Data in Brief 57 (2024) 110979

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

A dataset of color fundus images for the detection and classification of eye diseases

Shayla Sharmin[a], Mohammad Riadur Rashid[a], Tania Khatun[a,*], Md Zahid Hasan[a], Mohammad Shorif Uddin[b], Marzia[c]

[a] Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh
[b] Department of Computer Science and Engineering, Jahangirnagar University, Dhaka, Bangladesh
[c] Faculty of Medicine, University of Dhaka, Bangladesh

ARTICLE INFO

Article history:
Received 15 July 2024
Revised 10 September 2024
Accepted 24 September 2024
Available online 4 October 2024

Dataset link: Eye Disease Image Dataset (Original data)

Keywords:
Eye disease recognition
Health analytics
Computer vision
Deep learning
Image processing
Machine learning

ABSTRACT

The retina is a critical component of the eye responsible for capturing visual information, making the importance of retinal health for clear vision. Various eye diseases, such as age-related macular degeneration, diabetic retinopathy, and glaucoma, can severely impair vision and even lead to blindness if not detected and treated early. Therefore, automated systems using machine learning and computer vision techniques have shown promise in the early detection and management of these diseases, reducing the risk of vision loss. In this context, to facilitate the development and evaluation of machine learning models for eye disease detection, we introduced a comprehensive dataset which was collected during a span of eight months from Anawara Hamida Eye Hospital & B.N.S.B. Zahurul Haque Eye Hospital using Color Fundus Photography machine. The dataset has two categories of data: color fundus photographs and anterior segment images. The color fundus photographs categorized into nine classes: Diabetic Retinopathy, Glaucoma, Macular Scar, Optic Disc Edema, Central Serous Chorioretinopathy (CSCR), Retinal Detachment, Retinitis Pigmentosa, Myopia, Healthy and anterior segment images has one class: Pterygium. This dataset comprises 5335 primary images. By providing a rich and diverse collection of color fundus photographs, this dataset serves as a valuable resource for researchers and clinicians

* Corresponding author.
E-mail address: tania.cse@diu.edu.bd (T. Khatun).

https://doi.org/10.1016/j.dib.2024.110979

in the field of ophthalmology for the automatic detection of nine different classes of eye diseases.

Specifications Table

Subject	Computer Science
Specific subject area	Eye Disease Diagnosis, Medical Imaging, Clinical Decision Support, Telemedicine, Computer Vision
Type of data	JPG
Data collection	In collaboration with Health Informatics Lab-Daffodil International University, and the authorities of Anawara Hamida Eye Hospital & B.N.S.B. Zahurul Haque Eye Hospital the dataset was collected. It is important to highlight that this dataset is completely new, and there has been no previous research conducted with it.
Data source location	**Location:** Anawara Hamida Eye Hospital and B.N.S.B. Zahurul Haque Eye Hospital in Faridpur district **Zone:** Faridpur **Country:** Bangladesh
Data accessibility	**Repository name:** Mendeley Data **Data identification number:** 10.17632/s9bfhswzjb.1 **Direct URL to data:** https://data.mendeley.com/datasets/s9bfhswzjb/1

1. Value of the Data

- Across the world, eye conditions pose a significant concern, with Bangladesh experiencing a notable impact where 1.5% of adults suffer from blindness and 21.6% demonstrate impaired vision [1]. This issue is particularly pronounced in developing countries, where technological resources and financial support are scarce. The primary challenge in these regions remains the early identification and rapid, accurate classification of diseases. Detecting eye diseases early can prevent many individuals from experiencing blindness. Computer-assisted diagnosis has the potential to expedite the process of obtaining eye diagnoses, establishing a benchmark for clinical experts or screening purposes.
- In healthcare, identifying diseases early is crucial for more effective treatment, drawing significant interest from researchers due to its wide-ranging advantages and reducing the burden on healthcare professionals [2]. Enhanced diagnostic accuracy and the potential for advanced diagnosis are facilitated by research conducted on datasets of eye diseases.
- Researchers can harness the aggregated dataset for a multitude of tasks in computer vision and machine learning, extending beyond its initial applications, and deep learning approaches Researchers can easily classify various retinal ailments including glaucoma, diabetic retinopathy, Disc Edema, Glaucoma, central serous chorioretinopathy, Macular Scar, Myopia, Pterygium, Retinal Detachment, Retinitis Pigmentosa degeneration. This involves three primary stages: (a) Preparing the data, (b) Deriving relevant features, and (c) Implementing multi-classification of eye diseases [3]. Ensuring precise classification of images necessitates extracting crucial characteristics from the segmented blood vessels and employing diverse machine and deep learning techniques to accurately diagnose eye conditions. This dataset will serve as an invaluable resource for researchers, facilitating the acceleration of their research endeavors (Table 1).
- Our dataset stands as one of the largest and most comprehensive resources in the field of eye disease research. Our dataset encompasses nine distinct types of eye diseases and healthy, comprising a total of 5335 original images.

Table 1
Comparison with related datasets.

Dataset	Number of classes	Name of classes	Number of data
[18]	1	Diabetic Retinopathy	757 color fundus images
[19]	1	Diabetic Retinopathy	111 color fundus images
[20]	4	Hypertensive Retinopathy, Diabetic Retinopathy, Papilledema and Healthy	100 color fundus images
Ours	10	Diabetic Retinopathy, Glaucoma, Macular Scar, Optic Disc Edema, Central Serous Chorioretinopathy (CSCR), Retinal Detachment, Retinitis Pigmentosa, Myopia, Healthy and Pterygium	5318 color fundus images and 17 anterior segment images

2. Background

The motivation behind creating this dataset was to address challenges in identifying and categorizing common eye disorders in diagnostic processes for eye diseases. This dataset aligns with contemporary efforts in eye care to improve diagnostic and treatment methods using technological advancements. Additionally, the limited availability of comprehensive datasets in Bangladesh's eye care sector posed challenges for developing accurate identification and classification models. We have assembled a collection of 5335 images representing various eye conditions. This dataset facilitates the identification and classification of eye conditions and acts as an important asset for efficiently and accurately training and validating deep learning models. By providing researchers and practitioners with access to the raw data and advocating for clarity and replicability, this dataset publication contributes to advancing research in eye disease diagnosis.

3. Data Description

The incidence of eye diseases is on the rise in Bangladesh due to challenges in accurately identifying and classifying them. A significant shortage of experts further exacerbates the issue, leading to more patients and diseases going undiagnosed. Properly recognizing disease signs, especially in eye images is essential for effective diagnosis. Automation, using computer vision algorithms, is revolutionizing the healthcare sector, offering a more accurate and efficient alternative to manual diagnosis, which is often time-consuming, prone to errors, and costly due to the lack of specialized experts in eye diseases. In this paper, we introduced a dataset combination of Color Fundus Photographs and Anterior Segment of eye for Eye Disease detection and classification. The dataset comprises images captured directly by a computerized fundus imaging system. Fig. 1 illustrates the dataset organization.

The dataset includes images representing several ocular illnesses, including macular scars, pterygium, disc edema, diabetic retinopathy, glaucoma, retinal detachment, myopia, and retinitis pigmentosa. These images were collected during the period spanning from July 2023 to February 2024 from the Anawara Hamida Eye Hospital and B.N.S.B. Zahurul Haque Eye Hospital in Faridpur, Bangladesh. So, this dataset representing Bengali ethnicity. To capture the images, hospital authorities used Topcon TL-211 and Topcon TRC-50DX model fundus cameras. The resultant image sizes of 3900 $\times$ 2600 and 2004 $\times$ 1690 pixels were captured and stored in the JPG format. To increase the efficiency of computer vision algorithms, we enlarge every image to 2004 $\times$ 1690 pixels. For easier classification and analysis, each image in the collection has been annotated with the disease name.

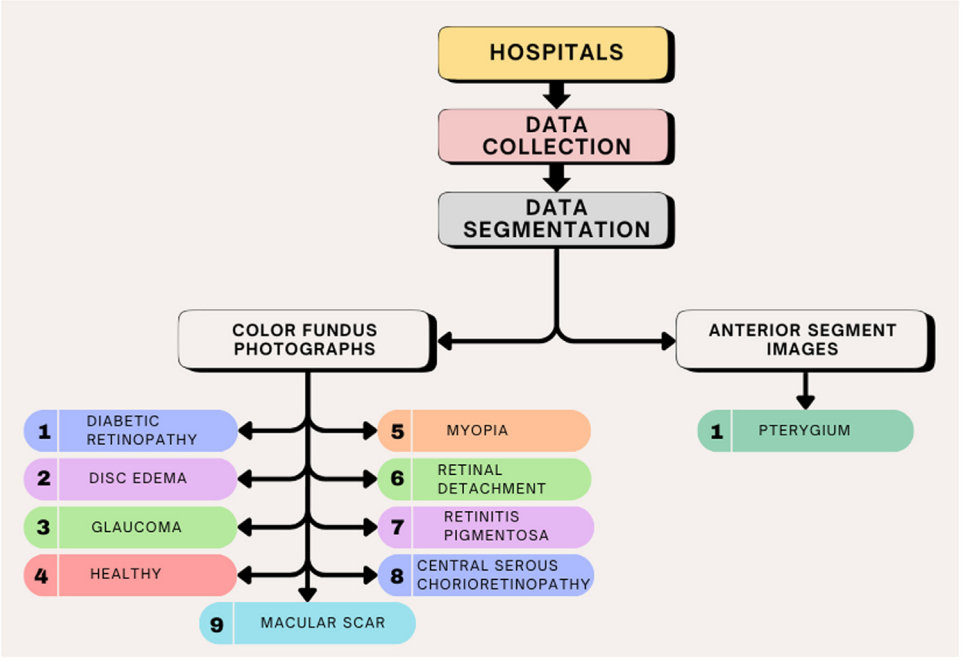


Fig. 1. Organization of the dataset.

While collecting images from the hospitals, we encountered a few challenges, such as,

- Securing data was challenging due to institutional regulations and procedures. Obtaining authorization to access and use the data required significant time and effort in negotiating agreements.
- During data collection, challenges arose from inconsistent data quality and the requirement for precise segmentation. Factors such as the equipment used, the patient's condition, and the anthropologist's skill level influenced the quality of the fundus photographs. Additionally, grouping all disease-related photos in a single directory made accurate disease-specific image segmentation particularly challenging.
- One challenge we faced during data collection was limited resources. Hospitals lacked dedicated infrastructure and personnel for data sharing and management. Acquiring photographs from these hospitals necessitated considerable time, effort, and financial investment.

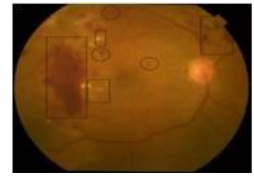
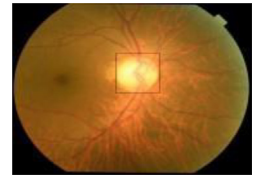
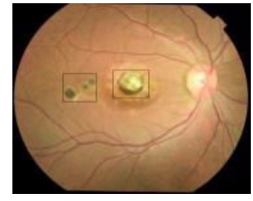
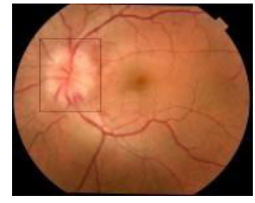
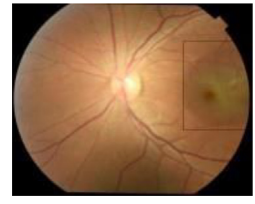
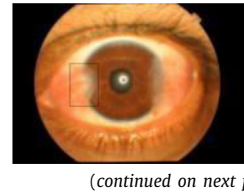
The dataset is divided into ten categories: Diabetic Retinopathy, Glaucoma, Macular Scar, Disc Edema, Central Serous Chorioretinopathy, Healthy, Pterygium, Retinal Detachment, Retinitis Pigmentosa, and Myopia. Table 2 outlines the specific classes included in the dataset.

The color fundus photographs dataset holds promise across various applications:

Disease Diagnosis and Clinical Decision Support: Utilizing machine learning and computer vision methods, automated or semi-automated analysis of fundus photographs can aid in early disease detection and treatment. Fundus photograph datasets serve as valuable tools for medical professionals in making clinical assessments. Algorithms trained on these datasets can provide decision support by identifying abnormalities, assessing disease severity, and predicting disease progression.

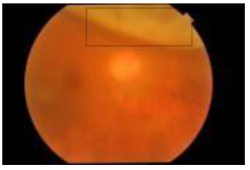
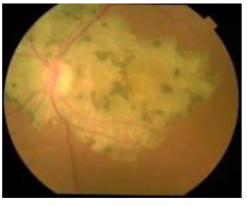
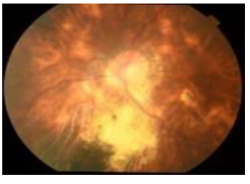
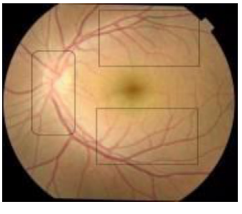
Medical Education and Research: Fundus photograph datasets serve as valuable educational resources for medical students, optometrists, and ophthalmologists, as well as other healthcare professionals. These datasets can be employed to develop training modules, virtual simulators, and educational content that instruct students on the anatomy, pathology, and diagnostic inter-

Table 2
Details about the ten classes of sample images in the dataset.

Class Name	Description	Visualization
Diabetic Retinopathy	Several visual effects may be observed: Microaneurysms: small, round, red dots. Hemorrhages: small dots, scars, and large red patches due to blood leakage from capillaries. Hard Exudates: These are yellow-white lipid buildups that form in the retina as a result of leakage from compromised blood vessels [4].	
Glaucoma	In fundus photography, the vertical cup-to-disc ratio (CDR) stands as a commonly adopted clinical metric used by medical practitioners. Glaucoma is a long-term, progressive optic nerve disorder distinguished by changes in the optic nerve head and the retinal nerve fiber layer [5,6].	
Macular Scar	The classic method of macular scar detection is retinal color photography to detect drusen and geographic atrophy. Damage to the retinal pigment epithelium and an abnormal inflammatory response are key factors in the development of geographic atrophy [7,8].	
Optic Disc Edema	Optic disc edema may present with optic nerve head bulging, hyperemia, vascular congestion, and peripapillary hemorrhages. Also, disc edema is visually apparent as a prominent and blurred optic disc margin. [9]	
Central Serous Chori-oretinopath y (CSCR)	The diagnosis of CSCR typically starts with a routine retina examination, revealing a central shallow serous detachment of the retina.The combination of central serous retinal detachment, retinal pigment epithelium detachment, and retinal pigment epithelium abnormalities are characteristic features of CSCR [10].	
Pterygium	Wing-shaped conjunctivitis, which is caused by an aberrant tissue layer on the conjunctiva, is referred to as pterygium. This eye condition manifests as a triangle-shaped growth that invades the cornea and grows close to the nasal area [11].	

(*continued on next page*)

Table 2 (*continued*)

Class Name	Description	Visualization
Retinal Detachment	A serious condition called retinal detachment (RD) occurs when the retinal pigment epithelium separates from the neurosensory retinal layer underneath [12,13].	
Retinitis Pigmentosa	In retinal imaging, pigmented patches on the posterior portion of the retina can be seen. These patches, along with any accompanying symptoms, may grow and move over time [14].	
Myopia	It is primarily marked by axial lengthening. As myopia progresses and the eye's axis extends, the observable retinal and choroidal abnormalities in the fundus worsen, leading to various significant complications [15].	
Healthy	In normal fundus photographs, the retina appears transparent, allowing visualization of the underlying structures. The reddish-orange glow observed in funduscopy is primarily due to the reflection of light from the choroid, a layer of tissue rich in blood vessels located between the retina and the sclera (the white outer layer of the eye) [16,17].	

pretation of the eye. Additionally, fundus photograph databases are crucial for ophthalmology and related research fields. Researchers can leverage these datasets to develop and validate new prognostic models, treatment methods, and diagnostic instruments, aiming to improve patient outcomes and advance scientific knowledge.

Telemedicine and Remote Monitoring: Fundus photograph datasets support telemedicine initiatives by enabling remote screening, diagnosis, and monitoring of eye diseases. Integrating automated analytical algorithms with remote fundus photography can enhance access to eye care services, particularly in underserved or remote areas where the availability of ophthalmologists is limited.

4. Experimental Design, Materials and Methods

4.1. Camera Specification

The dataset was captured by two Topcon digital fundus cameras (TRC-50DX and TL-211) which are linked to separate Nikon DSLR cameras (D7500 and D90 respectively). The TRC-50DX offers more features (color fundus, red-free, fluorescein imaging, adjustable flash intensity) and

Table 3
Statistics of the retinal disease dataset.

Name of Class	Original Image Number
Central Serous Chorioretinopathy	101
Diabetic Retinopathy	1509
Disc Edema	127
Glaucoma	1349
Healthy	1024
Macular Scar	444
Myopia	500
Pterygium	17
Retinal Detachment	125
Retinitis Pigmentosa	139
Total	5335

captures higher resolution images (up to 5600 × 3728) compared to TL-211 (max resolution 4288 × 2848). The sensor on the TRC-50DX camera (D7500) has smaller pixels (4.2 µm) compared to the TL-211 (D90) sensors (5.51 µm) (Table 3).

4.2. Process

Data Labeling: During the preliminary data pre-processing stage, we meticulously labeled the data, precisely categorizing each image into its respective class or category under the observation of experts. Deep learning models rely on labeled data for training and refinement; without precise labels, the models cannot learn effectively or make accurate predictions.

Image Resizing: We found that to create a consistent and coherent dataset, we needed to standardize the photo sizes, as they varied initially. This method decreased computational requirements and guaranteed suitability for training deep learning models.

Limitations

The classification of any other disease would not be possible for this dataset because it solely relates to and is primarily focused on eye disease. The dataset has geographical limitations because of the data sources. The dataset originates from two hospitals in Bangladesh representing Bengali ethnicity.

Ethics Statement

We confirm that the authors have read and followed the ethical requirements for publication in Data in Brief. The dataset was collected in collaboration with Health Informatics Lab-Daffodil International University, and the authorities of Anawara Hamida Eye Hospital & B.N.S.B. Zahurul Haque Eye Hospital. In addition, every patient's data stays unknown and his or her illness status is with the utmost confidentiality.

Data Availability

Eye Disease Image Dataset (Original data) (Mendeley Data).

CRediT Author Statement

Shayla Sharmin: Conceptualization, Data curation, Methodology, Writing – review & editing; **Mohammad Riadur Rashid:** Data curation, Methodology; **Tania Khatun:** Writing – review & editing, Methodology, Data curation; **Md Zahid Hasan:** Supervision, Writing – review & editing; **Mohammad Shorif Uddin:** Supervision, Writing – review & editing; **Marzia:** Data curation, Validation.

Acknowledgements

We are very grateful to the authorities from Anawara Hamida Eye Hospital and B.N.S.B. Zahurul Haque Eye Hospital for their valuable cooperation in accomplishing the task.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] I. Sutradhar, et al., Eye diseases: the neglected health condition among urban slum population of Dhaka, Bangladesh, BMC Ophthalmol. 19 (2019) 38, doi:10.1186/s12886-019-1043-z.
[2] K. Nara, V. Yada, S. Bashetty, S. Pallerla, S. Mudavath, Retina capturing using portable device and classification of images for prevalence of eye disease, in: Proceedings of the International Conference on Communications and Cyber Physical Engineering 2018, Singapore: Springer Nature Singapore, 2024, pp. 863–871.
[3] Q. Abbas, I. Qureshi, J. Yan, et al., Machine learning methods for diagnosis of eye-related diseases: a systematic review study based on ophthalmic imaging modalities, Arch. Comput. Methods Eng. 29 (2022) 3861–3918, doi:10.1007/s11831-022-09720-z.
[4] W.L. Alyoubi, W.M. Shalash, M.F. Abulkhair, Diabetic retinopathy detection through deep learning techniques: a review, Inform. Med. Unlocked 20 (2020) 100377.
[5] Fu, H., Cheng, J., Xu, Y., & Liu, J. (2019). Glaucoma detection based on deep learning network in fundus image. Deep learning and convolutional neural networks for medical imaging and clinical informatics, 119-137.
[6] L.J. Coan, B.M. Williams, V.K. Adithya, S. Upadhyaya, A. Alkafri, S. Czanner, G. Czanner, Automatic detection of glaucoma via fundus imaging and artificial intelligence: a review, Surv. Ophthalmol. 68 (1) (2023) 17–41.
[7] Y. Kanagasingam, A. Bhuiyan, M.D. Abràmoff, R.T. Smith, L. Goldschmidt, T.Y. Wong, Progress on retinal image analysis for age related macular degeneration, Prog. Retin. Eye Res. 38 (2014) 20–42.
[8] R.D. Jager, W.F. Mieler, J.W. Miller, Age-related macular degeneration, N. Engl. J. Med. 358 (24) (2008) 2606–2617.
[9] S. Urfalioglu, G. Ozdemir, M. Guler, G.G. Duman, The evaluation of patients with optic disc edema: a retrospective study, North Clin. Istanb. 8 (3) (2021) 280.
[10] F. Semeraro, F. Morescalchi, A. Russo, E. Gambicorti, A. Pilotto, F. Parmeggiani, C. Costagliola, Central serous chorioretinopathy: pathogenesis and management, Clin. Ophthalmol. (2019) 2341–2352.
[11] N.S.M. Zamani, W.M.D.W. Zaki, A.B. Huddin, A. Hussain, H.A. Mutalib, A. Ali, in: Automated Pterygium Detection Using Deep Neural Network, 8, IEEE Access, 2020, pp. 191659–191672.
[12] W. Wei, Atlas of Retinal Detachment: Diagnosis and Differential Diagnosis, Springer, 2019.
[13] S. Yadav, S. Das, R. Murugan, S. Dutta Roy, M. Agrawal, T. Goel, A. Dutta, Performance analysis of deep neural networks through transfer learning in retinal detachment diagnosis using fundus images, Sādhanā 47 (2) (2022) 49.
[14] R. Rashid, W. Aslam, A. Mehmood, D.L.R. Vargas, I.D.L.T. Diez, I. Ashraf, A Detectability Analysis of Retinitis Pigmetosa Using Novel SE-ResNet Based Deep Learning Model and Color Fundus Images, IEEE Access, 2024.
[15] C. Wan, H. Li, G.F. Cao, Q. Jiang, W.H. Yang, An artificial intelligent risk classification method of high myopia based on fundus images, J. Clin. Med. 10 (19) (2021) 4488.
[16] Akram I. and Rubinstein A., "Common retinal signs. An overview," Optometry Today, Jan. 28, 2005. [Online]. Available: https://www.aop.org.uk/
[17] V.S. Mary, E.B. Rajsingh, G.R. Naik, Retinal Fundus Image Analysis for Diagnosis of Glaucoma: a Comprehensive Survey, IEEE Access, 2016.
[18] V.E. Castillo, et al., Dataset from fundus images for the study of diabetic retinopathy, Data Brief 36 (2021) 107068, doi:10.1016/j.dib.2021.107068.
[19] P. Bidwai, S. Gite, A. Gupta, K. Pahuja, K. Kotecha, Multimodal dataset using OCTA and fundus images for the study of diabetic retinopathy, Data Brief 52 (2024) 110033.
[20] M.U. Akram, S. Akbar, T. Hassan, S.G. Khawaja, U. Yasin, I. Basit, Data on fundus images for vessels segmentation, detection of hypertensive retinopathy, diabetic retinopathy and papilledema, Data Brief 29 (2020) 105282.