drusen, and 26,315 normal images. The test dataset comprised 250 images from each class. We divided the training dataset into a sub-training dataset and a validation dataset, which included 4,000 images extracted randomly from 1,000 images of each class. The sub-training dataset includes the remaining training data. The image format for the OCT dataset is joint photographic experts' group 8-bit. Figure 3 portrays some classified OCT dataset images.

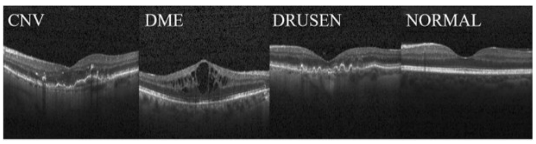

Figure 3. OCT images in the OCT dataset

Table 3. Diabetic people and ophthalmologist ratio

Year	No of patient with diabetes	No of ophthalmologist Available in India	Ratio per ophthalmologist
2021	101 million (10.1crore)	25,000	25/101,000 that is 4,040 diabetics have 1 ophthalmologist
2025	158 million (15.8 crore)	30,000	30/158,000 that is 5,266 diabetics have 1 ophthalmologist
2030	216 million (21.6 crore)	35,000	35/216,000 that is 6,171 diabetics have 1 ophthalmologist

Table 3 illustrates the growing disparity between the rising number of diabetic patients and the limited availability of ophthalmologists in India. As diabetes prevalence increases sharply from 2021 to 2030, the patient-to-ophthalmologist ratio worsens significantly, highlighting the urgent need for scalable, AI-assisted retinal screening and diagnostic solutions. Considering the above situation i.e., in 2025 one ophthalmologist will treat 5,266 diabetic people is a point of worry. If we ignore the facts, it will further deteriorate of 6,171. So, we come to conclusion that we must save the precious time of the ophthalmologist.

Normal retina: normal retina, with preserved foveal contour and absence of any retinal fluid/edema. The following are the parameters of a healthy retina:

i) Normal retina: indicates that the retina appears to be in a typical, healthy state without any visible abnormalities in image itself.

ii) Preserved foveal contour: the fovea, the part of the retina responsible for sharp central vision, and maintains its normal shape and structure.
iii) Absence of retinal fluid/edema: there is no fluid buildup or swelling in the retina, which is a positive sign that there are no underlying issues such as retinal detachment or macular edema.
iv) DME: DME is a complication of diabetes that affects the retina, specifically the macula, which is the central part of the retina responsible for sharp, detailed vision. In DME, there is a buildup of fluid in the macula due to leakage from damaged blood vessels. This fluid accumulation can cause adverse retinal thickening and can lead to vision impairment. When you mention "retinal-thickening-associated intraretinal fluid" it indicates that in the imaging, if we pointing to areas where there is both thickening of the retinal tissue and the presence of fluid within the retina. The occurrence of this fluid accumulation, commonly known as intraretinal fluid, is essential for diagnosing and monitoring DME. To effectively handle DME, it is essential to regulate blood sugar levels. Additional treatments that could be utilized comprise corticosteroids, anti-vascular endothelial growth factor (anti-VEGF) injections, or laser therapy to enhance vision and reduce fluid leakage.
v) Drusen: early AMD is often associated with drusen, which are small, yellowish spots that form beneath the retina. Drusen, visible as arrowheads on scans, serve as an initial sign of AMD and may be related to changes in visual quality. Drusen are commonly regarded as the initial stage of AMD, and their presence can help in diagnosing and monitoring the progression of the disease. Drusen may not have a major impact on vision, but people may experience little changes in their vision, such as trouble reading or seeing tiny details.
vi) CNV: in the context of AMD, CNV refers to choroidal neovascularization. This condition entails the formation of new, irregular blood vessels in the choroid layer located beneath the retina. The newly formed blood vessels are known as the neovascular membrane (indicated by arrowheads), which can lead to leaks and bleeding. The fluid accumulating between the retina and the underlying choroid due to leakage from these abnormal capillaries is referred to as subretinal fluid (indicated by arrows). Sudden vision deterioration and distortion can result from the swelling of this fluid and damage to the retina. To manage symptoms and prevent further vision loss, treatment for CNV, typically associated with wet AMD, must be initiated promptly. Anti-VEGF injections are commonly utilized in therapies to target and halt the progression of these abnormal blood vessels. Routine imaging monitoring is crucial to assess treatment effectiveness and implement necessary modifications.

Overall, these observations suggest that the retina is in good condition. If this is part of a report or assessment, it generally implies that there are no signs of retinal disease or damage at the time of examination. Figure 4 shows the samples of normal, Figure 5 shows the samples of DME, Figure 6 shows the samples of drusen, and lastly Figure 7 shows the samples of CNV.

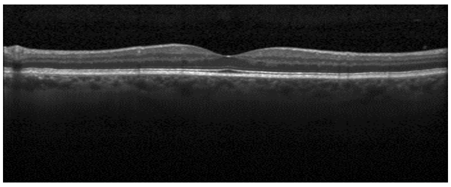

Figure 4. Sample of normal

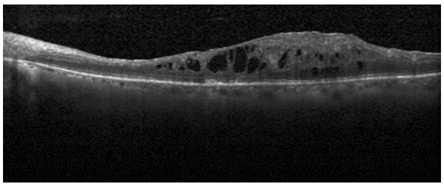

Figure 5. Sample of DME

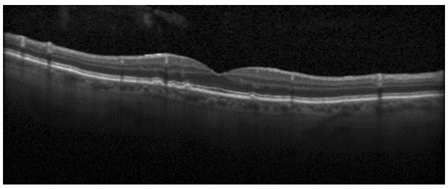

Figure 6. Sample of drusen

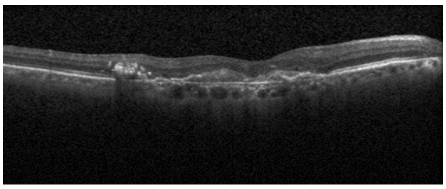

Figure 7. Sample of CNV

7. CONCLUSION

In response to the increasing need for automated tools in ophthalmology, this research presents a solid foundation for translating retinal OCT images into diagnostic text by combining structured data gathering with DL algorithms. The framework guarantees that high-quality OCT data is ready for model training by employing sophisticated image processing methods like contrast enhancement, noise reduction, and edge detection. The project effectively tackles class imbalances within the dataset of 84,484 retinal images through the application of preprocessing and data augmentation techniques. This enhances the model's ability to classify retinal disorders like DME, drusen, and CNV. The integration of DL, feature extraction, and selection algorithms automates diagnosis while enhancing scalability and efficiency in clinical environments. This method meets the essential need for precise and prompt identification of retinal disorders due to the rising incidence of ailments like DR. The framework can significantly alleviate the workload of ophthalmologists, enhance clinical decision-making, and boost patient outcomes by facilitating early treatment and intervention through optimizing the diagnostic process. Future research should focus on expanding the dataset variety and exploring more advanced DL architectures to improve the accuracy and relevance of the system across various clinical situations.

ACKNOWLEDGEMENT

We would like to thank to Oracle research for supporting this project giving financial assistant of Rs.250,000 thousand in the form of cloud services. B89009 and B88206 Cloud services offered by Oracle Research of cost 250,000/-.

FUNDING INFORMATION

The authors declare that no financial support or funding was received for the research, authorship, and/or publication of this article.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Uday Mande	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Shafi Pathan	✓				✓			✓		✓	✓		✓	
Pankaj Chandre	✓				✓	✓	✓		✓	✓				
Sharvari Mande	✓			✓	✓					✓		✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known competing financial or non-financial interests that could have influenced the work reported in this paper.

INFORMED CONSENT

This section not applicable, as this study did not involve human participants or the use of identifiable personal data.

ETHICAL APPROVAL

This section not applicable, as this study did not involve human participants or animal subjects.

DATA AVAILABILITY

The authors confirm that the data supporting the findings of this study are available within the article [and/or its supplementary materials].

REFERENCES

[1] A. A. Jammal *et al.*, "Detecting retinal nerve fibre layer segmentation errors on spectral domain-optical coherence tomography with a deep learning algorithm," *Scientific Reports*, vol. 9, no. 1, Jul. 2019, doi: 10.1038/s41598-019-46294-6.

[2] R. Kapoor, B. T. Whigham, and L. A. Al-Aswad, "Artificial intelligence and optical coherence tomography imaging," *Asia-Pacific Journal of Ophthalmology*, vol. 8, no. 2, pp. 187–194, 2019, doi: 10.22608/APO.201904.

[3] R. B. Nussenblatt, S. C. Kaufman, A. G. Palestine, M. D. Davis, and F. L. Ferris, "Macular thickening and visual acuity: measurement in patients with cystoid macular edema," *Ophthalmology*, vol. 94, no. 9, pp. 1134–1139, Sep. 1987, doi: 10.1016/S0161-6420(87)33314-7.

[4] J. G. Fujimoto, W. Drexler, J. S. Schuman, and C. K. Hitzenberger, "Optical coherence tomography (OCT) in ophthalmology: introduction," *Optics Express*, vol. 17, no. 5, pp. 3978–3979, Mar. 2009, doi: 10.1364/OE.17.003978.

[5] A. G. Podoleanu, "Optical coherence tomography," *Journal of Microscopy*, vol. 247, no. 3, pp. 209–219, Sep. 2012, doi: 10.1111/j.1365-2818.2012.03619.x.

[6] S. S. Damre, B. D. Shendkar, N. Kulkarni, P. R. Chandre, and S. Deshmukh, "Smart healthcare wearable device for early disease detection using machine learning," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 4s, pp. 158–166, 2024.

[7] A. Tayal, J. Gupta, A. Solanki, K. Bisht, A. Nayyar, and M. Masud, "DL-CNN-based approach with image processing techniques for diagnosis of retinal diseases," *Multimedia Systems*, vol. 28, no. 4, pp. 1417–1438, Aug. 2022, doi: 10.1007/s00530-021-00769-7.

[8] A. Jadhav, R. Bhosale, B. Shendkar, S. Jagadale, D. Lokare, and P. Chandre, "AI powered document verification - a behavioral and multi domain approach to automated fraud detection," *2025 IEEE 5th International Conference on ICT in Business Industry & Government (ICTBIG),* 2025, pp. 1-7, doi: 10.1109/ICTBIG68706.2025.11323606.

[9] P. Gholami, P. Roy, M. K. Parthasarathy, and V. Lakshminarayanan, "OCTID: optical coherence tomography image database," *Computers and Electrical Engineering*, vol. 81, Jan. 2020, doi: 10.1016/j.compeleceng.2019.106532.

[10] V. Bidve *et al.*, "Use of explainable ai to interpret the results of nlp models for sentimental analysis," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 35, no. 1, pp. 511–519, Jul. 2024, doi: 10.11591/ijeecs.v35.i1.pp511-519.

[11] T. Tsuji *et al.*, "Classification of optical coherence tomography images using a capsule network," *BMC Ophthalmology*, vol. 20, no. 1, Dec. 2020, doi: 10.1186/s12886-020-01382-4.

[12] J.-C. Mwanza and D. L. Budenz, "New developments in optical coherence tomography imaging for glaucoma," *Current Opinion in Ophthalmology*, vol. 29, no. 2, pp. 121–129, Mar. 2018, doi: 10.1097/ICU.0000000000000452.

[13] J. H. B. Im, Y.-P. Jin, R. Chow, and P. Yan, "Prevalence of diabetic macular edema based on optical coherence tomography in people with diabetes: a systematic review and meta-analysis," *Survey of Ophthalmology*, vol. 67, no. 4, pp. 1244–1251, Jul. 2022, doi: 10.1016/j.survophthal.2022.01.009.

[14] E. Hassan *et al.*, "Enhanced deep learning model for classification of retinal optical coherence tomography images," *Sensors*, vol. 23, no. 12, Jun. 2023, doi: 10.3390/s23125393.

[15] S. Goel *et al.*, "Deep learning approach for stages of severity classification in diabetic retinopathy using color fundus retinal images," *Mathematical Problems in Engineering*, vol. 2021, pp. 1–8, Nov. 2021, doi: 10.1155/2021/7627566.

[16] P. P. Srinivasan *et al.*, "Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images," *Biomedical Optics Express*, vol. 5, no. 10, Oct. 2014, doi: 10.1364/BOE.5.003568.

[17] J. R. Malgheet, N. B. Manshor, and L. S. Affendey, "Iris recognition development techniques: a comprehensive review," *Complexity*, vol. 2021, no. 1, Jan. 2021, doi: 10.1155/2021/6641247.

[18] K. Oh, H. M. Kang, D. Leem, H. Lee, K. Y. Seo, and S. Yoon, "Early detection of diabetic retinopathy based on deep learning and ultra-wide-field fundus images," *Scientific Reports*, vol. 11, no. 1, Jan. 2021, doi: 10.1038/s41598-021-81539-3.

[19] A. Moraru, D. Costin, R. Moraru, and D. Branisteanu, "Artificial intelligence and deep learning in ophthalmology-present and future (review)," *Experimental and Therapeutic Medicine*, Aug. 2020, doi: 10.3892/etm.2020.9118.

[20] Y. Tong, W. Lu, Y. Yu, and Y. Shen, "Application of machine learning in ophthalmic imaging modalities," *Eye and Vision*, vol. 7, no. 1, Dec. 2020, doi: 10.1186/s40662-020-00183-6.

[21] W. L. Alyoubi, W. M. Shalash, and M. F. Abulkhair, "Diabetic retinopathy detection through deep learning techniques: a review," *Informatics in Medicine Unlocked*, vol. 20, 2020, doi: 10.1016/j.imu.2020.100377.

[22] S. E. H. Andrab and A. Gupta, "Diabetic retinopathy disease classification using retina images: a review," *International Journal for Research in Applied Science and Engineering Technology*, vol. 10, no. 5, pp. 1452–1458, May 2022, doi: 10.22214/ijraset.2022.42432.

[23] D. Nagasato *et al.*, "Automated detection of a nonperfusion area caused by retinal vein occlusion in optical coherence tomography angiography images using deep learning," *PLoS ONE*, vol. 14, no. 11, Nov. 2019, doi: 10.1371/journal.pone.0223965.

[24] S. J. Chiu, X. T. Li, P. Nicholas, C. A. Toth, J. A. Izatt, and S. Farsiu, "Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation," *Optics Express*, vol. 18, no. 18, pp. 19413–19428, 2010, doi: 10.1364/OE.18.019413.

[25] F. G. Venhuizen *et al.*, "Automated staging of age-related macular degeneration using optical coherence tomography," *Investigative Ophthalmology & Visual Science*, vol. 58, no. 4, pp. 2318–2328, 2017, doi: 10.1167/iovs.16-21020.

[26] D. S. Kermany *et al*., "Identifying medical diagnoses and treatable diseases by image-based DL," *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018, doi: 10.1016/j.cell.2018.02.010.

[27] E. Vaghefi, S. Hill, H. M. Kersten, and D. Squirrell, "Multimodal retinal image analysis via deep learning for the diagnosis of intermediate dry age-related macular degeneration: a feasibility study," *Journal of Ophthalmology*, vol. 2020, pp. 1–10, 2020, doi: 10.1155/2020/7493419.

[28] L. Fang, D. Cunefare, C. Wang, R. H. Guymer, S. Li, and S. Farsiu, "Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search," *Biomedical Optics Express*, vol. 8, no. 10, pp. 4581–4596, 2017, doi: 10.1364/BOE.8.002732.

[29] S. K. Devalla *et al*., "A deep learning approach to digitally stain optical coherence tomography images of the optic nerve head," *Investigative Ophthalmology & Visual Science*, vol. 59, no. 1, pp. 63–74, 2018, doi: 10.1167/iovs.17-22617.

[30] S. P. K. Karri, D. Chakraborty, and J. Chatterjee, "Transfer learning based classification of optical coherence tomography images with diabetic macular edema and dry age-related macular degeneration," *Biomedical Optics Express*, vol. 8, no. 2, pp. 579–592, 2017, doi: 10.1364/BOE.8.000579.

[31] T. Schlegl, P. Seeböck, S. M. Waldstein, U. S.-Erfurth, and G. Langs, "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery," *Information Processing in Medical Imaging*, pp. 146–157, 2017, doi: 10.1016/j.media.2019.01.003.

[32] 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," *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 618-626, doi: 10.1109/ICCV.2017.74.
[33] A. G. Howard *et al*., "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.

BIOGRAPHIES OF AUTHORS

Uday Mande is a research scholar at MIT ADT University, Loni Pune. He has obtained his B.E. degree in CSE from Shri Guru Gobind Singhji COE, Nanded in 1993 from Marathwada University, Aurangabad, India, M.E. degree in Computer Science and Engineering from Pune University Maharashtra, India in the year 2006. He is currently working as an assistant professor in Department of Computer Science and Engineering, MIT School of Computing, MIT ADT, Pune, India. He has published 40 plus papers at international journals and conferences with 37 citations. He has guided more than 30 plus under-graduate students and 20 plus postgraduate students for projects. His research interests are data mining, AI, machine learning, and information security. He can be contacted at email: uday.mande@mituniversity.edu.in.

Shafi Pathan is a professor at Department of Computer Engineering, School of Computing, MIT Art, Design and Technology University, Pune, India. He completed his Ph.D. (CSE) from JNTU Anantapur, India. He has completed a university funded research project on public key cryptography for cross-realm authentication in Kerberos. He has worked as head of the Publicity Committee for International Conference at the Global ICT Standardization Forum for India. He was guest editor for a special issue of ICINC 2016 by IGI Global International Journal of Rough Data Sets and Analytics. Visited Curtin University (Australia), Dubai Campus, Murdoch University, Dubai Campus, University of West London (UK), RAK Campus, University of Bolton (UK), RAK Campus, De Montfort University (UK), Dubai Campus at Dubai Campus, He has authored 13 books and 7 book chapters for CRC and Springer. He has published more than 63 research articles in national and international journals. He has 6 Indian patents and 3 copyrights on his credit. He can be contacted at email: shafi.pathan@mituniversity.edu.in.

Pankaj Chandre has obtained his B.E. degree in Information Technology from Sant Gadge Baba Amravati University, Amravati, India, M.E. degree in Computer Engineering from Mumbai University Maharashtra, India in the year 2011 and Ph.D. in Computer Engineering from Savitribai Phule Pune University, Pune, India in the year 2021. He is currently working as an associate professor in Department of Computer Science and Engineering, MIT School of Computing, MIT ADT, Pune, India. He has published 60 plus papers at international journals and conferences. He is guiding 8 Ph.D. research scholar at MIT Art Design and Technology University, Pune, India. He has guided more than 30 plus under-graduate students and 20 plus postgraduate students for projects. His research interests are network security and information security. He can be contacted at email: pankaj.chandre@mituniversity.edu.in.

Sharvari Mande completed M.B.B.S from Rajarshee Chhatrapati Shahu Maharaj Government Medical College, Kolhapur. Have presented her research 'Disparity in cost spent on tubectomies and vasectomies in India in year 2019-20' in international surgical week, Malaysia 2024 international college of surgeons-US 2024 (second prize). She can be contacted at email: sumande6128@gmail.com.