Section A-Research paper

A Survey on Detection of Ophthalmic Diseases using Applied Deep Learning Techniques

Akalbir Singh Chadha[1], Aryan Kenchappagol[2], Rutuja Jangle[3], Yashowardhan Shinde[4] and Dr Ajitkumar Shitole[5]

Department of Computer Engineering, International Institute of Information Technology Pune, India 411057[1],
Department of Computer Engineering, International Institute of Information Technology Pune, India 411057[2],
Department of Computer Engineering, International Institute of Information Technology Pune, India 411057[3],
Department of Computer Engineering, International Institute of Information Technology Pune, India 411057[4],
Associate Professor,Department of Computer Engineering, International Institute of Information Technology Pune, India 411057[5]

Abstract

There are 2.2 billion visually impaired people in the world, and 1 billion of them have avoidable vision impairment, according to the World Health Organization's World Report on Vision 2019. Inequalities in eye care coverage, treatment, and rehabilitation are a global problem. Early diagnosis of ocular disorders can prevent vision loss. Retinal fundus scans can be very useful while diagnosing ophthalmic diseases. Using these retinal scans, it is possible to detect over 46 different diseases. This use case of retinal scans can be of great benefit to doctors. Using just one simple retinal scan doctors can diagnose multiple diseases at once. This research aims to review and summarise different deep-learning approaches that can be used to solve the problem of diagnosing diseases using retinal scans. The focus is on reviewing deep learning techniques that can be used for multi-disease detection and nerve segmentation.

Keywords

Retinal scans, Disease detection, nerve segmentation, autoencoders, deep learning.

1. Introduction

Many people in the modern world have trouble with their eyesight for a variety of reasons. A number of them have trouble determining the precise nature of the ophthalmic disorder they have experienced. Some recent studies show that Macular degeneration is more common in Whites than in African Americans, but Glaucoma strikes African Americans more frequently as they age [1]. Moreover, the retinal images are diagnosed for the presence of drusen in order to perform risk assessment of a fundus for detection of at an early-stage Age-related Macular Degeneration [3]. In addition, Ophthalmologists often disregard sight-threatening uncommon disorders such as central retinal artery blockage or anterior ischemic optic neuropathy. To aid these people, it is proposed to implement a system that will enable doctors to distinguish between various ophthalmic diseases by analysing scanned images of the patient's retinal fundus. Multiple public colour fundus image datasets have been collected over the past two decades, with a particular emphasis on diabetic retinopathy, glaucoma, and age-related macular degeneration. This challenge aims to bring together the medical image analysis community to create approaches for automatic eye illness classification. Moreover, the challenge aims to construct generalizable models for screening the retina, unlike past initiatives that focused on specific disorders. Due to the recent development of artificial intelligence, systems that can perform complex tasks like image classification and segmentation have become an area of interest for researchers all over the world.

2. Literature Survey

The literature survey is divided into 3 sub-parts. The first part of this section focuses on different ophthalmic conditions that can be diagnosed using retinal fundus scans. The second part discusses machine-learning and deep-learning methods that are employed to solve image classification tasks and the third part focuses on different machine-learning methods that can be used to solve bio-medical tasks such as segmentation and multi-disease detection.

2.1 Ophthalmic Disease-Based Literature Survey

Gohdes DM et.al. explain that age-related eye diseases cause reduced vision and, in severe cases, blindness. This study compares the frequency of age-related eye diseases among Whites, African Americans, and Hispanics aged 40 and above in the U.S. This paper summarises the risk factors and prevalence of age-related eye diseases, including diabetic retinopathy, glaucoma, macular degeneration, and cataract, across demographics. Eye therapy can be expensive and isn't always effective. The report discusses age-related eye problems' public health effects. The author also emphasised preventative measures, early disease detection, and public awareness [1]. Age-related macular degeneration causes blindness in the over-50 population (ARMD). ARMD, a posterior eye disorder, impairs central vision and can cause blindness. Dry ARMD affects 33% of 65-year-olds. Another kind is wet ARMD. Shahrokh Ramin et al. use scientometrics to analyse macular degeneration research's distinctive traits and trends. This study outlined future research priorities and knowledge gaps. This study identified information gaps to help design large-scale research. This paper's citation analysis considers local and worldwide ratings (GCS). The number of macular degeneration research articles grew by 14.46% between 1993 and 2013 [2]. Bhuiyan Alauddin et.al defines Drusen as a condition in which cellular debris accumulates behind the retina. Drusen can lead to AMD (ARMD). Age-related macular degeneration causes blindness in people over 50. This study intends to diagnose Age-Related Macular Degeneration early by automatically recognising, categorising,

and quantifying drusen in colour retinal images. Canon D60 retinal images were used to find drusen. Fifty photographs were initially analysed, but only 12 were used to demonstrate Drusen's accuracy. The system's sensitivity and specificity were determined by comparing generated and ground truth images. The model detected drusen with 100% accuracy, 74.94% sensitivity, and 81.17% specificity. Intermediate drusen subtype accuracy was 79.59% and soft was 82.14%. Remote areas can also examine fundus photos and send patients to ophthalmologists [3]. Rehak J et.al explains that Branch retinal vein blockage causes impaired peripheral vision (BRVO). BRVO is categorised into major and macular based on the obstructed retinal veins. As expected, people over 65 are most affected. Compression at the A/V crossing, degenerative vascular wall alterations, and aberrant haematological factors induce BRVO. According to visual prognosis, 50% to 2/3 of BRVO patients improve without treatment. Different patients' VA data were compared using the chi-square test with Yates adjustment. Eyes with a VA of 20/50 or greater have a good visual prognosis without therapy. Small sample sizes, limited follow-up periods, and difficulty to identify clinical entities are methodological flaws of BRVO therapies. Randomised clinical trials were once used to treat BRVO. Macular grid photocoagulation is a successful treatment for ME in patients with BRVO and 20/40 or worse VA. [4]. Dickson D et.al. suggest that Vitreous haze (VH) arises when vitreous and protein cells collect on the fundus. Uveitis is common. Uveitis, an inflammatory eye condition, is a major global contributor to reduced vision. According to some estimates, uveitis causes 25% of blindness worldwide. To examine Vitreous Haze, this paper outlines the use of Ultra-Wide Field (UWF) retinal imaging, a scanning laser technique. Carl Zeiss FF450 camera took the colour fundus photos. The same person who took the fundus photos also used an Optos UWF P200Tx retinal camera. After comparing 92 separate eye scans, researchers concluded that regular Zeiss images couldn't hold a candle to the quality of the composite images made by the UWF scanner. UWF photographs were 0.27 as sensitive as Zeiss images, and 0.88 as specific [5]. Yanping Zhou et.al. examined choroidal and retinal thicknesses in myopic tessellated eyes. A total of 115 eyes with high myopia were recruited for this study and characterised as having a tessellated fundus or a normal fundus. Quantification of RT and CT was accomplished by utilising optical coherence tomography with improved depth imaging (EDI-OCT) (EDI-OCT). Tessellation and subfoveal CT were compared using logistic regression models (SFCT). Statistically, tessellated fundus eyes had a thicker CT but a thinner RT. In every macular region, tessellated eyes had thinner choroids than controls. After adjusting for age, gender, and exercise level, the SFCT highly predicted tessellation. Binary logistic regression produced an OR of 0.975 (0.960-0.990) and a 95% CI of 0.991-0.995. (0.984–0.999, Cox regression). AUC = 0.824 for SFCT tessellation detection. SFCT has the highest sensitivity and specificity (81.8% and 74.2%). When compared to non-tessellated highly myopic eyes, the CT of tesselated highly myopic eyes is thinner. CT can predict high myopia in 6-year-olds [6]. Myopic eyes commonly have a tessellated fundus, which indicates retinochoroidal changes. No one knows the aetiology or significance of tessellated fundus. Yoshihara N et.al objectively assessed the degree of tessellation in the fundi of non-pathologically myopic eyes and discovered a positive association between tessellation and choroidal thickness (CT) and axial length (AL) (AL). This study utilised a cross-sectional design. They employed ophthalmoscopic tessellation to split eyes into three categories. ImageJ was used to evaluate digital colour fundus tessellation. It was decided to construct three tessellated fundus indices (TFIs) and compare them to the three categories chosen subjectively. OCT scans quantified subfoveal and nasal CTs. CT was used to determine the importance of all TFIs. The TFIs' AL association was also computed. One hundred normal people (mean age 25.8 3.9) had their right eye checked. 57, 27, and 16 eyes were assigned to the non-tessellated, slightly tessellated, and severely tessellated groups, respectively. The subjective categorizations correlated with TFI values (P0.05, Kruskal-Wallis test). Subfoveal and nasal CT correlate well with all TFIs (R = 0.20 to 0.24, P 0.05). No notable association was identified between TFIs and ALs. TFIs can be used to categorise tessellation degrees because there is a good correlation between subjective and objective classifications. Results imply that CT differences explain tessellation [7].

2.2 Image Classification Architecture-Based Literature Survey

R. Ribani et.al. talks about how transfer learning may influence the future of machine learning in academics and industry. It's used since gathering and categorising data is expensive and time-consuming, and recent privacy issues make it impractical to use real user data. Transfer learning allows quick prototyping of new machine learning models by utilising pre-trained models from a source task. In this overview, they discuss the many terms used to define transfer learning. To provide a clear vision for future studies on this topic, they have included the perspectives of numerous authors of past surveys [8]. Kolesnikov A et.al. explain how transferring pre-trained representations can boost sampling efficiency and reduce hyperparameter tweaking time. The authors look at pre-training on large supervised datasets and then tweaking the model for a given job. Big Transfer scales up pre-training (BiT). Combining a few well-chosen components and employing a simple approach yields great performance on over 20 datasets. BiT works nicely with 1 sample per class or 1,000,000 examples. BiT's top-1 accuracy on ILSVRC-2012, CIFAR-10, and the 19-task Visual Task Adaptation Benchmark is 87.5%. (VTAB). BiT obtains 76.8% accuracy on ILSVRC-2012 with 10 examples per class and 97% accuracy on CIFAR-10 with 10 examples per class. They have analysed transfer performance

factors [9]. Attention-based neural networks like ViT have beaten all previous methods on computer vision benchmarks. Zhai, X et.al. explain that knowing how a model scales is critical for future versions. Scaling Transformer language models have been researched, but Vision Transformers have not. ViT models and data are vertically and horizontally scaled, and error rate, data, and computation links are characterised. They improve ViT's design and training methods to reduce its memory dependence and improve its output models. With this strategy, the two-billion-parameter ViT model achieves 90.45% top-1 accuracy on ImageNet. It obtains 84.86% top-1 accuracy on ImageNet with 10 instances per class [10]. Q. Xie et. al present a simple self-training method that, with just 3.3 billion poorly labelled Instagram images, outperforms the most advanced model on ImageNet by 2.0%. On robustness test sets, it improves ImageNet-aesthetics. C's Top-1 accuracy increased from 61.0% to 83.7%, and mean corruption errors in ImageNet-C and ImageNet-P decreased from 45.7% to 28.3% and 27.8% to 12.2%, respectively. To get there, they use an EfficientNet model that was trained on labelled ImageNet images to create pseudo labels on 300M unlabeled images. They then trained a more durable EfficientNet as a student model using the combined labelled and falsely labelled images. The student assumes the position of the teacher during this repetition.[11]. Deeper neural network training is harder. K. Zang et. al. used residual learning to train deep networks. Instead of learning unique functions, they have reformulated the layers such that they learn residual functions. Empirical evidence shows that residual networks are easier to optimise and can benefit from more depth. Residual nets with 152 layers are 8x more complex than VGG networks yet still less complex. Residual nets can reduce the ImageNet benchmark error rate to 3.57 percent when employed in an ensemble. This won the 2015 ILSVRC classification. 100- and 1000-layer CIFAR-10 analyses are shown. Many visual identification tasks require representational depth. The deep representations increase COCO object detection by 28%. The ILSVRC and COCO 2015 submissions1 utilised deep residual nets and won on ImageNet detection, ImageNet localization, COCO detection, and COCO segmentation. [12]. Transformers are gaining popularity in computer vision, although Dai Z et. al. still trail behind convolutional networks. This paper shows that while Transformers offer larger model capacity, their generalisation may be inferior to convolutional networks due to a lack of inductive bias. CoAtNets (pronounced "coat" nets) combine the benefits of two architectures: 1. Vertically stacking convolution and attention layers improve generalisation, capacity, and efficiency and 2. Simple relative attention can naturally combine convolution and self-attention. Experiments show that CoAtNets can outperform state-of-the-art models despite varying degrees of hardware and data availability [13]. MobileNetV2 beats earlier work across a wide range of model sizes, tasks, and benchmarks. Sandler M. et al. also describe SSDLite, a novel framework for object detection using mobile models. They also show how to use DeepLabv3 Mobile to build mobile semantic segmentation models. DeepLabv3's inverted residual structure uses shortcut connections between narrow bottleneck layers. In the intermediate expansion layer, lightweight depthwise convolution filter, and non-linearize features. To sustain representational capacity, they reduced fine-grained non-linearities. This improves performance and design. Their technique separates input and output domains from formation expressiveness, giving it an easy analytical starting point. This study benchmarks ImageNet classification, COCO object identification, and VOC picture segmentation. Multiply-adds (MAdds), number of operations, real delay, and number of parameters are weighed [14]. Convolutional Neural Networks (ConvNets) are often created on a minimal budget and scaled up when funds become available. In this research, Tan M. et. al. examine model scaling and find that balancing network depth, width, and resolution enhances performance. Using a simple but effective compound coefficient, they create a new scaling approach that applies the same scale to an image's depth, width, and resolution. This strategy can increase MobileNets and ResNet's size. Using neural architecture search, they design a new baseline network and scale it up to produce EfficientNets, which are more accurate and efficient than ConvNets. EfficientNet-B7 is 8.4 times smaller and 6.1 times quicker than the best available ConvNet while attaining 84.3% top-1 accuracy on ImageNet. The EfficientNets translate well and achieve state-of-the-art accuracy on CIFAR-100 (91.7%), Flowers (98.8%), and 3 other datasets. [15]. Xie S et. al. offer a simple, modularized network for image classification. The network is constructed by repeatedly connecting same-topology transformations. This simple approach produces a multi-branch architecture with few hyper-parameters. This approach exposes "cardinality" in addition to height and width (the size of the set of transformations). On ImageNet-1K, they show that increasing cardinality improves classification accuracy, even while keeping complexity. Cardinality expansion is more efficient than digging deeper or adding width. ResNeXt models helped us place second in the ILSVRC 2016 classification challenge. ResNeXt outperforms ResNet on ImageNet-5K and COCO. Source code and models are free online1 [16]. Chollet F. et. al. propose a paradigm for Inception modules in convolutional neural networks to bridge ordinary convolution with depth-wise separable convolution. A depth-wise separable convolution is like an Inception module with infinite towers. Researchers present a new Inception-like architecture for deep convolutional neural networks using depth-wise separable convolutions. Xception performs marginally better than Inception V3 on the ImageNet dataset (for which Inception V3 was designed) and much better on a larger dataset with 350 million images and 17,000 classes. Performance gains aren't due to Xception's enhanced capacity but to more efficient model parameter use. [17]. EfficientNetV2 is a new class of convolutional networks that improve on existing models in training time and parameter

efficiency. Tan M. et. al. develop these models using training-aware neural architecture search and scaling to optimise training speed and parameter efficiency. New operators like Fused-MBConv were introduced to the search space. This research shows that EfficientNetV2 models can be trained faster and are 6.8x smaller than state-of-the-art models. Increasing image size during training can save training time but reduces accuracy. They provide an improved technique of progressive learning that dynamically alters regularisation (e.g., data augmentation) as image size expands. EfficientNetV2, which uses deep learning, improves on ImageNet and CIFAR-Cars/Flowers [18]. Very deep convolutional networks have improved image recognition in recent years. The Inception design achieves high performance with low computation. Recent state-of-the-art performance in the 2015 ILSVRC challenge was achieved by combining residual connections with a more conventional architecture; its performance was comparable to the latest generation Inception-v3 network. This raises the question of whether Inception with residual connections is effective. Research shows that residual connections dramatically speed up Inception network training. Although more expensive, Inception networks with residual connections outperform those without. The authors also introduce simplified residual and non-residual topologies for Inception networks. Single-frame ILSVRC 2012 classification benefits from these tweaks. Szegedy C. et. al. also show how activation scaling can stabilise large residual Inception network training. They get a top-5 error of 3.08% on the ImageNet CLS test set using three residuals and one Inception-v4 [19].

2.3 Disease Detection-Based Literature Survey

According to P. Tang et al. accurate skin lesion categorization is challenging due to inter-class similarity and intra-class variation and the weak generalisation ability of single deep Convolutional Neural Networks trained with little data. They introduce a Global-Part Convolutional Neural network (GP-CNN) model that considers both local and global contexts. The Global-Part model combines G-CNN with P-CNN. A G-CNN is trained using downscaled dermoscopy pictures to extract global-scale information and build the Classification Activation Map (CAM). P-CNN is trained with CAM-guided cropped photo patches to acquire micro-level skin lesion data. They also offer a data-transformed ensemble learning technique that can improve classification performance by merging discriminant information from GP-CNNs trained with original, colour constancy, and feature saliency altered images. Their experiments have revealed that the suggested method may achieve SOTA skin lesion classification performance without using external data [20]. Imran Ul Haq et al. have developed CAD, a system that helps radiologists detect disease phases and normal/abnormal tissues. The classification problem in CAD systems is notoriously challenging because of the poor contrast and noise of mammography images, the shape and location variations of tumours, and the remarkable similarities between normal and malignant regions of interest (ROI). They provide a new deep convolutional neural network (DCNN) technique based on feature fusion and ensemble learning to improve mammogram anomaly identification and classification. According to their research, ensemble learning improves normal and tumour ROI classification for more believable findings, while feature fusion helps detect discriminative characteristics. In mammography classification, spatial dropout and depth-wise separable convolution are considered solutions to overfitting and limited datasets. The suggested model is evaluated using MIAS and BCDR. On MIAS, the model exhibited high sensitivity, specificity, and accuracy [21]. In the article [22]. C. Jia et al. use ensemble learning in convolutional neural networks to recognize emotions. Their model integrates three independent neural network predictions using the SVM classifier. The model's expression recognition accuracy on FER2013 was 71.27%. Due to its high test accuracy and low prediction time, the technique provides high-performance, real-time facial recognition [22]. P. Monkam et al. in this article use multiple-view 3-D convolutional neural networks to classify non-nodules and micro-nodules in CT images (3D-CNNs) for lung cancer detection. LIDC/1010 IDRI's CT scans yield 34494 volumetric image samples, including 13179 micro-nodules and 21315 non-nodules. 20203, 16163, 12123, 883, and 443 are crop sizes used to investigate pulmonary nodule prospects. Five 3D-CNN models are then applied to a single nodule size. Combining five 3D-CNNs with an ELM network yields classification results. The proposed system's accuracy, AUC, F-score, and sensitivity are evaluated. The suggested system has a 97.35% accuracy rate, 0.98% AUC, 96.42% F-score, and 96.57% sensitivity. 2D-CNNs, single 3D-CNN models and state-of-the-art techniques trained on the same dataset produce worse outcomes. ELM outperforms autoencoder, averaging, majority voting, and AND. Extending lung cancer studies to include micro-nodules shows that an intuitive approach for distinguishing micro-nodules from non-nodules in CT images is possible. According to their research combining multiple-view 3D-CNNs with ensemble learning improves identification performance [23]. Das A et al. explain that COVID-19 illness is a major 20th-century health problem. In developing nations like Bangladesh, India, and others, COVID-19 is still under-recognized. Medical image analysis employing deep learning classifiers predicts COVID-19 cases better than RT-PCR. Recent lung illness investigations using COVID-19 chest X-rays have shown promise. Automatic COVID-19 diagnosis uses pharmaceutical records and medical images to emphasise coronavirus testing methods and control community spread. Convolutional Neural Network (CNN) data patterns can indicate tiny differences between healthy and infected X-rays, despite problems in manual interpretation due to COVID-19 infection (CNN). In this paper, the authors use CNN-based deep features to improve CNN's detection performance (EL-CNN-DF). Lung segmentation is COVID's second and most important step. U-Net, based on AAF, does this (AAF-U-Net).

After segmenting the lungs, EL-CNN-DF is used. This strategy replaces the fully connected CNN layer with SVM, AE, and Naive Bayes classifiers (NB). These classifiers employ a high-ranking approach to recognize COVID-19. SA-TSA is used to boost segmentation and detection performance. The increased CNN's precision utilising SA-TSA was 1.02 percentage points, 4.63 percentage points, 3.38 percentage points, 1.62 percentage points, 1.51 percentage points, and 1.04% percentage points greater than that of SVM, autoencoder, NB, Ensemble, RNN, and LSTM, respectively [24]. Anjali R. Shah et al. wrote "Diabetic Retinopathy: Research to Clinical Practice." In the US, diabetic retinopathy (DR) is the leading cause of vision loss. The CDC thinks DR will double by 2050. The document summarises prior diabetes research and its clinical impact. Examine future directions and present paradigms. DR understanding and treatment have improved, but the condition's rising incidence requires novel management measures. [25]. Ling Dai et al. have created DeepDR, a deep learning algorithm, that may identify diabetic retinopathy in its early to late stages, simplifying screening. DeepDR can perform real-time lesion detection, image quality assessment, and grading using fundus pictures of diabetic patients. For microaneurysms, cotton wool patches, hard exudates, and haemorrhages, the AUC is 0.901, 0.941, 0.954, and 0.967. The area under the curves for diabetic retinopathy classifications are 0.943, 0.955, 0.960, and 0.972. The range of AUC for diabetic retinopathy grading is 0.916 to 0.970, this demonstrates that their approach is effective [26]. D. Jude Hemanth et al. present a hybrid strategy to employ retinal fundus images to detect diabetic retinopathy. Combining deep learning with image processing improves the hybrid technique. Doctors need imaging and computer vision systems, and intelligent diagnosis tools are the next step. This study's proposed approach uses histogram equalisation and contrast-limited adaptive histogram equalisation. Diagnosis is performed using convolutional neural network classification. Their approach was evaluated using 400 retinal fundus images from the MESSIDOR database and had 97% accuracy, 94% sensitivity (recall), 98% specificity, 94% precision, 94% F-score, and 95% GMean. A comparison with prior studies shows that the new method is effective at diagnosing diabetic retinopathy from retinal fundus pictures [27]. Alexander K. Schuster et al. examine glaucoma's epidemiology, risk factors, diagnostics, and treatments. According to this review, Europe's 40-to-80-year-olds have 2.93 percent glaucoma. Over the 90s have a 10% incidence rate. Diagnostic procedures include perimetry, tonometry, ophthalmoscopy, and imaging. Glaucoma treatment includes topical medicines, laser treatments, and surgery. Lowering intraocular pressure in glaucoma patients reduces visual field aberrations by 7. According to their research diagnosing glaucoma requires evaluating ocular function, intraocular pressure and morphology. Regularly examine individually-designed blood pressure-lowering medicine [28]. Huazhu Fu et.al. introduce two deep learning-based glaucoma detection algorithms. M-Net solves optic disc and optic cup segmentation simultaneously. M-Net is a multi-scale U-shaped convolutional network with a side-output layer that provides a segmentation probability map. Using a segmented optic disc and cup, the vertical cup-to-disc ratio (CDR) assesses glaucoma risk. DENet is an ensemble network that is disc aware, meaning that it takes into account both the global context of a fundus picture and the local optic disc. Without the need for image segmentation, DENet can detect glaucoma. They compare two deep learning algorithms using similar approaches on several glaucoma datasets [29]. Ajitha, S., et al. explain how glaucoma causes irreversible vision loss and blindness by increasing intraocular pressure. Preventing eyesight loss requires a quick, correct diagnosis. Manual glaucoma identification needs skill and practice. The author offers a powerful and accurate glaucoma detection method utilising a convolutional neural network (CNN). In this study, 1113 fundus photographs from four databases were used to diagnose glaucoma. This dataset trains a 13-layer CNN to identify glaucomatous or normal characteristics. 12012 fundus photographs accompany the training photos. Their model using the SoftMax classifier had 93.86% accuracy, 85.42% sensitivity, 100% specificity, and 100% precision. The SVM classifier model achieved 95.61, 89.58, 100, and 100% accuracy, sensitivity, and specificity, respectively. These findings reveal that deep learning can identify glaucoma from fundus images, meaning that the suggested method can help ophthalmologists make quick, accurate, and trustworthy diagnoses [30].

3. Datasets

This section discusses different data sources that are publicly available for solving tasks like disease detection and nerve segmentation using retinal scans.

Table 1 Different open-source datasets available for retinal fundus images

Dataset	Description	Use
Retinal Fundus Multi-Disease Image Dataset (RFMiD) [31]	Original colour fundus images. Total of 3200 PNG image files	Multi-Disease Detection
High-Resolution Fundus (HRF) Image Database [32]	15 images of healthy patients, patients with diabetic retinopathy, and glaucomatous patients each.	Nerve Segmentation
DRIVE: Digital Retinal Images for Vessel Extraction [33]	Total of 40 JPEG colour fundus images; including 7 abnormal pathology cases.	Nerve Segmentation
PAPILA: Dataset with fundus images and clinical data of both eyes of the same patient for	244 patient records with information such as optic cup and disc segmentations, clinical data.	Glaucoma Detection

glaucoma assessment [34]		
FIVES: A Fundus Image Dataset for Artificial Intelligence based Vessel Segmentation [35]	800 high-resolution multi-disease colour fundus photographs.	Nerve Segmentation
Retinal fundus images for glaucoma analysis: the RIGA dataset [36]	Includes three datasets [1] MESSIDOR: 3220 images [2] Bin Rush Ophthalmic: 1365 images [3] Magrabi Eye Center: 665 Images	Glaucoma Detection, Nerve Segmentation
ORIGA-light : An Online Retinal Fundus Image Database for Glaucoma Analysis and Research [37]	650 retinal images annotated by trained professionals from Singapore Eye Research Institute.	Glaucoma Detection
Indian Diabetic Retinopathy Image Dataset (IDRiD) [38]	[A] Segmentation: 81 Images [B] Disease Grading: 516 Images [C] Localization: 516 Images	Diabetic Retinopathy Detection
The University of Auckland Diabetic Retinopathy (UoA-DR) [39]	200 Images for Diabetic Retinopathy. Resolution: 2124 x 2056 pixels Size: approx. 500MB	Diabetic Retinopathy Detection

4. Algorithms

This section discusses the algorithms that can be used for disease detection and nerve segmentation tasks. Networks like Convolutional Neural Networks are used for classification tasks as high computing power is required to train these models, and methods like transfer learning are applied. Moreover, to achieve higher accuracy and more reliable results, methods like ensemble learning are also applied. For nerve segmentation, generative networks like Autoencoders and Generative Adversarial networks are used. The algorithm for disease classification using ensemble learning is shown in Algorithm 1.

Algorithm 1: Disease Detection

Input: Retinal Fundus Scan
Output: Disease Prediction

1. *img ← read_image()*
   # Normalize, Crop, Resize Image
2. *img ← preprocess_scan(img)*

   # Set of trained classifiers
3. *ensemble_models ← [clf_1, clf_2, clf_3 …clf_n]*
4. *predictions ← []*
5. *for classifier in ensemble_models:*
6. *pred ← classifier.prediction(img)*
7. *prediction.add(pred)*

   # Majority Voting / Weighted Average
8. *final_prediction ← aggregate_predictions(predictions)*
9. *end*

The algorithm takes a retinal fundus scan image and predicts the type of disease using an ensemble approach. The steps for this algorithm can be summarised as follows:

- Read the input image and pre-process it so it can be fetched as input to the network.
- Initialise a list of all the classifiers chosen for ensemble learning.
- Using each classifier, predict the type of disease of the given retinal scan.
- After all such predictions are obtained, aggregate them to obtain a final prediction using majority voting or weighted average.
- Lastly, output the type of disease as the final prediction for the input retinal scan.

4.1 Convolutional Neural Network (CNN)

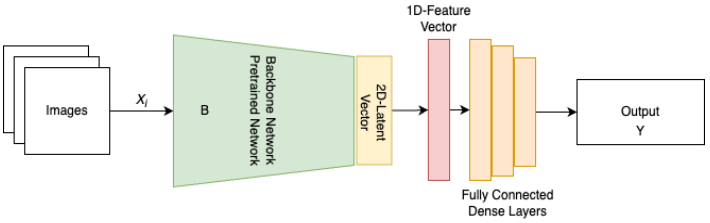

Figure 1 The architecture of CNNs

Consider an input image $\boldsymbol{x}_i$, this image is passed onto different CNNs with various Backbone Networks let these networks be, **B(x).** The output of these backbone networks is a 2D-feature vector, this 2D vector is then converted to a 1D vector and passed on to the fully connected dense layers. Let the fully connected layers be represented by $\boldsymbol{f}(\mathbf{x})$, and let the output be **Y**.
This can be represented as

$$Y = f\big(B(x_i)\big) \dots (1)$$

4.2 Ensemble Learning

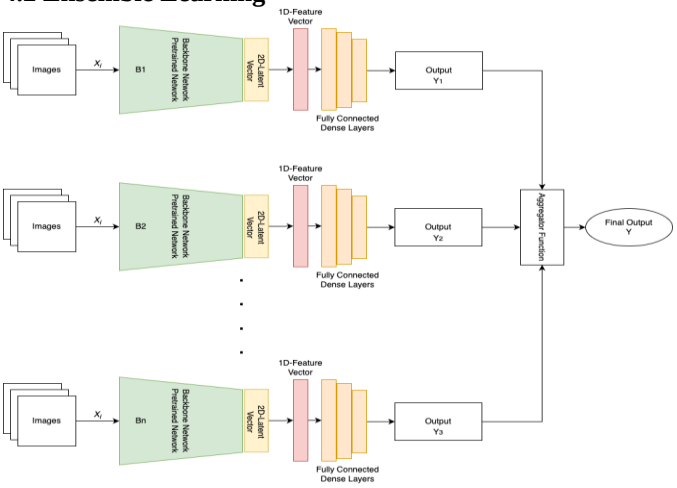

Figure 2 Ensemble Learning

Consider an input image x_i, this image is passed onto different CNNs with various Backbone Networks let these networks be, $B_1(x)$, $B_2(x)$, …..$B_n(x)$. The output of these backbone networks is a 2D-feature vector, this 2D vector is then converted to a 1D vector and passed on to the fully connected dense layers. Let the fully connected layers be represented by $f(x)$, and let the output be y_i. These outputs are passed on to an aggregator function $G(x)$ which gives the final output Y.
This can be represented as

$$Y = G\left(f\left(B_1(x_i)\right), f\left(B_2(x_i)\right), \dots, f\left(B_3(x_i)\right)\right) \dots (2)$$

4.3 Autoencoders

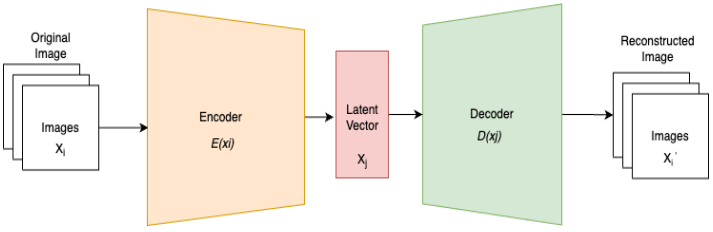

Figure 3 The architecture of Autoencoders

Consider an input image x_i, this image is passed to the encoder E(x) whose function is to extract features from the image. The output of this Encoder is a latent vector x_j which is then used for image reconstruction. This latent vector is passed to the decoder D(x) which reconstructs the latent vector to its original dimensions.
This can be represented as,

$$E: X_i \rightarrow X_j \, D: X_j \rightarrow X_i \, E, D = MSE \dots (3)$$

Mean Squared Error (MSE)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - Y_i^{'})^2 \dots (4)$$

4.4 Generative Adversarial Networks (GANs)

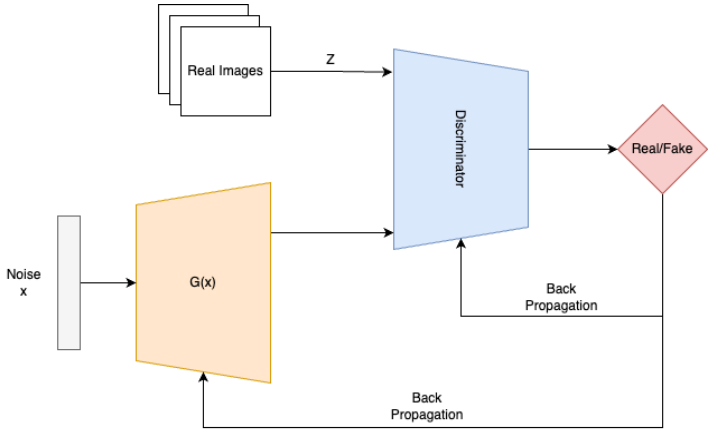

Figure 4 The architecture of GANs

Consider a random noise z, this noise is given as an input to the generator network G. The output of the Generator network G(z) and real images (x) are fed to the discriminator model which classifies the samples as real/fake and back propagates the loss. The min-max loss used in GANs can be represented as follows.

$$L = min\ max\ G\ D\ [\mathbb{E}_{x\sim Pdata}\ log\left(D(x)\right) + \mathbb{E}_{z\sim P(z)} log\left(1 - D\left(G(z)\right)\right)] \dots (5)$$

5. Conclusion and Future Work

In this research paper, numerous ophthalmic diseases and methods for diagnosing them have been studied. The survey takes into consideration two main tasks related to retinal fundus scans, image classification, and segmentation, and the data available for these tasks. Most of these ophthalmic diseases have also been the subject of research into image classification using deep neural networks like convolutional neural networks paired up with methods like ensemble learning and transfer learning and nerve segmentation using generative models like autoencoders and GANs. The paper also presents several data sources available for tasks like disease detection and nerve segmentation.

In the future, the goal is to develop a deep-learning framework around the diagnosis of diseases using retinal fundus scans. This framework can act as an assistive tool for doctors to make diagnoses. The framework can include a wide range of functionalities like risk prediction, disease detection, and nerve

segmentation. Furthermore, the framework can have features like patient management and report generation for extending its real-world use. An efficientcloud-based tool can be designed to overcome the high compute requirement of the machine learning models [40, 41] this will also make the application more accessible and reliable. The system can also be extended using IoT and Cloud to create more robust model to classify various ophthalmic diseases accurately [42].

6. References

[1] Gohdes DM, Balamurugan A, Larsen BA, Maylahn C. Age-related eye diseases: an emerging challenge for public health professionals. Prev Chronic Dis. 2005 Jul;2(3):A17. Epub 2005 Jun 15. PMID: 15963319; PMCID: PMC1364526.

[2] Ramin S, Soheilian M, Habibi G, Ghazavi R, Gharebaghi R, Heidary F. Age-Related Macular Degeneration: A Scientometric Analysis. Med Hypothesis DiscovInnovOphthalmol. 2015 Summer;4(2):39-49. PMID: 26060829; PMCID: PMC4458325.

[3] Bhuiyan, Alauddin& Kawasaki, Ryo & Sasaki, Mariko & Lamoureux, Ecosse &Ramamohanarao, Kotagiri&Guymer, Robyn & Wong, T-Y &Kanagasingam, Yogesan. (2013). Drusen Detection and Quantification for Early Identification of Age Related Macular Degeneration using Color Fundus Imaging. Clinical and Experimental Ophthalmology. 4. 10.4172/2155-9570.1000305.

[4] Rehak J, Rehak M. Branch retinal vein occlusion: pathogenesis, visual prognosis, and treatment modalities. Curr Eye Res. 2008 Feb;33(2):111-31. doi: 10.1080/02713680701851902. PMID: 18293182; PMCID: PMC2430176.

[5] Dickson D, Agarwal A, Sadiq MA, Hassan M, High R, Nguyen QD, Sepah YJ. Assessment of vitreous haze using ultra-wide field retinal imaging. J Ophthalmic Inflamm Infect. 2016 Dec;6(1):35. doi: 10.1186/s12348-016-0105-0. Epub 2016 Sep 29. PMID: 27687961; PMCID: PMC5042918.

[6] Yanping Zhou, Minlu Song, Minwen Zhou, Yiming Liu, Fenghua Wang, Xiaodong Sun, "Choroidal and Retinal Thickness of Highly Myopic Eyes with Early Stage of Myopic Chorioretinopathy: Tessellation", *Journal of Ophthalmology*, vol. 2018, Article ID 2181602, 9 pages, 2018. https://doi.org/10.1155/2018/2181602

[7] Yoshihara N, Yamashita T, Ohno-Matsui K, Sakamoto T. Objective analyses of tessellated fundi and significant correlation between degree of tessellation and choroidal thickness in healthy eyes. PLoS One. 2014 Jul 28;9(7):e103586. doi: 10.1371/journal.pone.0103586. PMID: 25068821; PMCID: PMC4113439.

[8] R. Ribani and M. Marengoni, "A Survey of Transfer Learning for Convolutional Neural Networks," 2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T), 2019, pp. 47-57, doi: 10.1109/SIBGRAPI-T.2019.00010.

[9] Kolesnikov, A. et al. (2020). Big Transfer (BiT): General Visual Representation Learning. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, JM. (eds) Computer Vision – ECCV 2020. ECCV 2020. Lecture Notes in Computer Science(), vol 12350. Springer, Cham. https://doi.org/10.1007/978-3-030-58558-7_29

[10] Zhai, X., Kolesnikov, A., Houlsby, N. and Beyer, L., 2022. Scaling vision transformers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 12104-12113).

[11] Q. Xie, M. -T. Luong, E. Hovy and Q. V. Le, "Self-Training With Noisy Student Improves ImageNet Classification," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 10684-10695, doi: 10.1109/CVPR42600.2020.01070.

[12] He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[13] Dai, Z., Liu, H., Le, Q.V. and Tan, M., 2021. Coatnet: Marrying convolution and attention for all data sizes. Advances in Neural Information Processing Systems, 34, pp.3965-3977.

[14] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L.C., 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520).

[15] Tan, M. and Le, Q., 2019, May. Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning (pp. 6105-6114). PMLR.

[16] Xie, S., Girshick, R., Dollár, P., Tu, Z. and He, K., 2017. Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1492-1500).

[17] Chollet, F., 2017. Xception: Deep learning with depth wise separable convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1251-1258).

[18] Tan, M. and Le, Q., 2021, July. Efficientnetv2: Smaller models and faster training. In International Conference on Machine Learning (pp. 10096-10106). PMLR.

[19] Szegedy, C., Ioffe, S., Vanhoucke, V. and Alemi, A.A., 2017, February. Inception-v4, inception-resnet and the impact of residual connections on learning. In Thirty-first AAAI conference on artificial intelligence.

[20] P. Tang, Q. Liang, X. Yan, S. Xiang and D. Zhang, "GP-CNN-DTEL: Global-Part CNN Model With Data-Transformed Ensemble Learning for Skin Lesion Classification," in IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 10, pp. 2870-2882, Oct. 2020, doi: 10.1109/JBHI.2020.2977013.

[21] Imran Ul Haq, Haider Ali, Hong Yu Wang, Cui Lei, Hazrat Ali, Feature fusion and Ensemble learning-based CNN model for mammographic image classification, Journal of King

Saud University - Computer and Information Sciences, Volume 34, Issue 6, Part B, 2022,Pages 3310-3318, ISSN 1319-1578, https://doi.org/10.1016/j.jksuci.2022.03.023.
[22] C. Jia, C. L. Li and Z. Ying, "Facial expression recognition based on the ensemble learning of CNNs," 2020 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), 2020, pp. 1-5, doi: 10.1109/ICSPCC50002.2020.9259543.
[23] P. Monkam et al., "Ensemble Learning of Multiple-View 3D-CNNs Model for Micro-Nodules Identification in CT Images," in IEEE Access, vol. 7, pp. 5564-5576, 2019, doi: 10.1109/ACCESS.2018.2889350.
[24] Das, A. Adaptive UNet-based Lung Segmentation and Ensemble Learning with CNN-based Deep Features for Automated COVID-19 Diagnosis. Multimed Tools Appl 81, 5407–5441 (2022). https://doi.org/10.1007/s11042-021-11787-y
[25] Shah, A.R., Gardner, T.W. Diabetic retinopathy: research to clinical practice. Clin Diabetes Endocrinol 3, 9 (2017). https://doi.org/10.1186/s40842-017-0047-y
[26] Dai, L., Wu, L., Li, H. et al. A deep learning system for detecting diabetic retinopathy across the disease spectrum. Nat Commun 12, 3242 (2021). https://doi.org/10.1038/s41467-021-23458-5
[27] Hemanth, D.J., Deperlioglu, O. &Kose, U. An enhanced diabetic retinopathy detection and classification approach using deep convolutional neural network. Neural Comput&Applic 32, 707–721 (2020). https://doi.org/10.1007/s00521-018-03974-0
[28] Schuster AK, Erb C, Hoffmann EM, Dietlein T, Pfeiffer N: The diagnosis and treatment of glaucoma. DtschArztebl Int 2020; 117: 225–34. DOI: 10.3238/arztebl.2020.0225
[29] Fu, H., Cheng, J., Xu, Y., Liu, J. (2019). Glaucoma Detection Based on Deep Learning Network in Fundus Image. In: Lu, L., Wang, X., Carneiro, G., Yang, L. (eds) Deep Learning and Convolutional Neural Networks for Medical Imaging and Clinical Informatics. Advances in Computer Vision and Pattern Recognition. Springer, Cham. https://doi.org/10.1007/978-3-030-13969-8_6
[30] Ajitha, S; Akkara, John D1,; Judy, M V. Identification of glaucoma from fundus images using deep learning techniques. Indian Journal of Ophthalmology: October 2021 - Volume 69 - Issue 10 - p 2702-2709 doi: 10.4103/ijo.IJO_92_21
[31] Samiksha Pachade, Prasanna Porwal, DhanshreeThulkar, ManeshKokare, Girish Deshmukh, Vivek Sahasrabuddhe, Luca Giancardo, GwenoléQuellec, Fabrice Mériaudeau, November 25, 2020, "Retinal Fundus Multi-disease Image Dataset (RFMiD)", IEEE Dataport, doi: https://dx.doi.org/10.21227/s3g7-st65.
[32] Budai, Attila; Bock, Rüdiger; Maier, Andreas; Hornegger, Joachim; Michelson, Georg. Robust Vessel Segmentation in Fundus Images. International Journal of Biomedical Imaging, vol. 2013, 2013
[33] J.J. Staal, M.D. Abramoff, M. Niemeijer, M.A. Viergever, B. van Ginneken, "Ridge based vessel segmentation in colour images of the retina", IEEE Transactions on Medical Imaging, 2004, vol. 23, pp. 501-509.
[34] Kovalyk O, Morales-Sánchez J, Verdú-Monedero R, Sellés-Navarro I, Palazón-Cabanes A, Sancho-Gómez JL. PAPILA: Dataset with fundus images and clinical data of both eyes of the same patient for glaucoma assessment. Sci Data. 2022 Jun 9;9(1):291. doi: 10.1038/s41597-022-01388-1. PMID: 35680965; PMCID: PMC9184612.
[35] Jin, K., Huang, X., Zhou, J. et al. FIVES: A Fundus Image Dataset for Artificial Intelligence based Vessel Segmentation. Sci Data 9, 475 (2022). https://doi.org/10.1038/s41597-022-01564-3
[36] Proc. SPIE 10579, Medical Imaging 2018: Imaging Informatics for Healthcare, Research, and Applications, 105790B (6 March 2018); doi: 10.1117/12.2293584
[37] Zhang Z, Yin FS, Liu J, et al. ORIGA(-light): an online retinal fundus image database for glaucoma analysis and research. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference. 2010 ;2010:3065-3068. DOI: 10.1109/iembs.2010.5626137. PMID: 21095735.
[38] Prasanna Porwal, Samiksha Pachade, Ravi Kamble, ManeshKokare, Girish Deshmukh, Vivek Sahasrabuddhe, Fabrice Meriaudeau, April 24, 2018, "Indian Diabetic Retinopathy Image Dataset (IDRiD)", IEEE Dataport, doi: https://dx.doi.org/10.21227/H25W98.
[39] Renoh Johnson Chalakkal, Waleed H. Abdulla, and S. Sinumol. 2017. Comparative Analysis of University of Auckland Diabetic Retinopathy Database. In Proceedings of the 9th International Conference on Signal Processing Systems (ICSPS 2017). Association for Computing Machinery, New York, NY, USA, 235–239. https://doi.org/10.1145/3163080.3163087
[40] Devare, M. H. (2019). Cloud Computing and Innovations. In G. Kecskemeti (Eds.), Applying Integration Techniques and Methods in Distributed Systems and Technologies (pp. 1-33). IGI Global. https://doi.org/10.4018/978-1-5225-8295-3.ch001
[41] Devare, M. H. (2019). Convergence of Manufacturing Cloud and Industrial IoT. In G. Kecskemeti (Eds.), Applying Integration Techniques and Methods in Distributed Systems and Technologies (pp. 49-78). IGI Global. https://doi.org/10.4018/978-1-5225-8295-3.ch003
[42] AjitkumarShitole, Dr. Manoj Devare, "Optimization of IoT Enabled Physical Location Monitoring using DT and VAR", International Journal of Cognitive Informatics and Natural Intelligence, ISSN: 1557-3958, Vol: 15, Issue: 4, Oct 2022