Results in Optics 16 (2024) 100700

Contents lists available at ScienceDirect

Results in Optics

journal homepage: www.sciencedirect.com/journal/results-in-optics

Diabetic retinopathy detection through generative AI techniques: A review

Vipin Bansal [a,*], Amit Jain [a], Navpreet Kaur Walia [b]

[a] University Institute of Computing, Chandigarh University, India
[b] Department of Computer Science and Engineering, Chandigarh University, India

ARTICLE INFO

Keywords:
Diabetic Retinopathy
Generative AI
Anomaly Detection
Explainable AI
Computer Vision
OCT
Fundus

ABSTRACT

Diabetes, a burgeoning health issue, especially among the youth, stems from poor dietary habits and unhealthy lifestyles. India stands as the second most afflicted country, witnessing a rapid diabetes epidemic. Diabetic Retinopathy (DR), a significant complication, threatens vision loss and blindness. Early detection, alongside lifestyle adjustments, can manage DR effectively. Traditional methods for DR detection are time consuming, costly and require specialized skills. Computer assisted screening systems, leveraging technologies like Fundus images and Optical Coherence Tomography (OCT), streamline DR detection, with Artificial Intelligence (AI) playing a pivotal role. Technological advancements and abundant data fuel significant progress in AI-based DR screening, promising enhanced accuracy and efficiency, even in remote regions. In healthcare, "normal" and "abnormal" statuses characterize patient health. AI applications in healthcare often focus on anomaly detection, leveraging distinct data distributions. Generative architectures, originally designed for content generation, find application across various domains, including healthcare. By adjusting architecture and data pipelines, controlled and specific samples can be generated, offering solutions for anomaly detection.

This paper reviews fundamental aspects of diabetes and DR, exploring the utilization of generative AI in analyzing retinal data for DR detection. It also discusses recent advancements in Generative AI and their potential to enhance AI solutions in healthcare.

1. Introduction

Many factors can influence an individual's lifestyle, including their profession, daily routine, eating habits, geo-graphical conditions, religious faith, and more. In the case of the Information Technology (IT) industry, there are many opportunities for Indian people, but this profession can also have an impact on an individual's lifestyle. Working in the IT industry often involves working on shifts, which can disrupt an individual's daily routine and sleep patterns. Additionally, many IT professionals work long hours and may not have enough time to prioritize their personal health, leading to unhealthy eating habits and a lack of exercise. Due to this, someone can face lots of health issues, diabetes, hypertension (Padma et al., 2015) are a few of them. An unhealthy lifestyle marked by factors such as inadequate dietary choices, smoking, erratic daily routines, sedentary behaviour, and other elements can give rise to a wide array of health issues. These concerns may involve metabolic conditions like diabetes and obesity, musculoskeletal issues like arthritis, cardiovascular diseases such as heart strokes, hypertension (high blood pressure), and even mental health disorders like depression and anxiety (Mathur and Mascarenhas, 2019). Furthermore, poor lifestyle decisions can also play a role in fostering violence and other societal issues. Conversely, individuals who embrace a healthy lifestyle, engaging in regular physical exercise, maintaining a nutritious diet, ensuring proper rest, and obtaining sufficient sleep, are more likely to enjoy favorable mental and physical health outcomes. They may also contribute positively to their communities. It's worth noting that an unhealthy lifestyle is often associated with various diseases, with diabetes being a notable example. Diabetes is a chronic disease that can happen due to multiple reasons. Genetics and bad lifestyles are some of the prominent reasons for diabetes. The body of a diabetic patient won't be able to regulate blood sugar levels properly and that can contribute to other levels of health complications. When the body's immune system starts affecting those pancreas cells that generate insulin for the body that can lead to occur type-1 diabetes whereas when the human body is not able to produce sufficient insulin that can control blood sugar level or it becomes immune to the insulin that leads to occur type-2 diabetes.

* Corresponding author at: I-032, Spaze Privy The Address, Sector-93, Gurgaon, Haryana, 122505, India.
E-mail address: vipin_bansal1@yahoo.com (V. Bansal).

https://doi.org/10.1016/j.rio.2024.100700
Received 20 March 2024; Received in revised form 9 May 2024; Accepted 23 May 2024
Available online 4 June 2024

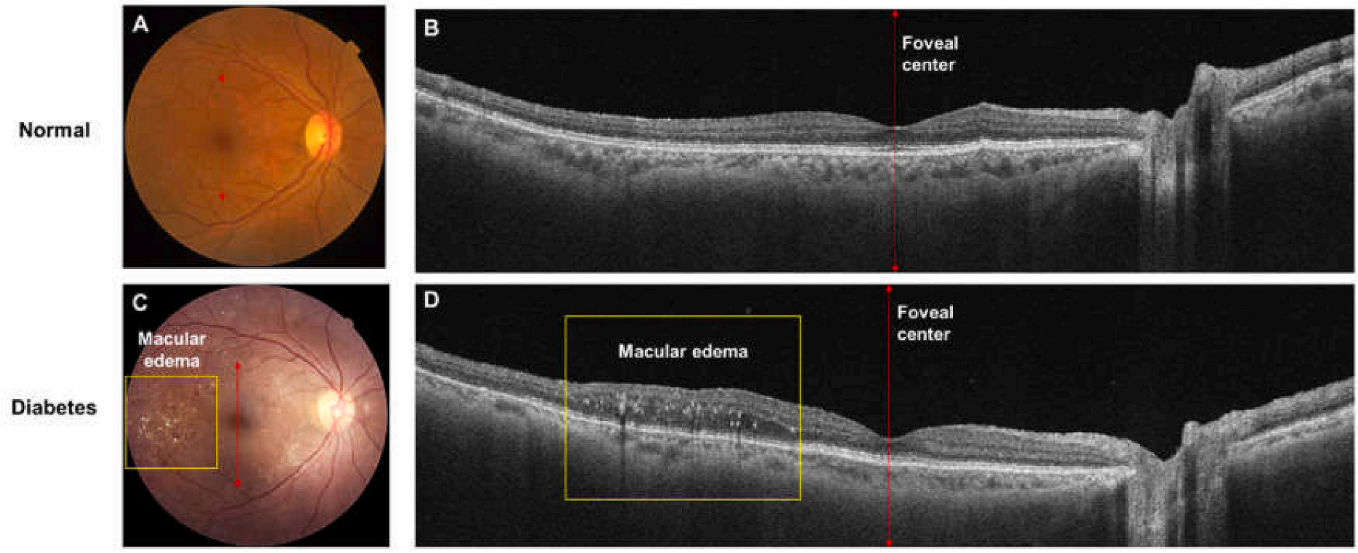

Fig. 1. Healthy vs Unhealthy retina (Fundus/OCT) (Chua et al., 2020.).

Generally, type-1 diabetes is detected in childhood or adolescence whereas type-2 diabetes is linked with poor lifestyle habits for example lack of physical activity, lack of nutritious diet etc.

According to the 2022 report from the IDF Diabetes Atlas on Type 1 diabetes (Ogle et al., 2021), there were approximately 8.75 million individuals globally affected by Type 1 diabetes. Of this population, around 1.52 million (17 %) are under the age of 20. In the year 2022, a total of 530,000 new cases of Type 1 diabetes were diagnosed across all age groups, with approximately 201,000 cases diagnosed in individuals under the age of 20. Another general report from the IDF Diabetes Atlas in 2021 (I. D. Federation, 2021) indicates that 537 million individuals between the ages of 20 and 79 are affected by diabetes worldwide, constituting 10.5 % of the world's population within this age group. Projections suggest an 11.3 % increase by the year 2030. The IDF estimates that approximately 50 % of individuals with diabetes globally remain undiagnosed, and one person succumbs to this condition every five seconds.

Studies have shown that diabetes is becoming more prevalent in younger people due to unhealthy lifestyle habits (Lascar et al., 2017). As mentioned in IDF Diabetes Atlas (I. D. Federation, 2021; Lascar et al., 2017) this could be a concerning trend as diabetes in severe cases can contribute to various other health complications including heart and kidney-related diseases, nerve damage and blindness.

- Cardiovascular disease, A diabetic patient is always at higher risk of heart-related diseases like stroke and other cardiovascular conditions.
- Kidney disease, A diabetic patient with continuous high blood sugar levels can damage the kidney and or even gradually lead to its failure.
- Eye problems, Diabetic patients can face various eye-related alignments due to the damage of blood vessels in the eyes. It can lead to severe vision-related problems to even blindness.
- Nerve damage, Continuous high blood sugar level can affect the body's nerves which can lead to neuropathy and raise other complications.

According to the Centers for Disease Control and Prevention report (C. f. D. C. a. Prevention, 2022), of adults diagnosed with diabetes, 39 % suffered from kidney disease, 69 % observed high blood pressure, 44 % had high cholesterol and 12 % reported vision impairments. Early detection and management of diabetes are crucial for preventing or minimizing these complications (Herman et al., 2015). This is why it becomes crucial for an individual to be aware of the risk factors for diabetes and to get regular check-ups with their healthcare provider. DR is one of the common complications of diabetes that happens due to the impact on the blood vessels in the eye retina. If left untreated, it can gradually lead to vision loss and blindness. The current standard method for diagnosing DR is a dilated eye exam performed by a qualified eye care professional (Fisher et al., 2016). This process can be time-consuming and costly, and it requires specialized training and expertise to identify the signs of DR accurately. However, with the advancements in technology, there are now computer-assisted screening systems that can detect DR using retinal images. These systems use machine learning algorithms to analyze the images and identify the presence of DR, making the process faster, more accurate, and less costly. The field of medical imaging is increasingly leveraging AI and machine learning to improve diagnostic accuracy and efficiency. Deep learning algorithms have shown promising results in analyzing medical images such as X-rays, MRI scans, retinal images etc. (Kalakota and Davenport, 2019) for conditions like diabetic retinopathy. Various computer vision-based solutions to detect DR have already been developed using OCT and Fundus images. Fundus or OCT images typically exhibit a high-dimensional representation of the retina, and individuals with DR may develop lesions on the retina. A sample image effectively illustrates the distinction between a healthy and an unhealthy retina in Fig. 1. These algorithms can detect and classify abnormalities in medical images with high accuracy without any need for invasive procedures.

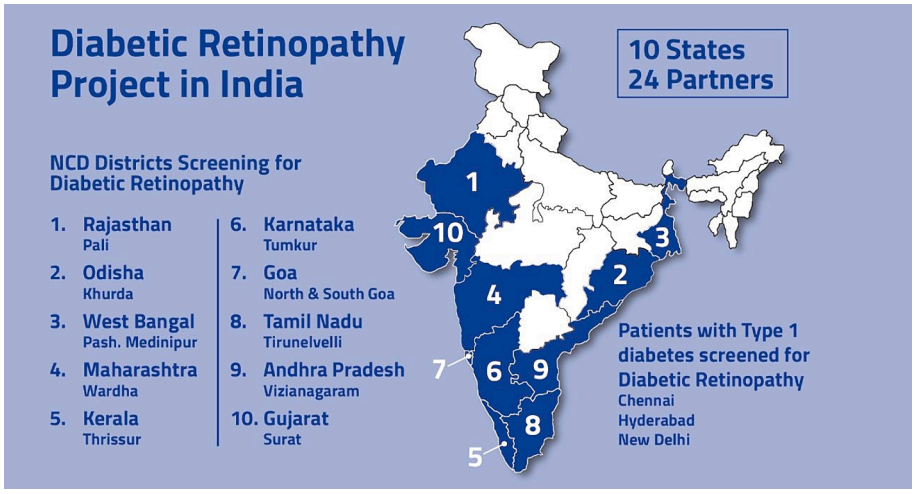

Fig. 2. State-wise distribution: The Trust's Diabetic Retinopathy Project (Diabetic Retinopathy, 2018).

This paper undertakes an in-depth analysis of Generative-AI approaches applied in recent years, focusing on the utilization of retina images for anomaly identification. The review strives to elucidate the techniques employed and scrutinizes existing research gaps. Through an exploration of these gaps, the paper aims to unveil potential avenues for future work in the realm of Generative-AI.

2. Motivation

IDF (I.D. Federation, 2021) report says, 74.2 million people in India affected by diabetes and by the 2045, it is expected to touch by 124.9 million. As per WHO reports (WHO, 2023), 1 in 11 is affected by diabetes and India is second most affected in the world after China. The prevalence of diabetes in India has been increasing at an alarming rate (nearly 25 more million people are at a higher risk and can develop diabetes in future), and it is a major health concern for the country.

The Sustainable Development Goals (SDG)-3.8 (Who, 2021), targets the coverage of essential health services globally. Eye health is equally essential for achieving SDG's. As per the IAPB report (a. v. I. V. A. a. T. L. G. H. C. o. G. E. H., 2021), due to the unavailability of basic eye care services around 1.1. billion people globally living with vision loss. Poverty is one of the other reasons for vision loss and vision loss problems spread in countries where the income is between the low and middle range.

India is the second most affected country and that is the reason the Indian government have started various programmes to develop and integrate services for the prevention, early detection, and management of DR (India: the Global Burden of Disease Study 1990–2016, 2018). Project "The Trust's Diabetic Retinopathy project" is being implemented in 53 facilities in 10 States in India, refer Fig. 2.

Diabetes can lead to various complications and that can make a big impact on the quality of patient's life. It is crucial to develop innovative solutions using AI and other technologies to facilitate early detection and effective management of diabetes and its complications. The past research work being done using different generative architecture especially using Generative Adversarial Network (GAN) (Goodfellow et al., 2014) for analysing diabetic retinopathy using fundus images is a step in the right direction and shown a possibility to make an impact on the medical industry.

Diabetic Retinopathy is a growing problem worldwide, and researchers have an important role to play in developing solutions that can help prevent, manage, and treat the disease. AI and ML technologies offer a promising avenue for researchers to explore, as they can potentially improve the accuracy and efficiency of diagnosis, treatment, and monitoring of diabetes and its associated complications. Generative models, as detailed in a study (Gm et al., 2020), possess a unique capability to understand the distribution of training data and can generate new data samples within the same distribution. They are versatile in generating various types of media data, including text, audio, and images. Leveraging this novel data generation capability, researchers have extensively explored generative AI to address diverse and intricate problems, with anomaly detection being a notable application. In healthcare, many challenges can be addressed using generative AI architecture (Takyar, 2023). DR, which relies on retinal images for detection, has extensively utilized generative AI models for precise and accurate identification. Generative-AI methods have been employed to highlight lesion marks in retinal images. Although continuous improvement contributes to better detection and identification of minor lesions, there remains ample opportunity for further advancements in this field.

By working in this domain, researchers can make a significant contribution to society by improving the health and well-being of millions of people affected by diabetes and can facilitate in achieving SDG's.

3. Literature screening and selection

A structured literature review and meta-analysis method were employed to conduct a descriptive analysis utilizing eligible study items. The systematic arrangement of the analysis aimed to categorize and assess existing publications comprehensively, covering the breadth of the study. The initial step in delineating the study topics involved precisely calculating the inclusion ratio of prior work. The availability of healthy patient data is easier than unhealthy patient data, hence anomaly detection can be a useful approach for identifying and characterizing abnormal cases in the patient population. That is the reason the chosen boolean keywords for the literature review align with this specific criterion. The search encompassed papers published between 2017 and 2023 (mid), utilizing the search string "[Retinopathy OR OCT] AND [Unsupervised OR Generative OR Anomaly]", refer to Table 1. Yearwise and Journal wise distribution of paper is presented in Fig. 3. Two datasets, namely WebofScience and Scopus, were explored,

Table 1
Dataset and Searched Keywords.

Dataset	Searched Keywords
Web of Science and Elsevier	[Retinopathy OR OCT OR Fundus] AND [Unsupervised OR Generative OR Anomaly]

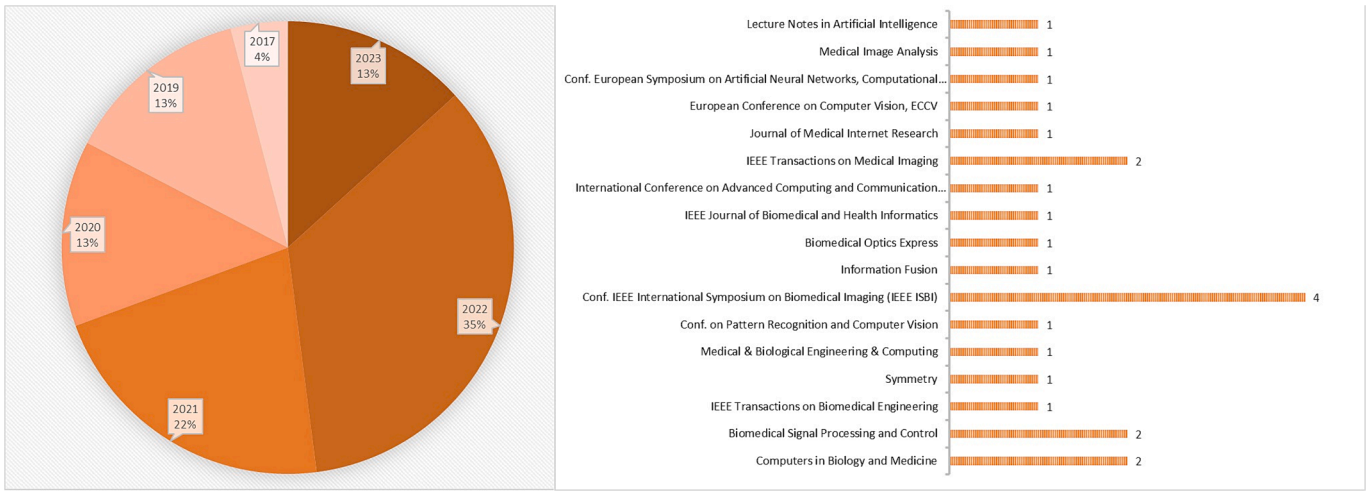

Fig. 3. Year-wise and Journal-wise Paper Distribution.

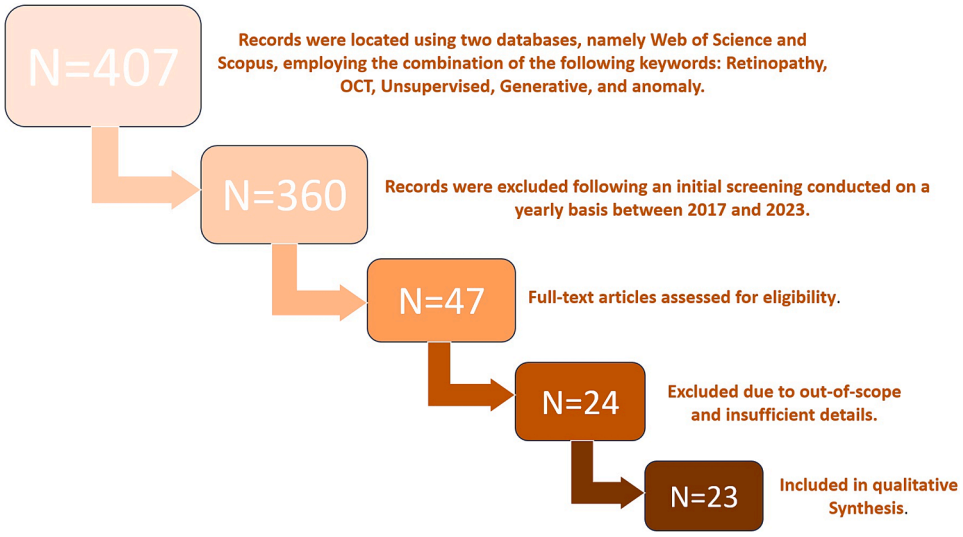

Fig. 4. Study selection for review.

resulting in a total of approximately 407 papers. These papers were further filtered as illustrated in Fig. 4.

This work is mainly focused on the Anomaly detection approach which is based upon a generative network instead of classifying the disease i.e., segregate the unhealthy records from the healthy records and facilitate the medical practitioners in their decision-making and explanation of the problem. Considering these criteria's, 23 relevant papers have been filtered out and considered for further studies.

In summary, the criteria used for including and excluding research papers are as follows:

Inclusion Criteria:

- Primary research papers are scientific papers.
- The articles considered were published between 2017 and mid-2023.
- The query terms were assessed within the titles and abstracts.

Exclusion Criteria:

- Papers in languages other than English.
- Scientific papers that are not accessible.
- Research papers unrelated to AI in diabetic retinopathy.

4. Literature review

Numerous studies within the field of artificial intelligence utilize fundus images, a medical imaging technique capturing the back part of the eye, encompassing the retina, optic disc, macula, and blood vessels, for image analysis. Fundus images generally portray a high-dimensional representation of the retina, and individuals affected by DR can develop lesions on the retina as mentioned on Fig. 1. Deep learning solutions have demonstrated efficient capabilities in detecting such lesions, enabling the segregation of images from healthy and unhealthy individuals. This facilitates medical practitioners in the detection of DR.

AI can address and manage the detection of DR by framing it as a binary classification problem, distinguishing between healthy and unhealthy images. The progress in technology allows for the development of large and intricate architectures. Generative-AI based structures, known for their complexity and reliance on substantial infrastructure, significantly contribute to solving image segregation challenges. This literature review specifically focuses on research involving Generative-AI for detecting DR.

Liu et al. (2023) discussed that healthy images have symmetricity which is broken by unhealthy images when some kind of liaison is formed. Researcher proposed a deep learning based solution that weakly

supervised biomarker identification and segmentation with image-level annotation in retinal fundus images. The model consists of two main components: a localization network and a segmentation network. The localization network is used to identify potential biomarker regions in the input fundus image, while the segmentation network is used to refine the boundaries of these regions and produce a final segmentation mask. The whole approach is divided into three phases; pre-training, fine-tuning, and distillation model.

During the pretraining phase, only normal images are employed. These normal images undergo augmentation involving the introduction of random noise, thereby labeling them as unhealthy images. Through the application of contrastive loss, an encoder is trained to acquire the features associated with normal images.

Fine-tuning involves the utilization of a pre-trained model, supplemented by an additional cross-entropy loss. The model undergoes further refinement by incorporating normal and biomarker images during the fine-tuning process.

In the distillation phase, the student network is trained using the teacher network. In this context, the pre-trained Encoder architecture serves as the teacher network, and an identical architecture is specified for the student network, forming a teacher-student model (Abbasi et al., 2020). The weights of the teacher network remain constant, while the weights of the student network undergo fine-tuning. Training exclusively employs normal images, enabling the student network to comprehend the features associated with normal images. As the teacher and student networks share the same architecture, the discrepancy at each layer is minimized during model training, and the accumulated difference is considered the total loss for optimization.

In the testing phase, when a retina image is input to the teacher model, the diverse feature presentations of the teacher model, trained on both normal and biomarker images, contrast with the perspective of the student network, which has been trained exclusively on healthy images. At each layer, the disparity serves to generate a heatmap, pinpointing the biomarker on an eye afflicted with Diabetic Retinopathy. Ultimately, the activation mask is employed to elucidate the specific problematic area within the image.

Liu et al. (2023), researchers have used an erasure strategy where part of an image is masked for fine-tuning. Whole approach is divided into two stages,

- The reconstruction network is designed to identify abnormal areas housing the biomarker.
- In the second stage, the classification task is centered on categorizing the biomarker level within the properly defined region.

The reconstruction module comprises an image generator based on GAN and a distinct encoder model. The generator module of GAN accepts input in the format of 128*128*1 to produce an image with dimensions of 128*128*1. The input real image is grayscale and shares the same dimensions (128*128*1). Throughout the generation process, the generator can observe either an actual image (limited to healthy images) or random noise with dimensions of 128*128*1. The resulting generated image is then sent to the discriminator to distinguish between real and fake images.

In the second stage, dedicated to classification, the classification network adopts a pretrained discriminator architecture. The last fully connected layer is substituted with 8 outputs to signify the 8 biomarker classes. During the backpropagation process, an attention map is obtained, aiming to pinpoint the biomarker region. This highlighted attention area is then removed, and the input is once again fed for classification prediction, with the objective of minimizing the probability associated with the biomarker class. Following the training of the GAN, the encoder training commences. The encoder processes input in the format of 128*128*3 and produces an output of 128*128*1. It takes an actual grayscale image, as well as light-to-dark and dark-to-light gradient versions of the image (Chiu et al., 2010), combining all three to form an input of 128*128*3. The output from the encoder is fed into the generator to generate an image. The residual difference between the actual image and the generated image serves as a basis for optimizing the encoder.

During the testing phase, the generator produces an image resembling a healthy one. The residual mask, calculated as the difference between the generated and actual image (with a specified threshold), aids in identifying abnormal areas. Through backpropagation in the classification model, a refined attention map is obtained along with the corresponding class. The combined output of the residual mask and attention map contributes to the precise detection of lesions.

Zhou et al. (2023), employing an ablation-based technique, which involves systematically removing or disabling specific components or features in a system or model to analyze their individual impact on performance, researchers aim to discern discrepancies between actual and generated images. The residual error resulting from this analysis is then utilized to categorize anomalies in the image. They have divided the overall process into 4 steps: A self-supervised segmentation network is implemented to eliminate background noise and identify the retinal region in normal + biomarker images. To segment the retinal region and remove the background, a straightforward pixel range-based method is initially employed, distinguishing between background and retinal portions which is based upon minimum and maximum pixel values. While effective for normal images, this approach encounters challenges with abnormal images characterized by inconsistencies in the retina. To address this, a UNet-based (Ronneberger et al., 2015) segmentation network is developed. Operating on data post-preprocessing (with dimensions h*w*2, where 2 signifies the upper and lower pixel value bounds for background differentiation), the segmentation process produces an image. This not only rectifies errors in anomaly scores due to background regions but also enhances the reconstruction quality of retinal areas.

The Spatial Variational Autoencoder (Kingma and Welling, 2019) (S-VAE) is employed to reconstruct input images, aiming to transform them into the nearest normal counterparts while maintaining their original retinal topology. An Encoder architecture is trained specifically for extracting a multidimensional latent space and preserving the spatial features of the input image. The multi-dimensional latent space is then fed into the decoder section to regenerate the original image. A Contextual Variational Autoencoder (C-VAE) is implemented to reconstruct input images, transforming them into images with normal retinal topology. Similar to S-VAE, C-VAE has been introduced. While the S-VAE effectively extracts features from normal images, it proves insensitive to abnormalities. In contrast, the C-VAE generates a one-dimensional latent space specifically crafted for anomaly detection by reconstructing a normal image from an abnormal one. The one-dimensional output of the C-VAE's latent space places greater emphasis on the global context of a normal retina. Finally, while the C-VAE aids in detecting anomalies through significant reconstruction errors, it has been observed to falter in cases of altered retinal topology. To address this limitation, an ablation-based method is employed to calculate anomaly attention maps, facilitating the localization of anomalous regions.

Zang et al. (2022), Researchers have delved into the concept of cycle consistency, a notion frequently linked with tasks involving image-to-image translation. This concept revolves around maintaining consistency in the translation process from one domain to another and subsequently back to the original domain. It involves a binary classification model used to categorize diabetic retinopathy images as Positive/Negative. Subsequently, this model guides the fine-tuning process of the Biomarker Activation Models (BAM). Initially, a classification model is trained for DR detection. The design and training of the two BAM generation frameworks are conducted with assistance from this classification model. Together, these two BAMs constitute a cyclic GAN-type architecture (Zhu et al., 2017), with both architectures based on UNET.

The main generator is created to transform positive DR images into

negative DR (forged image) representations. Consequently, when this generated image is input into the classifier, it can be classified as non-DR. Whereas, if a negative DR image is supplied to this main generator, it exclusively produces the negative DR image (preserved image), ensuring that the classification model predicts non-DR for the generated image. The assistant generator is formulated to produce a preserved positive DR image when a positive DR image is provided as input. Conversely, when given a negative DR image, the assistant generator generates a forged positive DR image. This functionality is in contrast to the main generator.

Combining these two architectures results in a cyclic GAN, and distinct loss functions are applied to train the main and assistant generators to prevent overfitting. The biomarker is produced by taking the difference between the positive class and negative class generated using the main generator.

Hemamalini and Kumar (2022), they have trained a hybrid segmentation and classification model, exploring the application of Fuzzy clustering and skimpy regularization to mitigate over-segmentation. Fuzzy clustering is a technique in machine learning and image processing that extends traditional clustering methods by allowing for soft assignments of data points to multiple clusters, enabling each point to belong to more than one cluster. Skimpy regularization refers to a form of regularization that selectively applies constraints, potentially using a minimal set of regularization techniques or parameters. Their hybrid segmentation technique incorporates clustering to group relevant pixels, where fewer pixels are assigned high membership values in a cluster, thereby preventing over-segmentation of lesions in images. This approach emphasizes only the pertinent pixels, enhancing segmentation accuracy. Additionally, a classifier model was developed, utilizing the output of the segmentation model for training and predicting diabetic retinopathy.

Naz et al. (2022) discussed and developed a solution to detect retinal abnormalities using fundus images by using a mix of unsupervised and supervised techniques. They trained an unsupervised segmentation model to learn image features and then used these features to train a custom CNN-based classification model. To consolidate relevant features, they used Fuzzy logic and k-means-based clustering techniques.

Li et al. (2022), A model based on image reconstruction has been trained for detecting anomalies in retinal images. The optimization process incorporates the use of the updated Spatial Similarity Retention Loss (SRL), a loss function explicitly crafted to maintain spatial relationships among features within images during deep learning tasks. SRL dynamically recognizes and prioritizes “hard” sample pairs, which are either excessively similar or dissimilar, throughout the training process. This targeted attention to challenging samples enhances feature discrimination and spatial consistency. By assigning higher weights to hard sample pairs, SRL ensures that they have a more substantial impact on the overall loss and contribute significantly to model optimization. It comprises three main components:

Nor-Net (Normal representation network): An Encoder-Decoder architecture designed for Normal OCT images to extract their features. Elastic Distortion Module: Utilized for generating Pseudo abnormal OCT images, serving as a form of image augmentation. They applied an elastic distortion technique (David et al., 2021) to the training data. Abr-Net (Abnormal representation network): Employs Pseudo-abnormal images to extract their features in the representation network.

The standard residual loss alone is inadequate, prompting the utilization of SRL, as specified in the image, for precise reconstruction and preservation of structural details. It discerns anomalies by evaluating deviations between the original and reconstructed images, usually employing a threshold-based method.

Mou et al. (2022), researchers aim to address the anomaly detection challenge in OCT images by employing custom-trained segmentation and classification models. The training data predominantly consists of healthy OCT images. Instead of opting for the conventional Deep Neural Network with fixed weights and biases, they have chosen to utilize a Bayesian Neural Network (BNN) (Wilson and Izmailov, 2020) coupled with the concept of Epistemic uncertainty. Epistemic uncertainty refers to a type of uncertainty linked to the model’s incomplete knowledge regarding the true underlying data distribution. In a BNN, uncertainty is associated with model parameters through probability distributions, treating weights and biases as probability distributions rather than fixed values. A key advantage of BNNs lies in their capacity to provide uncertainty estimates alongside predictions.

The methodology is segmented into three components:

A Bayesian neural network (BNN) architecture referred to as Multi-scale Bayesian U-Net (MBU-Net) is developed to acquire the epistemic uncertainty of OCT images. An algorithm for borderline uncertainty filtration is devised to diminish uncertainty estimation in healthy regions of OCT images. Ultimately, a threshold-based function is created to determine whether the input data is indicative of health or pathology.

The comprehensive architecture incorporates five UNET-based segmentation models (multi-scale models) trained at varying dropout values. To align with the Bayesian property, dropout is applied during prediction, optimizing resource utilization. The outputs from all five segmented models are employed, and a customized borderline filtration mechanism is implemented for accurate segmentation, eliminating uncertainty in border-line regions. Subsequently, thresholding-based classification is performed on the images.

Huang et al. (2022), They employed an unsupervised anomaly detection method designed for fundus images, with the objective of detecting and segmenting abnormal areas in the images without relying on manual annotation or ground truth data. Considering that healthy images exhibit pixel consistency, which can be disrupted by any lesion, the researchers exclusively utilized healthy images. Portions of the image grid were masked, and an UNET-based image reconstruction network was employed. Additionally, a separate Discriminator CNN-based architecture was used to optimize the overall model, reconstructing the input image from a corrupted version of the same image. The network learns to predict the missing patch, aiding in understanding the underlying structure of the image. The trained architecture excels in generating healthy images, a task challenging in the presence of lesions in the masked images, resulting in increased residual error. Thresholding based on residual error is employed to categorize images as healthy or unhealthy.

Hervella et al. (2022), researchers have utilized the fundus images for highlighting Microaneurysms (tiny abnormal blood vessels in the retina) and representing them as a heat map. The overall process is divided into three steps:

An UNET-based GAN model is trained, where the goal is to use fundus images as input and generate Fluorescein Angiography (FA) images. Consequently, the training requires fundus images paired with their corresponding FA images. Once this model is trained, a pretrained UNET architecture is employed for further fine-tuning specifically for microaneurysm detection. A ground truth microaneurysm heatmap serves as a label during training, with the objective of utilizing fundus images as input to generate the heatmap of microaneurysms. The maximum value in the generated heatmap can be indicative of the presence of microaneurysms.

Dipta et al. (2022), researchers employed a generative approach based on thresholding to classify healthy and unhealthy fundus images, relying on the reconstruction error. In their approach, they incorporated the concept of sparse coding. Sparse coding is a technique utilized for efficiently representing data and capturing its essential features. The objective is to identify a compact representation of data using a small number of non-zero components. This technique finds applications in image compression, denoising, restoration, and feature extraction.

The overall process is divided into two steps: An AutoEncoder is trained on healthy normal retinal images to extract the features of a normal image. Multi-scale deep feature sparse coding is implemented, with the central idea for anomaly detection using sparse coding being to reconstruct the image using the dictionary (weights) learned on normal

images and subsequently thresholding the residual error. Multi-scale features are extracted from the trained encoder part, encompassing features from different encoder layers to aid in extracting both fine-grained and global contextual features. These multi-scale features contribute to the formation of sparse coding, with the objective of extracting receptive fields (part of input image contributing in certain predictions, attention map of an image) at different levels for anomaly detection.

Wang et al. (2021), A generative-based approach was employed to detect anomalous fundus images utilizing Cyclic-GAN (Zhu et al., 2017). The objective is to generate normal fundus images from abnormal ones, and subsequently, the residual error between them is utilized to identify lesion marks and highlight fluid and exudation areas in the image. The model incorporates the concept of Dilated Convolution. Dilated Convolution enables the network to capture larger contextual information while preserving the output size, akin to viewing an image with a "wider lens" without actual zooming. This convolution inserts "holes" or "gaps" between the elements of the filter, effectively increasing its size without adding more parameters, all while maintaining the same output size as the input or even expanding it. A tailored network, comprising a dilated convolutional block-based discriminator combined with a UNET generator and a multi-scale structural similarity perceptual reconstruction loss (MS-SSIM), was implemented to address this challenge. This customized architecture has the capacity to capture more global structural variabilities, while the MS-SSIM can represent geometric differences using area statistics, aiding the model in emphasizing structural changes.

The architecture operates as a cyclic GAN, featuring two generator modules: To generate a normal image from an abnormal one. To generate an abnormal image from a normal one. Together, these modules constitute a cyclic GAN. During training, the connection between these two modules occurs in sequence depending on the type of input data.

Niu et al. (2022) have used an autoencoder architecture to identify key activation neurons that are directly related to disease prediction. The autoencoder consists of two parts: the DR Detection NET, which serves as the encoder, and the Activation Net, which serves as the decoder. The DR Detection NET takes in retinal OCT and fundus images and compresses them into a low-dimensional representation (or bottleneck layer). This bottleneck layer is fused with another dense layer to predict the severity labels of the Diabetic Retinopathy. The Activation Net is essentially a reverse architecture of the DR Detection NET, with the same layers and parameters but in reverse order. The Activation Net takes in the same low-dimensional representation as the bottleneck layer and reconstructs the original image, but its main focus is on identifying the activated neurons that are directly related to disease prediction. The researchers applied Gaussian blur to the last layer of the Activation Net in order to obtain a binary mask of different lesions. This binary mask highlights the areas where the key neurons are activated and can be used to identify the location and severity of the lesions in the retinal fundus images. By using the architecture trained above to guide the Patho-GAN, the researchers were able to generate synthetic images that are specifically tailored to the features of diabetic retinopathy.

Kumar and Bindu (2021), segmentation architecture based on AutoEncoder has been trained to identify lesions in fundus images. To generate the Ground Truth (GT) for the lesions, the researchers convert the retinal image into grayscale, assigning pixel values less than 50 as 0 and the rest as 1. This binary segmentation is treated as the GT. The dataset undergoes preprocessing using Contrast Limited Adaptive Histogram Equalization (Pizer et al., 1990) (CLAHE), an algorithm designed to improve contrast. Instead of processing the entire image at once, a tile-based approach is adopted. The image is divided into multiple tiles, and the GT tiles are accordingly split. This approach enables the network to concentrate on specific regions of the image, capturing finer details of the lesions and potentially enhancing segmentation accuracy. The UNet architecture is trained on these tiles for lesion detection. The predictions from individual patches are then amalgamated to create the final lesion segmentation map for the entire image.

Zhao et al. (2021), researchers have proposed and devised a technique for anomaly detection in medical images, employing a GAN based architecture to comprehend the distribution of images from healthy individuals. Abnormalities are detected through the residual outcome of the image and the latent features of GAN models. They have incorporated the concept of Translation Consistency, which denotes features or representations that remain constant even when an image is translated or shifted. In this context, translation implies moving an object or pattern within an image without altering its content.

The overall methodology is segmented into four steps:

1. Train an Encoder-Decoder based GAN architecture to grasp the distribution of normal images. 2. Feature-Space, train another GAN architecture using the latent features obtained from the aforementioned encoder. This architecture resembles a Decoder-Encoder setup where the goal is to generate the same latent space, adhering to the concept of Translation Consistency. A separate discriminator is employed for this GAN. 3. Self-Supervised Learning Module, detecting anomalies based on the normal dataset is challenging. To enhance robustness, an Encoder-Decoder architecture is trained. The input image is perturbed, and this architecture is expected to generate the original image. The same Encoder-Decoder as mentioned in step 1 is used, with the encoder sharing the same weights. 4. During inference, the residual error between the generated image and latent space feature is amalgamated to generate an anomaly score.

Han et al. (2021), researchers trained a GAN-based model and showcased the efficacy of this approach for abnormality detection through their results. An anomaly is flagged if the generated image significantly deviates from the input image. A straightforward UNET-based Encoder-Decoder Architecture is employed alongside a discriminator. By utilizing normal images during training, the objective is to understand the distribution of normal images, aiding in anomaly detection during the testing phase.

Zhou et al. (2020), researchers have devised an architecture named P-Net, which is a neural network specifically crafted to extract attributes pertaining to the structure and texture of an image. They utilized an existing dataset of retina images containing structural information such as vessel details and veins segmentation information. Employing the concept of domain adaptation, where the model is trained on one dataset (source) and evaluated on a different dataset (target) with distinct data distribution and characteristics, proved beneficial. This approach is particularly advantageous when the target domain data is limited or expensive to obtain, but there is an abundance of data available in the source domain. The entire process is segmented into three modules:

1.Structure Extraction Network: A GAN network is trained to extract the structural information from retina images. 2. Another GAN based reconstruction GAN-based network is trained for retina image generation. In this module, the input is the structural output from the previous step. The latent space is fused with, and the last layer encoder output from step 1 is combined before being fed into the decoder component. The last layer's output represents the texture features of the input image. 3. The generated retina image is once again fed into the Structure Extraction module, and the generated structural information is compared with the structural information obtained in step 1 for anomaly detection. Residual error should be high if data consistency is broken on the input image.

Zhang et al. (2020), an Encoder-Decoder-based GAN architecture is trained using healthy OCT images. The Generator module incorporates a memory module, functioning as an embedding mechanism used to represent patterns of normal images. The memory module is structured as a 2D matrix that stores extracted features from normal OCT images. This matrix is dynamically updated during training to capture variations in normal patterns. The latent feature z of the Encoder is obtained through a memory addressing strategy based on the cosine similarity of the query z and the memory item in the memory module. The Discriminator module follows the Generator-Encoder architecture with

slight modifications. The anomaly score is determined based on the residual value, and a thresholding value is applied to categorize healthy and unhealthy images.

Zhou et al. (2019), anomaly detection framework is applied to retina images utilizing an Encoder-Decoder based GAN. The model is exclusively trained on healthy retina images. Rather than relying on image residual error for anomaly detection, they employ the difference of latent vectors for detecting anomalies.

The overall process can be segmented into three parts: 1. Sparse-GAN, an Encoder-Decoder based GAN architecture is employed to generate OCT images. 2. Instead of directly utilizing the encoder-latent vector through the decoder for image generation, the latent vector is regularized with a novel sparsity regularizer termed as Sparsity Regularizer-net. This regularizer is implemented to control the sparsity of latent features, with the goal of focusing on key features. 3. To address the impact of speckle noise (a type of noise that can occur in digital images), another Encoder has been trained with the same SPARSE-GAN encoder architecture to extract the latent features of the generated image. The residual error between the latent features of the actual image (Sparse-GAN) and the generated image (Another Encoder) aids in detecting anomalies.

Additionally, an explanatory framework is developed using the Anomaly Activation Map to visualize and illustrate lesions in the abnormality detection framework. This framework relies on the residual error of Global Average Pooling of the latent space of the original and generated image.

Sutradhar et al. (2019), A blind-spot network is trained using an adversarial training technique, wherein the network is taught to differentiate between genuine fundus images and synthetic images generated by a GAN. The synthetic images are intentionally crafted to include subtle anomalies that may pose challenges for human experts in detection. The training exclusively utilizes data from healthy patients to understand the distribution of normal images. The Blind-Spot network is trained for image generation, addressing the issue where autoencoders might not efficiently learn meaningful semantic features of the input distribution, often leading to blurring in the reconstructed output. The data of normal images for training is partitioned into a 3*3 grid. The surrounding eight grids are fed to the AutoEncoder architecture, while the middle grid remains as is to generate the central patch. This approach leverages the consistency in normal images, making it easier to generate the middle grid. This consistency, however, is not applicable to abnormal images, resulting in a higher residual error. Each input segment is processed separately with independent convolutional neural networks, and their outputs are concatenated. These concatenated outputs are then fed into a common fully-connected network responsible for generating the target center patch.

Seebock et al. (2020), researchers employed a Bayesian-based generative deep learning architecture for anomaly detection in fundus images. The architecture utilized in this study was based on the UNET-based Bayesian deep learning model. The anatomical structure of normal images carries implicit information that distinguishes it from abnormal information. During training, the model was exclusively exposed to healthy images to mitigate epistemic uncertainty stemming from a lack of training data for abnormal images. Bayesian deep learning was implemented in this model to learn the distribution of weights rather than using hard weights, aiding in capturing the mean and variance of each weight. The researchers did not directly utilize the residual error of the generated image. Instead, to generate uncertainty maps, Monte Carlo (Gal and Ghahramani, 2016) dropout was applied during prediction. Dropout is a regularization technique that randomly removes some nodes during training to prevent overfitting. Monte Carlo dropout involves running multiple simulations with random dropout, and then taking the mean of these simulations to generate the uncertainty map. Epistemic uncertainty increases if the image for inference differs significantly from the training data, indicating an anomaly. Finally, a thresholding technique along with a novel post-processing method was employed on the uncertainty map to generate a segmentation map of the lesion in the fundus image. This segmentation map can aid in identifying and diagnosing abnormalities in medical images.

Schlegl et al. (2019), researchers have used a GAN based generative approach for Anomaly detection in fundus images and to classify anomalous images and to segmenting the anomalous region. They used a traditional Wasserstein GAN (Arjovsky et al., 2017) architecture and performed unsupervised learning on health images only to learn the normal variability present in normal data.

The overall process is segmented into two main parts: 1. GAN Training 2. Encoder Training Encoder training commences after the completion of GAN model training and is further divided into two components: 1. Izi (Image to Latent Space to Image): This involves an AutoEncoder architecture where a new Encoder layer is connected to a pretrained Decoder layer (the image generation layer of GAN). In this phase, the weights of the Encoder layer are optimized while keeping the weights of the Decoder layer fixed. The optimization is driven by the residual loss (input-generated image) and the loss from the Discriminator module. The Discriminator module exclusively uses the image features (removing fully connected layers) from both the generated and actual images, and the residual is utilized for optimization. Thus, both these losses collectively optimize the weights of the Encoder. 2. Ziz (Latent Space to Image to Latent Space): This involves an AutoEncoder-based architecture where the GAN generator is connected to the Encoder from the previous module. In this setup, a random latent input is fed to generate an image, and the Encoder's task is to take the generated image and reconstruct the same latent space. The residual error of the latent space is used to optimize only the Encoder (with the weights of the generator module fixed). It's important to note that in this context, the GAN generator module behaves as a latent space to image, and the Encoder module is treated as image to latent space. Anomaly scoring is used to classify the anomalous healthy and non-healthy fundus images.

Schlegl et al. (2017) have used deep convolution generative adversarial network to find out the anomaly score and thresholding of this score helps in detecting the anomaly on the fundus image. This is one the initial approach which has used DCGAN (Radford et al., 2016) model for image generation. An unsupervised image generation technique based on GAN is employed for anomaly detection. The distribution of normal images is learned using normal retinal images. Prior to training, an image undergoes preprocessing, involving the flattening of the retinal part and extraction of patches. Instead of utilizing the entire image, multiple random patches are created and employed for model training.

The overall process is delineated into the following steps: 1. Traditional GAN Setup: A conventional GAN is established where a Decoder-based image generation module exists. This module takes a random input (acting as latent space) for image generation. An Encoder-type module is employed for the Discriminator. 2. Inverse Mapping Challenge: The aforementioned GAN model lacks the capability to yield an inverse mapping, i.e., from image to latent vector. Therefore, for inference on an image X, the objective is to find the closest z (latent vector). Considering the smooth transition in the latent space, where two different close latent space points can generate visually similar images, a random z is fed to the Generator. The generated image is then compared with the input image X used for inference. A loss function is defined based on the generated image, aiding in moving z closer to the latent space of image X. This process is iteratively repeated for T steps, resulting in the best z that is close to the input image X. 3. Discriminator Analysis: The Discriminator is applied to both the images X and the image generated with the best z value. Instead of using fully connected output, the feature map is extracted without fully connected layers. Differences between both outputs (Discriminator output with X and z) determine whether the image has anomalies or not. The residual error is also utilized to highlight the anomalous region as an explanation.

After detailing the methodologies employed by each paper, the summary of the literature review is encapsulated in Table 2 below, aiming to encompass the essential features of each paper.

Table 2
Dataseta and Searched Keywords.

Paper	Dataset	Technique	Results	Remarks/Improvement
Liu et. al. (Liu et al., 2023)	Kermany's dataset (Kermany et al., 2018), Dataset from Wuhan Aier Eye Hospital	Encoder based architecture and Teacher-student (Abbasi et al., 2020) based distillation system is used.	AUC 0.98, Dice coefficient 0.50	The model lacks the generality required to accommodate various types of biomarkers.
Liu et. al. (Liu et al., 2023)	Publicdatasets (Kermany et al., 2018; Kermany et al., 2018)and Wuhan Aier Eye Hospital	GAN based Encoder-Decoder	Healthy images, Dice-coefficient is 0.92 and IoU is 0.78, Unhealthy images, Dice coefficient is 0.50 and IoU is 0.32	The output levels of the model exhibit inconsistency when subjected to various types of input.
Zhou et. al. (Zhou et al., 2023)	Cirrus	Classification model based upon Variational AutoEncoder architecture.	Accuracy 0.8692 ± 0.0107	Interpretability of the model results is missing and training computational cost.
Zang et. al. (Zang et al., 2022)	Custom dataset	UNET based GAN architecture trained using pretrained VGG19.	F1 score 0.63 (+-0.08) and the recall value is 0.65 (+- 0.08)	Very small dataset used whereas BAM based model performance relies heavily on the quality and size of the training data.
Hemamalini et. al. (Hemamalini and Kumar, 2022)	Messidor (E. Decenciere et al. 2014) and iDRiD (Porwal et al., 2018)	Image segmentation with clustering using Auto-Encoder architecture and CNN based classification model.	Accuracy 0.92 to 0.95	The performance of trained model might sensitive to the chosen parameters for outlier detection, regularization, and other components more generic optimization using different dataset can solve this problem
Naz et. al. (Naz et al., 2022)	DIARETDB1 (Kauppi et al., 2007); APTOS (APTOS, 2019); Liverpool	Image Segmentation and classification model used for anomaly detection.	Accuracy 0.98	Lacking in results interpretability
Li et. al. (Li et al., 2022)	ZhangLabData (Kermany et al., 2018) and OCTA500 (Li et al., 2020)	Encoder-Decoder model	Accuracy 0.91, AUC 0.96	Synthetic data consumed by model, further validation across diverse datasets and clinical scenarios is needed.
Mou et. al. (Mou et al., 2022)	UCSD(Kermany et al., 2018) Kermany (Kermany et al., 2018); and BFHJLU private dataset	Bayesian based UNET architectures.	Accuracy 0.94, Sensitivity 0.97	Monte Carlo sampling is computation- ally expensive and lacking in interpretability due to model uncertainty.
Huang et. al. (Huang et al., 2022)	EyeQ (Fu et al., 2019)	UNET based GAN architecture.	AUC value of 0.87	Scope of improvement in accuracy and generalizability across different grade of DR is required.
Hervella et. al. (Hervella et al., 2022)	MISP(Waniewski et al., 2012)E-Ophtha (Decenciere et al., 2013); ROC (Niemeijer et al., 2010); DDR (Li et al., 2019)	UNET based GAN model	AUROC of approx. 0.97	Due to dataset variability model performance of the model is not generalized among different datasets.
Dipta et. al. (Dipta et al., 2022)	Eye-Q (Schmidt-Erfurth et al., 2018), IDrid (Porwal et al., 2018), OCTID (Gholami et al., 2018)	Encoder-Decoder architecture where VGG-16 used as encoder backbone	AUC approx. 0.82	Sparse coding is computationally ex- pensive and lacking in interpretability of the results
Wang et. al. (Wang et al., 2021)	K's (Kermany et al., 2018) Chiu's dataset (Chiu et al., 2015)	UNET based Cyclic-GAN architecture	Accuracy0.96, Dice coefficient 0.8239	The model lacks generalization for lesion location across the retina, especially when the number of lesions is high
Niu et. al. (Niu et al., 2022)	SA-UNET(Guo et al., 2021); IDRid(Porwal et al., 2018); FGADAR (Zhou et al., 2021)	UNET based GAN architecture is used where VGG19 is used as backbone	MSE of 0.0114 across 3 datasets	Prioritizing severity detection of the lesion while overlooking the identification of crucial lesions.
Kumar et. al. (Kumar and Bindu, 2021)	IDRiD (Porwal et al., 2018)	UNET based architecture is used for image segmentation	Accuracy 0.99, Sensitivity 0.92	Enhancing the sensitivity of diabetic retinopathy detection is possible
Zhao et. al. (Zhao et al., 2021)	OCT and Chest X- Ray (Kermany et al., 2018)	Encoder-Decoder based GAN architecture	Accuracy 0.90, AUC 0.96	While the framework works for general anomaly detection, further research could explore adapting it for specific types of anomalies for improved accuracy
Han et. al. (Han et al., 2021)	ODIR-5 k(Ocular Disease Intelligent Recognition ODIR-5K, 2019) Messidor (Decenciere et al., 2014);JSIEC (Cen et al., 2021) and one private dataset	UNET based GAN architecture	AUC 0.89	Understanding the specific features the model relies on for anomaly detection remains a challenge
Zhou et. al. (Zhou et al., 2020)	RESC (Hu et al., 2019); iSEE (Yan et al., 2019)	UNET based GAN architecture, AUC 0.92 and 0.72 across two datasets	The model's generability across different levels of lesions in retinal images can be enhanced, and	

(*continued on next page*)

Table 2 (*continued*)

Paper	Dataset	Technique	Results	Remarks/Improvement
			its results exhibit significant variability across datasets	
Zhang et. al. (Zhang et al., 2020)	Wuhan Aier Eye Hospital	Encoder-Decoder based GAN architecture with memory module	AUC 0.87	Further research can be done to under- stand how the model utilizes the memory module and makes decisions about anomalies
Zhou et. al. (Zhou et al., 2019)	Kermany's dataset (Kermany et al., 2018)	Encoder-Decoder Based GAN architecture	AUC 0.92, Accuracy 0.84	The accuracy level can be further explored for improvement
Sutradhar et. al. (Sutradhar et al., 2019)	Messidor (Decenciere et al., 2014)	AutoEncoder based architecture with Patch based approach	.	Lack of interpretability can raise concerns about bias or errors in the model's decision-making process
Seebock et. al. (Seebock et al., 2020)	Datasetof Heidelberg Engineering Germany	UNET based Bayesian deep Learning architecture (Wilson and Izmailov, 2020)	DICE index of 0.789	Over segmentation of the lesion in retina image is the potential issue and could leads to False positive
Schlegl et. al. (Schlegl et al., 2019)	Private dataset	Wasserstein GAN (Arjovsky et al., 2017) based generative model	AUC 0.78, Sensitivity 0.8	Approach for calculating False positive can be explored further. Sensitivity can be improved
Schlegl et. al. (Schlegl, 2017)	.	DCGAN (Radford et al., 2016) based technique	AUC 0.89, Sensitivity 0.7	Inference approach is very slow and needs many iterations on the generator for single instance

5. Challenges found in the literature

- Medical datasets are often skewed towards healthy patients. In the context of medical imaging, anomaly detection techniques, used to identify and classify abnormal images. Such techniques may not provide a clear explanation or reasoning for why a particular image is classified as abnormal. This lack of interpretability can be a significant concern in clinical settings, where it is essential to understand the underlying reasons for such a detection made by the AI models.
- Lesion segmentation-based models can also have limitations, particularly in cases where the lesions are not well-defined or have ambiguous boundaries. In such cases, the model may over-segment the image, identifying regions as abnormal that are actually healthy tissue or noise. This can lead to FP (false positives) and reduce the accuracy of the model.
- Generative-based approaches have shown promise in detecting anomalies in medical images, but they may struggle to identify minor lesions or subtle abnormalities that are not easily distinguishable from normal tissue. Improving the sensitivity and specificity of these models will be an important area of research going forward.

6. Conclusion and future directions

Various generative architectures have been extensively researched due to its ability to generate realistic synthetic data especially various variants of GANs. GANs consist of two neural networks, a generator network, and a discriminator network, that work together in a game theoretic framework to generate synthetic data that is difficult to distinguish from real data. It has a wide range of applications, including image and video synthesis, data augmentation, image-to-image translation, super-resolution, and anomaly detection.

The idea behind using generative model for anomaly detection is to train it on normal images and then use the generator network to generate new images. Any generated images that deviate significantly from the normal images can be flagged as anomalies. To use generative model for anomaly detection in medical imaging, a dataset of normal images or healthy person must be collected and used to train it. The trained model is then used to generate new images, and any generated images that are significantly different from the normal images can be flagged as anomalies.

Anomaly detection using generative architectures has the potential to improve medical imaging by providing a more automated and accurate method for detecting abnormalities (Liu et al., 2023; Li et al., 2022; Huang et al., 2022; Hervella et al., 2022; Dipta et al., 2022; Wang et al., 2021; Niu et al., 2022; Zhou et al., 2020; Zhang et al., 2020; Zhou et al., 2019; Schlegl et al., 2019). This can lead to earlier detection and treatment of diseases, which can ultimately save lives.

However, utilizing various generative architectures for anomaly detection in medical imaging presents certain challenges, including the complexity of analyzing smaller lesions. Nevertheless, ongoing research endeavors aim to enhance these generative models. The emergence of Diffusion models (Yang et al., 2023) and vision transformers (Dosovitskiy et al., 2010), coupled with their promising outcomes, offer opportunities for further exploration in generating architectures for anomaly detection in medical imaging. This research is centred to improve the generative models for lesion detection and present a technique which can be used to detect minor lesion which are overlooked by the current techniques and perform a comparative study with the existing models.

With the invent of such tools, it can facilitate the professional like ophthalmologist by aiding in their decision-making and can facilitate the treatment in right direction.

7. Ethical and informed consent for data used

Participants were informed of the research objectives and the intended use of their data, providing voluntary consent with the freedom to withdraw at any point. Their data was securely maintained, with identifying information removed, and the research upheld their rights and dignity while adhering to ethical standards throughout all procedures.

CRediT authorship contribution statement

Vipin Bansal: Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Amit Jain:** Writing – review & editing. **Navpreet Kaur Walia:** Writing – review & editing.

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.

Data availability

No data was used for the research described in the article.

References

a. v. I. V. A. a. T. L. G. H. C. o. G. E. H. VLEG/GBD 2020 model, Eye Health: Why it matters, IAPB Vision Atlas, 2021.

Abbasi, S., Hajabdollahi, M., Karimi, N., Samavi, S. 2020. Modeling teacher-student techniques in deep neural networks for knowledge distillation. In: 2020 International Conference on Machine Vision and Image Processing (MVIP), Vols. 2166-6784, pp. 1-6.

APTOS (2019) diabetic retinopathy dataset, [Online]. Available: https://www.kaggle.com/c/aptos2019-blindness-detection/data.

Arjovsky, M., Chintala, S., Bottou, L. 2017, Wasserstein GAN.

C. f. D. C. a. Prevention, 2022. National Diabetes Statistics Report, Centers for Disease Control and Prevention.

Cen, L.-P., Ji, J., Lin, J.-W., Ju, S.-T., Lin, H.-J., Li, T.-P., Wang, Y., Yang, J.-F., Liu, Y.-F., Tan, S., Tan, L., Li, D., Wang, Y., Zhang, M., Pang, C.P., Chen, W., Chen, H., Ng, T.K., Huang, Y., Zhang, G., 2021. Automatic detection of 39 fundus diseases and conditions in retinal photographs using deep neural networks. Nature Commun. 12 (1).

Chiu, S.J., Li, X.T., Nicholas, P., Toth, C.A., Izatt, J.A., Farsiu, S., 2010. Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation. Optica Publishing Group 18, 19413–19428.

Chiu, S.J., Allingham, M.J., Mettu, P.S., Cousins, S.W., Izatt, J.A., Farsiu, S., 2015. Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema. Biomed. Opt. Express 6 (4), 1172–1194.

Chua, J., Sim, R., Tan, B., Wong, D., Yao, X., Liu, X., Ting, D., Schmidl, D., Ang, M., Garhofer, G., 2020. Optical coherence tomography angiography in diabetes and diabetic retinopathy. J. Clin. Med. 9 (6), 1723.

D. B. David , L. B. David, Y. Shapira , R. Leibu , D. Dori , R. Schneor, A. Fischer, S. Soudry, 2021. Elastic Distortion Transformation on an image.

E. Decenciere,‘ G. Cazuguel, X. Zhang, G. Thibault, J. Klein, F. Meyer, G. Quellec, A. Chabouis and Z. Viktor, 2013. TeleOphta: Machine learning and image processing methods for teleophthalmology, IRBM, 34(2) 196-203.

Diabetic Retinopathy, Diabetic Retinopathy and Retinopathy of Pre-maturity, 2018.

Dipta, S.D., Dutta, S., Shah, N.A., Mahapatra, D., Ge, Z., 2022. Anomaly detection in retinal images using multi-scale deep feature sparse coding. In: IEEE 19th International Symposium on Biomedical Imaging (ISBI), pp. 1–5.

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, 2020. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,” arXiv preprint arXiv: 2010.11929.

E. Decenciere,‘ X. Zhang, G. Cazuguel, B. Lay, B. Cochener, C. Trone, P. Gain, R. Ordonez, P. Massin, A. Erginay, B. Charton and J.-C. Klein,, 2014. FEEDBACK ON A PUBLICLY DISTRIBUTED IMAGE DATABASE: THE MESSIDOR DATABASE. Image Analysis and Stereology 33 (3), 231–234.

Fisher, M.D., Rajput, Y., Gu, T., Singer, J.R., Marshall, A.R., Ryu, S., Barron, J., MacLean, C., 2016. Evaluating adherence to dilated eye examination recommendations among patients with diabetes, combined with patient and provider perspectives. Am. Health Drug Benefits 9 (7), 385–393.

H. Fu, B. Wang, J. Shen, S. Cui, Y. Xu, J. Liu, L. Shao, 2019. Evaluation of Retinal Image Quality Assessment Networks in Different Color-Spaces, 48–56.

Y. Gal, Z. Ghahramani, 2016. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. In: ICML’16: Proceedings of the 33rd International Conference on International Conference on Machine Learning.

Gholami, P., Roy, P., Parthasarathy, M.K., Lakshminarayanan, V., 2018. OCTID: Optical Coherence Tomography Image Database. Comput. Vision Pattern Recogn. 81, 106532.

Gm, H., Gourisaria, M.K., Pandey, M., Rautaray, S.S., 2020. A comprehensive survey and analysis of generative models in machine learning. Computer Science Review 38 (1574–0137), 100285.

Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. ”generative Adversarial Networks no, 2661.

Guo, C., Szemenyei, M., Yi, Y., Wang, W., Chen, B., Fan, C., 2021. SA-UNet: Spatial Attention U-Net for Retinal Vessel Segmentation. In 2020 25th International Conference on Pattern Recognition (ICPR).

Han, Y., Li, W., Liu, M., Wu, Z., Zhang, F., Liu, X., Tao, L., Li, X., Guo, X., 2021. Application of an Anomaly Detection Model to Screen for Ocular Diseases Using Color Retinal Fundus Images: Design and Evaluation Study. J. Med. Internet Res. 23 (7).

Hemamalini, S., Kumar, V.D., 2022. Outlier Based Skimpy Regularization Fuzzy Clustering Algorithm for Diabetic Retinopathy Image Segmen-tation. Computer Science and Symmetry/asymmetry 14 (12).

W. H. Herman, W. Ye, S. J. Griffin, R. K. Simmons, M. J. Davies, K. Khunti, G. . E. Rutten, A. Sandbaek, T. Lauritzen, K. Borch-Johnsen, M. B. Brown and N. J. Wareham, ”Early Detection and Treatment of Type 2 Diabetes Reduce Cardiovascular Morbidity and Mortality: A Simulation of the Results of the Anglo-Danish-Dutch Study of Intensive Treatment in People With Screen-Detected Diabetes in Primary Care (ADDITION-Europe).,” Diabetes Care., vol. 38, no. 8, pp. 1449-55, (2015).

Hervella, A.S., Rouco, J., Novo, J., Ortega, M., 2022. Retinal microa-neurysms detection using adversarial pre-training with unlabeled multi-modal images. Information Fusion 79, 146–161.

Hu, J., Chen, Y., Yi, Z., 2019. Automated segmentation of macular edema in OCT using deep neural networks. Med. Image Anal. 55, 216–227.

Huang, Y., Huang, W., Wenhao, L., Xiaoying, T., 2022. Lesion2void: Unsupervised Anomaly Detection in Fundus Images. In: International Symposium on Biomedical Imaging (ISBI), pp. 1–5.

I. D. Federation, 2021. Diabetes around the world in 2021. IDF Diabetes Atlas.

India: the Global Burden of Disease Study 1990–2016, Lancet Glob. Health. 2018.

Kalakota, R., Davenport, T., 2019. The potential for artificial intelligence in healthcare. Future Healthcare Journal 94–98.

T. Kauppi, V. Kalesnykiene, J.-K. Kamarainen, L. Lensu, I. Sorri, A. Raninen, R. Voutilainen, J. Pietila, H. Kalviäinen, H. Uusitalo, 2007. The DIARETDB1 diabetic retinopathy database and evaluation protocol, Proceedings of the British Machine Vision Conference, vol. 1, pp. 15.1-15.10.

D. Kermany, M. Goldbaum and K. Zhang, 2018. Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification.

Kermany, D.S., Goldbaum, M., Cai, W., Valentim, C.C., Liang, H., Baxter, S.L., McKeown, A., Yang, G., Wu, X., Yan, F., Dong, J., Prasadha, M.K., Pei, J., Ting, M.Y., Zhu, J., Li, C., Hewett, S., Dong, J., Ziyar, I., Shi, A., Zhang, R., Lewis, H.X., Zhang, K., 2018. Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell 172, 1122–1131.e9.

Kingma, D.P., Welling, M., 2019. An introduction to variational autoen-coders. Found. Trendsin Mach. Learn. 12 (4), 307–392.

Kumar, E.S., Bindu, C.S., 2021. Segmentation of retinal lesions in fundus images: A patch based approach using encoder-decoder neural network. 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS).

Lascar, N., Brown, J., Pattison, H., Barnett, A.H., Bailey, C.J., Bellary, S., 2017. Type 2 diabetes in adolescents and young adults. Lancet Diabetes Endocrinol. 6 (1), P69–P80.

Li, M., Chen, Y., ji, Z., Xie, K., Yuan, S., Chen, Q., Li, S., 2020. Image projection network: 3D to 2D image segmentation in OCTA images. IEEE Trans. Med. Imaging 39 (11), 3343–3354.

Y. Li, J. Li, H. Shen, Z. Chen, S. Yu, Z. Zhang, P. C. Yuen, J. Han, T. Tan, Y. Guo, J. Lai and J. Zhang, ED-AnoNet: Elastic Distortion-Based Unsupervised Network for OCT Image Anomaly Detection, Pattern Recognition and Computer Vision, pp. 3-15, (2022).

Li, T., Gao, Y., Wang, K., Guo, S., Liu, H., Kang, H., 2019. Diagnostic as-sessment of deep learning algorithms for diabetic retinopathy screening. Inf. Sci. 501, 511–522.

Liu, X., Liu, Z., Zhang, Y., Wang, M., Tang, J., 2023. Weakly-supervised localization and classification of biomarkers in OCT images with inte-grated reconstruction and attention. Biomed. Signal Process. Control 79 (1746-8094), 104213.

Liu, X., Liu, Q., Zhang, Y., Wang, M., Tang, J., 2023. TSSK-Net: Weakly supervised biomarker localization and segmentation with image-level annotation in retinal OCT images. Comput. Biol. Med. 153 (0010-4825), 106467.

Mathur, P., Mascarenhas, L., 2019. Life style diseases: keeping fit for a better tomorrow. Indian J. Med. Res. 149, S129–S135.

Mou, L., Liang, L., Gao, Z., Wang, X., 2022. A multi-scale anomaly detection framework for retinal OCT images based on the Bayesian neural network. Biomed. Signal Process. Control 75, 103619.

Naz, H., Nijhawan, R., Ahuja, N.J., 2022. An automated unsupervised deep learning–based approach for diabetic retinopathy detection. Med. Biol. Eng. Comput. 60, 3635–3654.

Niemeijer, M., Ginneken, B.V., Cree, M.J., Mizutani, A., Quellec, G., Muramatsu, C., Wu, X., Cazuguel, G., You, J., Mayo, A., Li, Q., Fujita, H., Garcia, M., 2010. Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs. IEEE Trans. Med. Imaging 29 (1), 185–195.

Niu, Y., Gu, L., Zhao, Y., Lu, F., 2022. Explainable diabetic retinopathy detection and retinal image generation. IEEE J. Biomed. Health Inform. 26, 44–56.

Ocular Disease Intelligent Recognition ODIR-5K, (2019.

Ogle, G.D., Wang, F., Gregory, G.A., Maniam, J., 2021. Type 1 Diabetes Numbers in Children and Adults. International Diabetes Federation, Brussels, Belgium.

Padma, V., Anand, N.N., Gurukul, S.M., Javid, S.M., Prasad, A., Arun, S., 2015. Health problems and stress in Information Technology and Business Process Outsourcing employees. Pharm. Bioallied Sci. 7, S9–S13.

Pizer, S.M., Johnston, R.E., Ericksen, J.P., Yankaskas, B.C., Muller, K.E., 1990. Contrast-limited adaptive histogram equalization: speed and effectiveness. [1990] Proceedings of the First Conference on Visualization in Biomedical Computing.

Porwal, P., Pachade, S., Kamble, R., Kokare, M., Deshmukh, G., Sahasrabuddhe, V., Meriaudeau, F., 2018. Indian Diabetic Retinopathy Image Dataset (IDRiD): A database for diabetic retinopathy screening research. Data 3 (3).

A. Radford, L. Metz, Chintala, S. 2016. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. In: 4th International Conference on Learning Representations, ICLR 2016.

O. Ronneberger, P. Fischer and T. Brox, 2015. U-Net: Convolutional Net-works for Biomedical Image Segmentation, pp. 234–241.

T. Schlegl, P. Seebock, S. M. Waldstein, U. Schmidt-Erfurth, G. Langs, M. Niethammer, M. Styner, S. Aylward, H. Zhu, I. Oguz, P.-T. Yap and D. Shen, 2017. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In: International Conference on Information Processing in Medical Imaging.

T. Schlegl, P. Seebock, S. M. Waldstein, G. Langs, U. Schmidt-Erfurth, 2019. f-AnoGAN: Fast unsupervised anomaly detection with gener-ative adversarial networks, Med. Image Anal., 54, 30-44.

Schmidt-Erfurth, U., Sadeghipour, A., Gerendas, B.S., Waldstein, S.M., 2018. Artificial intelligence in retina. Prog. Retin. Eye Res. 67, 1–29.

Seebock, P., Orlando, J.I., Schlegl, T., Waldstein, S.M., Bogunovic, H., Klimscha, S., Langs, G., 2020. Exploiting epistemic uncertainty of anatomy segmentation for anomaly detection in retinal OCT. IEEE Trans. Med. Imaging 39 (1), 87–98.

S. Sutradhar, J. Rouco, M. Ortega, 2019. Blind-spot network for image anomaly detection: A new approach to diabetic retinopathy screening.

A. Takyar, 2023. From diagnosis to treatment: exploring the applications of generative AI in healthcare, LeewayHertz - AI Development Company, (2023). [Online]. Available: https://www.leewayhertz.com/generative-ai-in-healthcare/.

Wang, J., Li, W., Chen, Y., Fang, W., Kong, W., He, Y., Shi, G., 2021. Weakly supervised anomaly segmentation in retinal OCT images using an adversarial learning approach. Biomed. Opt. Express 12 (8), 4713–4729.
Waniewski, J., Hajeb Mohammad Alipour, S., Rabbani, H., Akhlaghi, M.R., 2012. Diabetic retinopathy grading by digital curvelet transform. Comput. Math. Methods Med.
WHO, 2021. SDG Target 3.8 — Achieve universal health coverage (UHC) [Online]. Available: WHO https://www.who.int/data/gho/data/major-themes/universal-health-coverage-major.
WHO, "Diabetes," WHO, (2023).
Wilson, A., Izmailov, P., 2020. Bayesian deep learning and a probabilistic perspective of generalization. NIPS'20 Proceedings of the 34th International Conference on Neural Information Processing Systems.
Yan, Y., Tan, M., Xu, Y., Cao, J., Ng, M., Min, H., Wu, Q., 2019. Oversampling for imbalanced data via optimal transport. In: Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, pp. 5605–5612.
Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., Zhao, Y., Zhang, W., Cui, B., Yang, M.-H., 2023. Diffusion models: A comprehensive survey of methods and applications. ACM Comput. 56 (4), 1–39.
Zang, P., Hormel, T.T., Wang, J., Guo, Y., Bailey, S.T., Flaxel, C.J., Huang, D., Hwang, T. S., Jia, Y., 2022. Interpretable diabetic retinopathy diagnosis based on biomarker activation map. Electr. Eng. Syst. Sci.
C. Zhang, Y. Wang, X. Zhao, Y. Guo, G. Xie, C. Lv, B. Lv, 2020. Memory-Augmented anomaly generative adversarial network for retinal OCT images screening. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI).
Zhao, H., Li, Y., He, N., Ma, K., Fang, L., Li, H., Zheng, Y., 2021. Anomaly detection for medical images using self-supervised and translation-consistent features. IEEE Trans. Med. Imaging 40 (12), 3641–3651.
Zhou, K., Gao, S., Cheng, J., Gu, Z., Fu, H., Tu, Z., Yang, J., Zhao, Y., Liu, J. 2019. Sparse-GAN: Sparsity-constrained Generative Adversarial Network for Anomaly Detection in Retinal OCT Image, CoRR, vol. abs/1911.12527.
K. Zhou, Y. Xiao. J. Yang, J. Cheng, W. Liu, W. Luo, Z. Gu, J. Liu and S. Gao, 2020. Encoding structure-texture relation with P-Net for anomaly detection in retinal images. In: European Conference on Computer Vision.
Zhou, X., Niu, S., Li, X., Zhao, H., Gao, X., Liu, T., Dong, J., 2023. Spatial–contextual variational autoencoder with attention correction for anomaly detection in retinal OCT images. Comput. Biol. Med. 152 (0010-4825), 106328.
Zhou, Y., Wang, B., Huang, L., Cui, S., Shao, L., 2021. A benchmark for studying diabetic retinopathy: segmentation, grading. IEEE Trans. Med. Imaging 40 (3), 818–828.
J.-Y. Zhu, T. Park, P. Isola, A. A. Effros, 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks.