Protovision: Utilizing Prototypical Networks for Retinal Diseases Classification Based on Few-Shot Learning

by

Sheikh MD. Nafis Noor Nabil

23341121

Sabir Ahmed

20301189

Naimul Haque Chowdhury

23341124

Farhana Eyesmeen Maria

23341127

A thesis submitted to the Department of Computer Science and Engineering

in partial fulfillment of the requirements for the degree of

B.Sc. in Computer Science

Department of Computer Science and Engineering

School of Data and Sciences

Brac University

Jan 2024

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.

2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Sheikh MD. Nafis Noor Nabil

23341121

Sabir Ahmed

20301189

Naimul Haque Chowdhury

23341124

Farhana Eyesmeen Maria

23341127

Approval

The thesis/project titled "Protovision: Utilizing Prototypical Networks for Retinal Diseases Classification Based on Few-Shot Learning" submitted by

1. Sheikh MD Nafis Noor Nabil (23341121)

2. Sabir Ahmed (20301189)

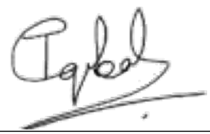
3. Naimul Haque Chowdhury (23341124)

4. Farhana Eyesmeen Maria (23341127)

of Fall, 2023 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on January 18, 2024.

Examining Committee:

Supervisor:



Dr. Muhammad Iqbal Hossain

Associate Professor

Department of Computer Science and Engineering

BRAC University

Co-Supervisor:

Rafeed Rahman

Lecturer

Department of Computer Science and Engineering

BRAC University

Thesis Coordinator:

Dr. Md. Golam Rabiul Alam

Professor

Department of Computer Science and Engineering

BRAC University

Head of Department:

Dr. Sadia Hamid Kazi

Chairperson and Associate Professor

Department of Computer Science and Engineering

BRAC University

Abstract

Classifying retinal diseases with a higher accuracy rate is one of the most important means in the medical field. In the case of image classification, finding a dataset becomes a significant challenge for such cases. As a result, the accuracy rate of classification keeps deteriorating. To address this issue of data scarcity and improve the accuracy rate, the Few-Shot method has been proposed. The few-shot learning algorithms integrated into upgraded image classification techniques have been used to enhance retinal images. VGG19 and ResNet50 have been used for feature extraction and VGG19 has given promising results comparatively. Nonetheless, a variation of training episodes was evaluated to acquire the optimal outcome. The proposed method was tested on 4 new classes that are completely different from the training classes and 82% test accuracy was obtained. This acquired result leaves a further scope for potential applications of Few-Shot learning techniques in this medical field.

Keywords: Meta-learning; Deep Learning; Few-Shot learning; Prototypical Network; Retinal Disease; Classification; Image Processing; Retinal Fundus Image.

Acknowledgement

Firstly, all praise to the Great Allah for whom our thesis has been completed without any major interruption.

Secondly, to our supervisor Dr. Muhammad Iqbal Hossain sir and our co-supervisor Mr. Rafeed Rahman sir for his kind support and advice in our work. We are immensely grateful for all the guidance and support.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

The most important sense for human beings is vision. Loss of vision or visual impairment degrades the quality of personal and social life of a person. Nowadays, retinal diseases have spread all over the world, and without proper diagnosis and quick treatment, people may end up losing their vision. This visual impairment or irreversible blindness is caused by diseases such as Diabetic Retinopathy (DR), Age-Related Macular Degeneration (AMD), Glaucoma or Cardiovascular Disease and so on[5]. These diseases can be life-threatening if not correctly diagnosed and treated at the right time. Hence early classification and the accuracy rate in the case of retinal diseases are very important phenomena. However, the accurate detection of these diseases is not easy and is not cost-friendly. As Bangladesh has a low-incoming population, it is more prone to the fatality of retinal diseases.

The wonders of artificial intelligence have started working in the medical field as well. Specifically, in the case of ophthalmic disease classification, AI has made notable progress over the years. The accessibility and efficiency of artificial intelligence in the realm of medical science is something that can vouch for its promising prospects in ophthalmic disease detection. AI advancement has provided multiple algorithms such as machine learning and deep learning which brought revolutionary changes in retinal disease detection[19]. Starting from the process to the accuracy rate of correct diagnosis, these algorithms have given nothing but positive prospects in the realm of ophthalmology.

Subsequently, deep learning algorithms are being frequently used for classification detection purposes. The few-shot learning algorithm is one of the subfields of deep learning that has brought prolific chances in the case of disease detection. The few-shot learning algorithm has improved the outlook on the classification problem's solution [16]. It is an extended form of meta-learning techniques of supervised learning[31]. Meta-learning enforces a model to acquire the skill of learning. It is often described as "learning to learn" which is a subcategory of supervised learning[4]. It aims to improve the outcomes and increase the efficiency of the algorithms through modification. Additionally, the Few-Shot Learning algorithm can identify a pattern from the dataset combined with a small number of training samples that contain supervised information. It is a remarkable approach that solves the problem of insufficient data in the case of training models. For instance, cases like medical fields require large datasets to develop a reliable detection model. However, data scarcity is a common problem in the medical field as collecting annotated datasets for emerging diseases like retinal diseases. Hence it has become an obstacle in the way of machine or deep learning algorithms. This is when the few-shot learning classifier is used to solve issues when certain disease detection cases lack enough images to feed the model during the training phase. In addition to that, for analyzing images such as CT scans or OCT from the medical dataset, image processing is used. It is a technique to differentiate features or abnormalities that is important for any kind of detection purposes. To use image classification, usually, neural network models like CNN are used for training purposes. Moreover, this paper focuses on retinal disease detection using a few shot-learning techniques that require deep learning, image processing, and image classification methods.

1.1 Problem Statement

Prevention of retinal diseases requires a proper diagnosis that can help in providing effective treatment as soon as possible. However, data scarcity has always been an issue in the medical field, especially in the context of retinal diseases. In real-life scenarios, it is hard to find sufficient image data not to mention the complexities of analyzing the image datasets. Hence in this paper, to reduce the complexity of image analysis and solve the issue of limited datasets in this medical sector, the few-shot learning approach will be used. This paper intends to overcome the challenges of data scarcity and complex image classification so that we can develop a robust model. Hopefully, this model will improve the work that has been done before in this regard of accurate diagnosis and early intervention in retinal diseases.

1.2 Research Objectives

1. Finding the limitations of the medical section of retinal disease detection.

2. Figuring out the suitable image classification technique for detecting retinal disease.

3. Understanding few-shot learning and building a robust hybrid model with fewer parameters, less time complexity, and more accurate disease detection.

4. Evaluate the model's performance through rigorous experiments using a limited amount of data.

5. Determine if the developed model can predict the diseases accurately.

1.3 Thesis Structure

Chapter 1 discusses a brief background of retinal diseases, data scarcity in the medical sector, the research problems that motivated our thesis, the contribution we have made through our work, and the overall thesis outline. **Chapter 2** gives a summary of the background study that was done on previously published related to our research topic. **Chapter 3** holds our complete background study of the models we have used in our research work. **Chapter 4** talks about the proposed methodology in detail. Starting from the workflow diagram, an overview of the full dataset along the pre-processing phases has been shown in this section. **Chapter 5** gives a brief idea about the implementation of our research work. **Chapter 6** portrays a clear view of the whole result analysis process and the future scopes that can be explored. Finally, **Chapter 7** concludes our research work with a brief discussion of the work we have done and how we can enhance the performance in the future.

Chapter 2

Literature Review

A literature review is an overview of all the works that have been published or are about to be published on a particular topic. Every research builds its base upon previous research works so that the previous results can be improved or new perspectives can emerge. Added to that, we have also gone through a huge amount of previous work that is related to our chosen topic. In this portion of our paper, we are going to cluster the summary of a few papers that are related to our research work.

In a paper [8] the researcher tried to develop a comprehensive fundus multi-disease screening method by combining data from common ophthalmic disease images. The images were taken from SYSU datasets, IChallenge-AMD, IChallenge-GON, and IChallengePALM [8]. The main intention was to simulate camera characteristics for data augmentation and construct a strong disease detection model using metric-based few-shot learning. At first, they presented a dataset that is a combination of different ophthalmic diseases. Also, the perspective of having less amount of datasets in the medical field has been taken into consideration. For instance, in the mentioned scenario, the data distribution was imbalanced and the number of datasets was scarce. Due to that issue, they used a Siamese network from the perspective of few-shot learning to solve the overfitting problem. Siamese-Resnet was done on unenhanced data and the types with insufficient samples. The classification effect of the modular Siamese-Resnet has improved quite a bit. Finally, the experimental results prove that Siamese-Resnet-50 is well-fit for the screening and classification task of ophthalmic diseases[8]. During the training process, data augmentation was done by style transfer and then a neural network model was trained so that it does not learn features other than the diseases. The original dataset was divided so that different styles could be trained accordingly. For each fundus photo that is under a particular style, k-1 style transfer models with various styles are used and then the generated fundus photos are added for augmentation[8]. That is how sample numbers of the data have expanded which contributed to improving the model performance. For a comprehensive evaluation of the model, accuracy, recall, and f1 score has been used. Lastly, this paper worked on building a classification model on a dataset made from a common image dataset of ophthalmic diseases and analyzed the feasibility of using metric-based few-shot learning that improves the accuracy rate of the model.

In this [27] paper, the authors proposed the prototypical network for classification based on the few-shot method. The researchers suggested an approach where the classifier must generalize in the context of new unseen classes that are not present in the training set given a small amount of examples per class. The authors have worked on multiple datasets such as Omniglot, miniImageNet, and CU-Birds dataset. They have extended their experiment starting from the prototypical network to zero-shot learning. They have applied approaches like matching networks, neural statisticians, or meta-learner LSTM on these datasets. However, when applied to prototypical networks, they have received state-of-the-art accuracy. In their paper, they used Euclidean distance to calculate the distance between the images and prototypes or classes and used the average cross-entropy to calculate the loss. The authors generalized a prototypical network to work on zero-shot learning, and the algorithm archives state-of-the-art results on the CUB-200 dataset. Moreover, writers have also conducted preliminary tests for learning a variance per dimension for each class and learned that embedding network has enough flexibility without any fitted parameters per class. In future, they hoped to try Bregman divergences instead of Euclidean distance. All the effectiveness makes the prototypical network a favorable approach for few-shot learning.

In the paper [28] the authors discussed retinal disease detection methods using OCT (Optical Coherence Tomography) images and employing deep learning and transfer learning techniques. Pre-training models as feature extractors and fine-tuning the top layers of the models were the two approaches that the authors compared. Some of the models that obtained better accuracy are InceptionV3, Xception, and DenseNet201[28]. The paper concludes that models that use feature extractors are less accurate than fine-tuning. Combining pre-trained models, hyperparameter optimization, and fine-tuning led to increased efficiency and reduced training time. Bayesian optimization is given priority in the case of hyperparameter optimization. After utilizing optimal values for the model's hyperparameters, the paper obtains better results. After fine-tuning, the classifier achieved an accuracy of more than 95%[28]. Compared to default values this method gives increased precision to classify retinal diseases. The limitations of the model are identified through error analysis, for instance, the misclassified data have been examined to enhance classification accuracy[28]. The application of this study can also be extended to other retinal diseases. To establish a balance between model performance and training time, future exploration of hyperparameters is advised. To summarize, it can be said that this study demonstrates that deep learning techniques are improved, have better potential, and provide promising solutions for retinal disease detection. These results can help patients in the upcoming years by enabling accurate and structured diagnoses.

The research work [7] proposes the detection and classification of diabetic retinopathy (DR) and glaucomatous conditions using models OF deep learning, such as VGG16 and AlexNet. For detection purposes, the authors used two datasheets (High-Resolution Fundus Images (HRF) and Messidor) to achieve early detection of

DR. The HRF dataset consists of 45 fundoscopic images that are categorized into three classes (Healthy, Diabetic Retinopathy, and Glaucomatous). The Messidor dataset had 1200 color(RGB) images and, it was a combination of four classes. [7]. The dataset's images differ in resolution and quality. The authors used grayscale conversion, GHE (Global Histogram Equalization), Otsu thresholding, and CLAHE (Contrast Limited Adaptive Histogram Equalization) to preprocess and upgrade the deep learning model's performance[7]. Grayscale conversion is used to simplify the analysis. On the other hand, GHE enhances contrast. For the segmentation of images by clustering-based thresholding, Otsu thresholding is used. CLAHE is applied to further improve contrast while maintaining infrequently occurring pixel intensities. Pretrained CNN architectures (AlexNet and VGG16) are used to classify the images. These models are trained and tested on the datasets. Their performance is evaluated based on 2-ary, 3-ary, and 4-ary classifications. The VGG16 model performed better than the AlexNet model on the HRF dataset, achieving higher training accuracy (98.70%) and validation accuracy (73.33%). For this dataset, AlexNet is a very good fit for 2 ary classifications[7]. Despite that, the accuracy result is insignificant. VGG16 outperformed AlexNet in terms of accuracy, although both models had struggled with the subtle features of diabetic retinopathy images. The authors also addressed overfitting and underfitting problems with methods like batch normalization and dropout. Limitations in computational resources restricted the model's ability to extract all possible image features, affecting classification accuracy.

In the paper [32] the authors tried to upgrade the application of deep learning in the diagnosis of rare diseases using optical coherence tomography (OCT). This is essential because in many cases, rare diseases are overlooked due to the lack of data. Few-shot learning (FSL) allows pattern extraction with a limited amount of data. According to the authors, disease detection using the concept of Few Shot Learning with OCT was rarely something that was taken under consideration. Their approach was based on FSL which used GAN (CycleGAN) for image generation without matching paired images. They had an image dataset of OCT acquired from an existing work of Kermany. They also collected OCT images of retinal diseases that were unidentified [32]. Their dataset contained OCT images of usual retinal diseases. But it also had OCT images of drusen, diabetic macular edema, and choroidal neovascularization [32]. There were nine classes in their preliminary training dataset, including normal retinas. They split the training and test randomly which showed no overlap. The authors had built CycleGAN augmentation models that were used for rare retinal disease[32]. These models were trained on two domains called normal retina and a specific disease. They used linear and elastic transformation augmentation techniques and the few-shot algorithm on the OCT images of rare diseases. The prepared images for each disease were resized for the CNN model. The t-SNE algorithm was applied to evaluate the effects of CycleGAN-based augmentation. After data augmentation, a CNN using the Inception-v3 model was trained for multi-class diagnosis of diseases. 0.1% of the training dataset was taken for validation data. For training, ADAM optimizer, and for generating the saliency map, the Grad-CAM technique was used. The authors also implemented few-shot learning methods such as prototypical networks and Siamese neural networks. Synthetic OCT image dataset of Stargardt disease and retinitis pigmentosa gave higher rejection rates due to overlapped features, low quality, and mode collapse [32]. The algorithm t-SNE displayed improved cluster formation and generalizability after CycleGAN-based augmentation. The model with transfer learning and GAN-based data augmentation performed the best. In the first validation, it achieved overall accuracy of 93.9% and in the second, it achieved 92.1% accuracy on the test dataset. The model performed better than the human expert in diagnosing rare diseases except for macular telangiectasia.

In this paper [3] the authors tried to figure out if deep learning methods based on low-shot are effective for automated retinal diagnostics. There are a lot of existing works done by researchers taking the aspect of limited availability of datasets into consideration. They thought that it might be beneficial in a situation involving a rare retinal disease or where potential biases may exist in the data that does not exist in the training set. For the baseline, they evaluated many DL algorithms and reached a strategy where they will pretrain a ResNet 50 network on imageNet (RES-FT) [3]. This baseline model was then compared to the existing algorithms and then the low-shot learning was applied. The image represented is encoded via the same ResNet 50. One of the three classifiers such as random forest, support vector machine, and k-nearest neighbors was applied to the processed images. For this research, the researchers used a public domain fundi to obtain 44346 participants from eyePACS [3]. Their preprocessing includes cropping each image to the circumscribed square around the retina, and padding where needed. Then they partitioned the data into three sets such as train, test, and validation sets. Given the low-shot approach, the training set was limited, with N samples per class. As a result, a total of 2N training samples per experiment was allocated. According to the Res-FT method, the dataset was split into an 80% training set and a 20% testing set [3]. The limitation of this research is that it focused only on one retinal disease which is diabetic retinopathy without considering other potential diagnostic scenarios. Another limitation is the biological variability in each disease. People from different geographical locations, ethnicities, and races have different amounts of melanocytes in their bodies. So, insufficient data in these cases may lead to low performance of the low shot algorithm. In conclusion, we can say that if enough data is present, we can use the usual DL approaches but in the case of low datasets, the low shot approach gives better results.

In this paper [17], the author analyzes the use of Few-Shot Transfer Learning for recognizing hereditary retinal diseases using limited OCT images. The scarcity of data is the main challenge of this paper. To overcome this issue the authors present a student-teacher learning framework but it has its challenges. To overcome them, the authors employed a preprocessing pipeline that applied image alignment and feature extraction. Previous works were done by using deep learning methods but they needed a large amount of data. To address this limitation, the authors introduced a three-part pipeline where the first part involved image processing (normalized to reduce noise), the second part focused on training the teacher-student model, and the last part expressed the few-shot learning and its limitation. The authors admitted the challenges of few-shot learning in supervised machine learning, particularly empirical risk minimization and its unreliability. They considered data, models, and algorithms to mitigate these issues. They used a ResNet-50 teacher model for the auxiliary dataset and employed the Soft Nearest Neighbor Loss (SNNL) technique [17]. Three types of fine-tuning were performed on the target dataset, and a RasNet-18 model was used as the student model to adapt to the limited data. The results showed that the student model outperformed the teacher model in detecting retinal diseases. According to the result, the student model outperformed the teacher model in detecting retinal diseases. Even with a small dataset, Few-shot learning is feasible for identifying hereditary retinal diseases. By utilizing a student-teacher learning framework and implementing preprocessing techniques, the authors successfully addressed the challenges associated with limited data availability in this domain.

In this study [23] an advanced method demonstrated named ONCE (Object detector with Novel Class Estimation) targets the challenges of detecting novel object classes with finite training samples. Two datasets are used, the COCO dataset with 80 object classes and the PASCAL VOL dataset with 20 classes[23]. The performance of ONCE was compared to two state-of-the-art models by the authors in the non-incremental setting. Alongside the base class data, 10 shots per novel class for training were used. The results indicate that ONCE approaches the performance of the existing models, demonstrating its effectiveness in few-shot object detection. On the other hand, for the incremental setup, an experimental setup for evaluating ONCE was presented while adding certain modifications to the previous dataset. However, during meta-testing the base class training data is not available, and there is a requirement for incremental updates for novel classes. The models are trained on COCO and PASCAL VOC tested, and both the same-dataset and cross-dataset scenarios were considered for evaluation performance. The method ONCE was compared with multiple alternate methods, such as fine-tuning, Feature-Reweight, and MAML in this paper. The performance of these methods was reported using average precision (AP) and average recall (AR)[23]. The authors conclude that ONCE achieves superior performance in detection for both classes. Furthermore, other methods fail to address the problem of catastrophic forgetting, while ONCE is efficient in this manner. The transfer performance of ONCE from the COCO to VOC dataset is also studied in this paper. The capabilities of the ONCE method prevail in most aspects, including the test domain. The DeepFashion2 benchmark is used by the authors to assess ONCE for landmark detection in addition to object detection[23].ONCE can effectively handle the hierarchical semantic structure of fashion landmark detection tasks and achieve competitive results where the dataset consists of clothing items with multiple landmark classes. In conclusion, the experimental findings support ONCE's efficacy in both incremental and non-incremental few-shot object detection. It outperforms existing methods in terms of detection accuracy as well as solving the issues of catastrophic forgetting. The versatility and potential in real-world applications of ONCE are further demonstrated by its applicability to fashion landmark identification and transferability to various datasets.

Another paper [14] shows it can achieve pre-diagnosis by analyzing the retina from OCT images. The authors stated multiple previous works using OCT images based on deep learning and specified works on CNN. There was no such work on OCT image classification based on VGG-16. Because of this, the authors utilized a deep learning model to diagnose diseases which was based on the visual geometry group 16 and pre-trained on the ImageNet dataset. They used the dataset containing different images of diseases like drusen, macular edema, choroidal neovascularization, and also normal retinas[14]. To the labeled image dataset, they applied image normalization. To enhance the quality of the images, the authors used OCT automatic real-time averaging. The dataset was divided into validation and training data. The predictions were based on the output generated by the VGG-16 network. This model learned to differentiate different OCT images of distinct groups. Thirteen convolution layers, three fully connected layers, and five max-pooling layers made up the VGG-16 network. In the multi-class comparison, the model achieved an accuracy of 98.6% on the validation data. It had 97.8% of sensitivity and 99.4% of specificity. Analysis of the ROC curve demonstrated a 100% area under the curve. The model's accuracy reached 98.6% on the validation data and 96.63% on training data[14]. The authors also implemented binary classifiers which showed 100% accuracy for CNV, 98.8% accuracy for DME, and 99.2% accuracy for DRUSEN. The proposed approach outperformed common approaches and achieved great results. Future work will involve developing software that professionals in medical centers can utilize to implement the suggested approach[14].

In a research [13] the researchers aimed to work on a classification problem of four classes. A model was developed to detect images like diabetic macular edema, choroidal neovascularization, DRUSEN, and normal retinal diseases through OCT images whereas each of the OCT images used ResNet50 as their core neural network. The dataset of a vast amount of retinal OCT images taken in real scenarios was processed to combine four improved ResNet 50 for automated classification of mentioned retinal diseases. For the dataset, 21,357 OCT retinal images were collected from the Shanghai First People's Hospital and the Shanghai Zhongshan Hospital[13]. The enhanced ResNet50 was combined with pooling layers, convolutional layers, and fully connected layers. Data augmentation was done to make performance better and avoid overfitting issues. Then the images from the dataset were resized by Gaussian pyramid-down sampling. Transfer learning has been used to improve the ResNet50 which was pre-trained on a huge dataset. The whole dataset was divided into various folds for cross-validation whereas each of them had unique training and test datasets[13]. Based on them the parameters of the model were optimized. The researchers did qualitative assessments and occlusion testing to identify the important parts of the images so that the classification process gets easier. For evaluation purposes, with 95% confidence intervals overall accuracy, sensitivity, and area under the curve was used. Then while evaluating the model's performance on an independent testing dataset for disease classification at the B-scan level, it had given a specificity rate of 98.5%, sensitivity rate of 96.3%, and accuracy rate of 97.3%, Also the ROC curve analysis had given the high area under the curve of 0.995. Moreover, three binary classifiers were employed to differentiate the diseases. The binary classifiers had acquired a high accuracy rate of 0.978 to 0.988, sensitivity range of 0.981 to 0.986, specificity range of 0.991 to 0.993, and AUCs range of 0.993 to 0.997[13]. The performance of the model was tested on DHU and UCSD datasets where in the case of DHU, the result was comparatively better. Lastly, the researcher lacked the vision to take the necessity of exploring the integration of multimodal data to improve the performance into consideration and to incorporate longitudinal images for enhancing the accuracy rate of the predictions.

In this paper [25]the authors have taken a neural network approach that is based on CNN. It was used to develop a model that classifies correct images from disease pathologies. It intends to acquire higher accuracy in this regard than collected from previous works. Hence in this work, the researchers have evaluated multiple deep neural network frameworks[25]. For training and testing purposes, different preprocessing techniques have been taken and various augmentation styles have been taken under consideration to increase the accuracy rate and the sample size of the dataset for enhancement purposes. Datasets were collected from the Messidor, Messidor-2, DRISHTI-GS, and publicly accessible datasets of retinal diseases. To incorporate the classification of the diabetic eye disease dataset, pre-trained CNNs on ImageNet named VGG and InceptionVN were used. Here VGG16 has shown better results comparatively[25]. Then again, these two models' accuracy was compared against the already existing one on the test dataset. Fine-tuning was evaluated for the accuracy rate. The highest accuracy rate was determined with the help of optimizers. While testing in DED detection tasks, the first CNN model showed that the sensitivity is 0.85 and the specificity is 0.96. Both of these were the maximum. Additionally, the second model gave results of 85% sensitivity and 98% specificity[25]. Both of the cases were up to the standard of the British Diabetic Association. This automated classification model had shown a better accuracy rate than previous works.

In this paper [1] the classification system had been built for detecting five different retinal diseases. These diseases are age-related macular degeneration (AMD), diabetic macular edema (DME), DRUSEN choroidal neovascularization (CNV), and normal retinal diseases. The system was built upon a novel CNN architecture that classifies four mentioned diseases and normal retinal diseases that are not rare. The process was done using a huge SD-OCT dataset of images and had a higher accuracy rate [1]. In the training phase, the collected dataset has been denoised and resized according to the necessity. Then the dataset was divided into 70% and 30% consecutively for the train and test set. The testing set had 1250 images which was divided into five classes of acquiring 250 images per class. Then it was fed to the CNN model. To train the model, an adaptive moment estimation optimizer has been used and the hyperparameters are chosen by trial and error. The results showed that the learning rate initially was 0.001 and 64 was the size of the mini-batch. Then again, the highest number of epochs was 100 whereas the squared gradient decay factor was 0.9900 and the gradient decay factor was 0.9000. Moreover, Epsilon was 1.0000e08, and the regularization of L2 was 1.0000e04 but the gradient threshold was infinity. Lastly, the total number of training iterations was 74,400 whereas for validation it was only 744. Additionally, the training set had given 0.996 accuracy but the test set showed 0.944 which was less than the training set. Despite that, the CNN architecture has given better results compared to the other methods. This proposed CNN architecture that used a softmax classifier had 0.953 accuracy overall. This had correctly 100% of AMD cases which was the highest. Then for CNV, DME and DRUSEN it was 0.9917, 0.9897 and 0.9915. Subsequently, the accuracy rate for usual retinal diseases was 95.30%.In the future, this work intends to focus on generalizing the system and reducing the error rate.

In the paper [15] the researchers tried to implement a method to do the task of retinal vessel image segmentation accurately. Here, the attention method was paired with a convolutional neural network by the authors. Important biological information is found in the retinal blood vessels. It can be used clinically to identify disorders like hypertension and diabetes, as deviations in the structure of the retinal vasculature can serve as early indicators of these conditions [15]. They showed some previous work on image segmentation based on supervised and unsupervised learning. Authors in this paper used U-net architecture which focuses on a small area at a time for accurate segmentation. The segmentation process is made more effective by the attention module, which gathers global data and enhances features. There are three parts of the attention mechanism. One is feature similarity calculation, the second is feature extraction and the last one is original feature enhancement [15]. In this method, fewer parameters are used compared to previous approaches. Five datasets were used to evaluate the performance. They are (i) STARE, (ii) DRIVE, (iii) IOSTAR, (iv) RC-SLO, and (v) CHASE DB1 [15]. The proposed method achieves MCC values of 0.9207, 0.8002, and 0.7791 on the STARE, DRIVE, and CHASE DB1 datasets, respectively [15]. Performance on the IOSTAR and RC-SLO datasets is noticeably superior to that of the current approaches. The outcomes indicate that the proposed method achieved satisfactory performance for the segmentation task on multiple datasets.

In the research [10] the researchers proposed a hybrid method to diagnose diabetic retinography from retinal fundus images [10]. If untreated, diabetic retinopathy, a consequence of diabetes, can cause vision loss. To detect this, they used a method that includes image processing along with deep learning, specifically Convolutional Neural Networks. 400 retinal fundus photos were taken by the researchers from the MESSIDOR database. [10]. They resized the images and separated them into color components. They employed the techniques of contrast-limited adaptive histogram equalization (CLAHE) and histogram equalization (HE) on each of the components. Then they applied CNN for the classification. The convolutional layer contained two feature maps which are obtained from the outputs of nodes in the 2x2 receptive area of the previous layer. 3D weight tensors represent the weights and values. The pixel values are collected within the feature map's context using the mean dimension, which summarizes the entries of the feature maps. The CNN contained eight layers in total. The model's performance was assessed using its accuracy, F-score, specificity, sensitivity, recall, and G-mean. At the initial stage of image resizing, accuracy 91%, F-score 82%, specificity 94%, and G-mean 87% were the average values for evaluation criteria. Following the use of histogram equalization (HE), the accuracy increased by 3%, the F-score by 7%, the recall by 7%, the sensitivity by 7%, the specificity by 2% and the G-mean by 5%. After utilizing CLAHE, the average classification values increased to 97 percent accuracy, 94 percent sensitivity, 98 percent specificity, 94 percent recall, 94 percent F-score, and 95 percent G-mean. [10]. The performance shows that the CLAHE method had the best results. This research was conducted entirely based on the MESSIDOR database. The authors recommend looking into alternative solution mechanisms, incorporating different image processing techniques, and utilizing additional medical databases.

In this paper [6] the authors tried to find more effective ways of image processing to detect retinal diseases. They proposed several steps to find the most efficient way which include pre-processing of the dataset, detection of optic disc, removal of blood vessel segmentation, elimination of fovea, extraction of features, and image classification. After this, they performed a simulation using MATLAB and the DUARETDBI dataset. Pre-processes have three steps such as the conversion of HSI (Hue Saturation Intensity), DE noising, and augmentation of contrast. This is done to make RGB images clearer. To remove noise researchers used a Nonlinear wiener filter [6]. They used CHT (circular hough disc) to detect optic discs. After this comes blood vessel segmentation which is a certain benchmark for localizing fovea, lesions, and optic nerve. In their paper, they stated that they used many algorithms to accurately segment the blood vessels but these algorithms did not extract extra information from the retinal images which are why they used the Spatially Constrained Possibilistic Fuzzy C-Means algorithm (SCPFCM). Fovea elimination was also essential as it is associated with hemorrhages. The authors identified the fovea center using morphological dilation operations performed on binary images. The authors focused on three characteristics in the feature extraction stage: microaneurysms feature extraction, exudates feature extraction, and retinal hemorrhage feature extraction. They selected relevant features from these extracted features based on the deep learning concept of the DBN (Deep belief network) [6]. DBN has two parts, DBN construction, and DBN training. They applied greedy layer-wise unsupervised learning in construction, and exponential loss function along with gradient descent. That was stated upon supervised learning in the phase of training. The researchers used classification of the features from feature selection processes. After that, to separate the feature they used the SVMGA algorithm. They used many algorithms to have better results. After using all those algorithms and different types of ways to enhance the image processing we get the result that this proposed image processing method has a 99% high sensitivity, 96% specificity, and 98.4% accuracy [6]). This result concludes that their research on finding a better way of image processing is successful as it has better accuracy in detecting diabetic retinopathy diseases.

In this [24], paper the authors tried to detect papilledema disease using U-Net and Dense-Net architecture by a deep learning method. The authors presented two phases of detecting this, the first one is OD (optic disc) image extraction and the other one is a classification of that extracted image from the blood vessel segmentation of the retina. Dense-Net has been used in classification tasks and they used U-Net in blood vessel segmentation. The authors used the STARE dataset for their research and used Dense-Net to classify the fundus images from that dataset to normal and papilledema categories [24]. U-Net is trained with 32*32 pixel patches from that dataset. After that, it shows better accuracy in classifying blood vessels using 23 convolutional layers [24]. To evaluate their research, the authors used a confusion matrix and dice coefficient and after evaluation, their model shows high accuracy of 99.17%, 97.83% specificity, and 98.63% sensitivity [24]. In conclusion, they stated that their work surpassed the previous works and that the proposed method successfully differentiated between moderate and severe papilledema stages. They also stated that in the future someone who is working will need a bigger dataset to successfully train and get better results from their deep learning methods.

Chapter 3

Background Studies

3.1 Few-shot learning

Traditional machine learning algorithms work with a sufficient amount of labeled data. These algorithms often need large amounts of labeled data to generalize unseen data. In that scenario, few-shot learning has taken place which aims to train models with less amount of labeled data. Few-shot learning works well in case of insufficient data[18]. There are a lot of approaches to few-shot learning and among them, meta-learning or learning to learn is one of the popular approaches that is being used nowadays. Meta-learning trains a model on different tasks in a way that the model can adapt to new tasks quickly with fewer numbers of examples.
Few-shot learning comes to the scenario where collecting sufficient amounts of labeled data is not possible such as in the medical field of Bangladesh. It has worked well for a lot of domains like reinforcement learning, computer vision, or natural language processing.

3.2 VGG19

VGG19 is the short form of Visual Geometry Group 19. VGG19 is a type of convolutional neural network that consists of 19 layers. It includes 16 convolutional layers with 3 fully connected layers. These convolutional layers are composed of 3x3 filters and the fully connected layers use 2x2 filters just like the figure 3.1. VGG19 is widely used as it serves the purpose of classification tasks with simplicity and effectiveness. It is often used as a baseline model for developing and evaluating architectures[2].

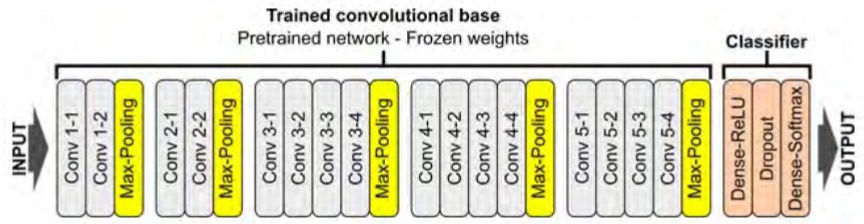


Figure 3.1: VGG19 architecture

Implementation

We used a VGG19 architecture which had pre-trained weights of ImageNet data. Initially, all the parameters in the model were frozen. The Avgpool layer was changed to an AdaptiveAvgPool2d layer to allow dynamic output size and the classifier was modified to a Flatten layer which is useful for converting 2D output from the previous layer to a 1D for the final layer. Then unfreezing the parameters after the index 34 allowed us to achieve specific trainable layers during fine-tuning. The 34th index of the VGG19 corresponds to the fully connected layer. Unfreezing these layers allows these layers to update their weights during training while keeping the pre-trained weights in the previous layers. As the lower layers capture the edges and textures, this change makes the model better for our few shot learning tasks. This was a lightweight model containing 2,359,808 trainable parameters.

3.3 ResNet50

ResNet-50 is mostly used to train very deep networks through deep neural network architecture[9]. The architecture of this model in shown in figure 3.2. This model introduced the concept of residual blocks that provide skip connections and allow the network to learn residual information. Because of ResNet50, the difference between the prediction and the expected output can be captured effectively. This helps mitigate the vanishing gradient problem and makes it simpler to train deep networks. In this model implementation, an input image is given to the network, and through the convolutional layers along with the residual blocks, the model learns to extract features. Then the model's fully connected layers transform the features into the expected output, making the model suitable for image classification[22].

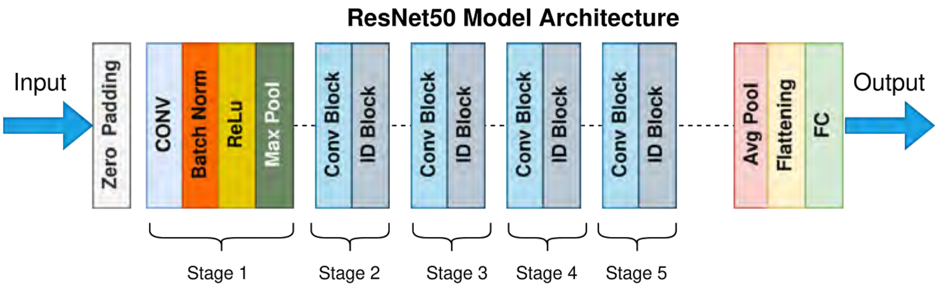


Figure 3.2: ResNet50 architecture

Implementation

ResNet-50 architecture was imported with pre-trained weights of ImageNet data. Initially, all the parameters of the model were frozen which kept the pre-trained weights unchanged. Then we fine-tuned the model unfreezing layer 4 or the last residual block enabling it to be updated during training.

3.4 InceptionV3

InceptionV3 is a well-known convolutional neural network model that consists of a layer of 48 factors into 7x7 convolutions similar to the figure 3.3. It is widely used for image classification and object recognition. Initially developed by Google as part of their convolutional neural network series, this model includes multiple convolutional layers of various sizes, enabling it to capture features at different scales. Inception V3 is known for its efficiency and high success rate in image classification tasks[29].

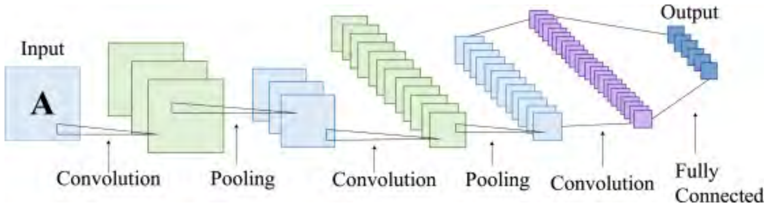


Figure 3.3: Inception V3 architecture

Implementation

The InceptionV3 model was loaded along with the pre-trained weights from the ImageNet dataset. We applied fine-tuning on the layers of the model, setting up some as non-trainable except for the last 10 layers. To convert the 4D output of the InceptionV3 base into a 1D vector, we used a flattened layer. We also added a ReLU activation and a dense layer with 256 units. A dropout layer with a rate of 0.3 was implemented to prevent overfitting. Additionally, a softmax activation function embedded into a final dense layer was added for classification purposes. We used an Adam optimizer with the same parameters as ResNet50. A ReduceLROnPlateau callback was defined to reduce the learning rate. At that time, the validation loss plateaus and an Early Stopping callback was implemented to stop training if the validation loss does not improve for 5 consecutive epochs.

3.5 Evaluation Metrics

Evaluation metrics are a way of measuring or evaluating the overall performance of any model[11]. These metrics make it easier to understand the performance of a model in various criteria.

Accuracy

Accuracy is a metric that is calculated by the ratio of correctly predicted instances to the total instances of the test set. This metric holds the overall performance of the model's predictions.

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \tag{3.1}$$

Precision

Precision is measured through the ratio of correctly predicted instances to the total number of positive instances.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{3.2}$$

Recall

Recall is the calculation of the the ratio of truly positive predictions to the total number of actual positives.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{3.3}$$

F1-Score

F1-Score is the measure of the harmonic mean of precision and recall. It provides the balanced accuracy of the model when there is a scenario of an imbalanced dataset.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3.4}$$

Chapter 4

Proposed Methodology

4.1 Workflow

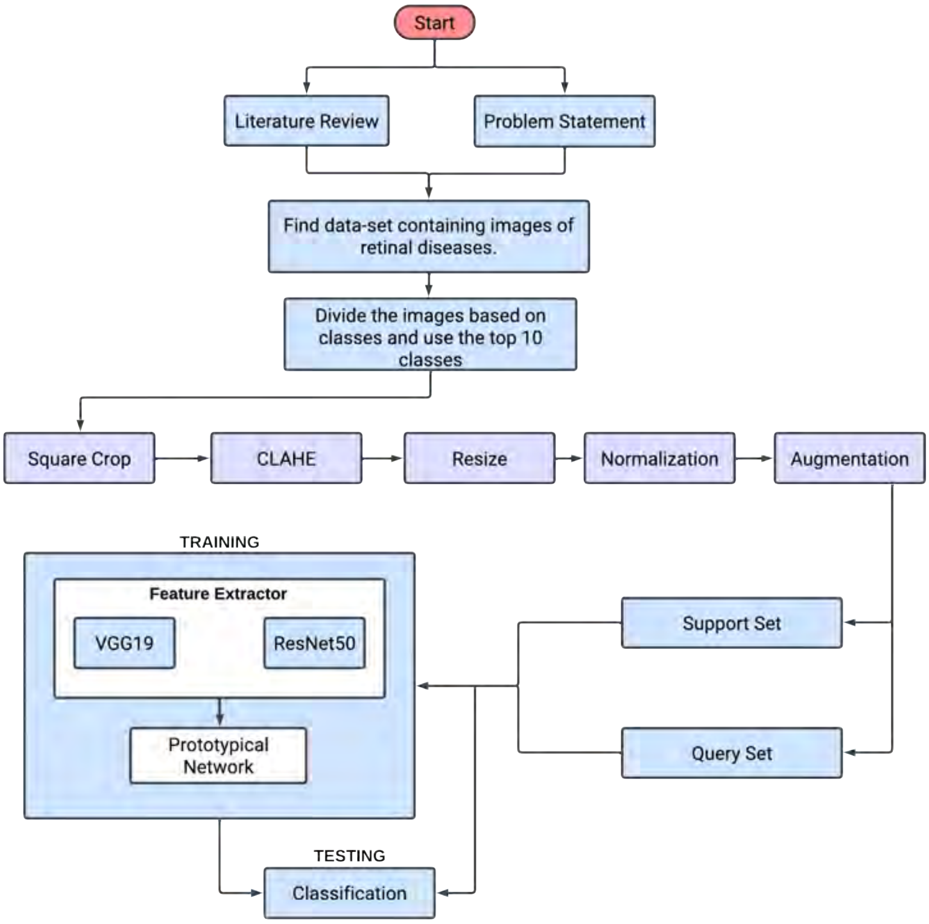


Figure 4.1: Workflow

4.2 Overview of RFMiD Dataset

For our study on 'Protovision: Utilizing Prototypical Networks for Retinal Diseases Classification Based on Few-Shot Learning,' we have used RFMiD (Retinal Fundus Multi-disease Image Dataset) as our primary source[20]. This dataset is comparatively new as it was last updated in 2023. Moreover, the RFMiD dataset is publicly available and is made up of various diseases, contributing to the application of versatile models for retinal disease detection. This is an open-access dataset from IEEE DataPort.

The RFMiD dataset consists of 3200 retina images, of which 1920 images were allocated for the training dataset, 640 images for testing purposes, and another 640 images for validation. This dataset also had a CSV file that contained the labels. Some images contained more than one disease. Three different fundus cameras were used to capture the fundus images, with resolutions of 4288x2848 (277 images), 2048x1536 (150 images), and 2144 x 1424 (1493 images). In total, 46 disease conditions were represented by various Classes. Although it contained many eye diseases, most of the classes had very few samples. The numbers are shown in table 4.1.

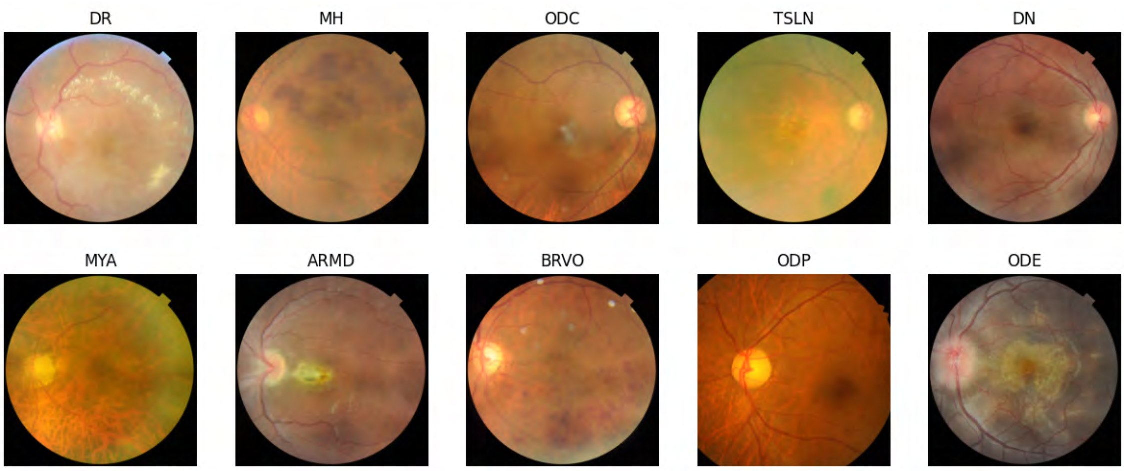


Figure 4.2: Visualization of the top ten diseases

Acronym	Full Name	Training	Validation	Total
DR	Diabetic Retinopathy	396	99	495
MH	Media Haze	135	34	169
ODC	Optic Disc Cupping	211	52	263
TSLN	Tessellation	125	31	156
ARMD	Age-Related Macular Degeneration	126	32	158
DN	Drusen	130	32	162
MYA	Myopia	71	18	89
BRVO	Branch Retinal Vein Occlusion	63	16	79
ODP	Optic Disc Pallor	50	12	62
ODE	Optic Disc Edema	46	11	57

Table 4.1: Image count per class of top ten retinal diseases

The ten classes shown in figure 4.2 from the total of 46 classes were taken for the classification task. These ten classes had the most images. The dataset is an imbalanced dataset which can be seen in figure 4.3. Every class was labeled accordingly.

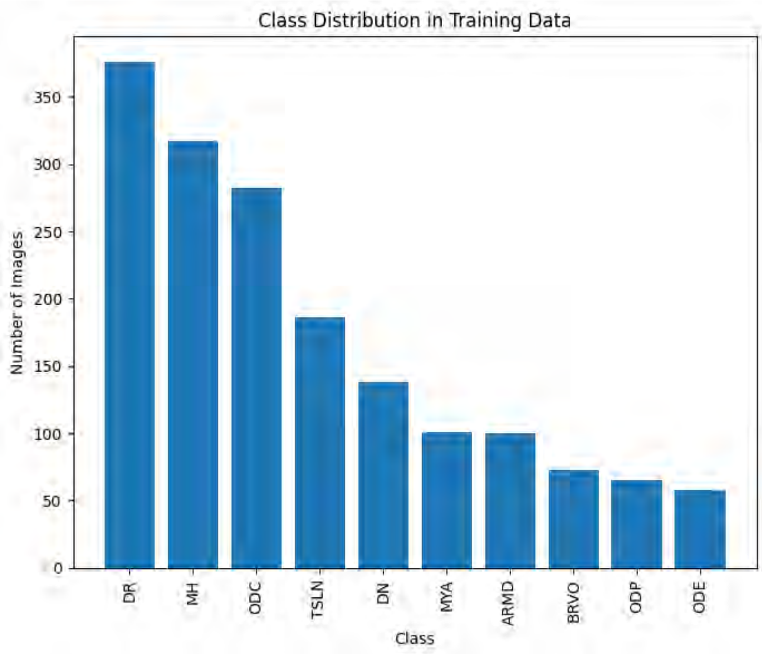


Figure 4.3: Number of images in each class

4.3 Description of the Dataset

4.3.1 Data Pre-processing

Contrast Limited AHE (CLAHE)

It is harder to find patterns and differentiate retinal fundus images as all of them are quite similar. To enhance the features we used the CLAHE on the training images so that the deep learning model can recognize the patterns better.

The Contrast Limited Adaptive Histogram Equalization (CLAHE) is a crucial image processing technique used for enhancing the local contrast of an image, particularly in scenarios where the global histogram equalization method may lead to noise amplification [26]. In this data preprocessing, the RGB image of the retinal fundus image is first converted to the LAB color space to separate luminance (L) information from color components. CLAHE is then applied exclusively to the L channel, employing a specified grid or block size for histogram equalization [26]. With its adaptive nature, CLAHE adjusts the contrast based on image characteristics, effectively handling the variability present in retinal images with diverse structures. This adaptability is vital for highlighting subtle features indicative of retinal diseases, assisting in early diagnosis, and tracking disease progression. After implementing CLAHE, the blood vessels and parts of the eye containing disease patterns were visible making it preferable for the model training (4.4).

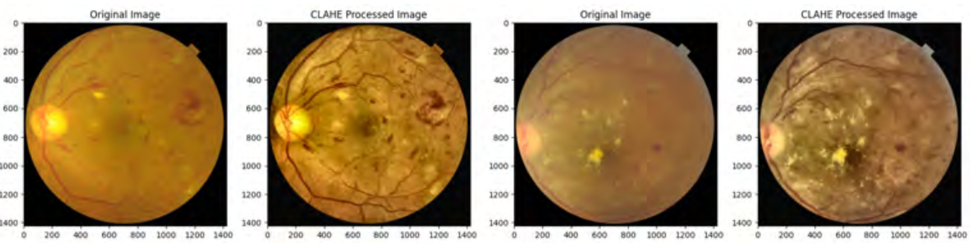


Figure 4.4: Original and CLAHE Processed Images

Augmentation

To improve the model training and achieve a satisfactory output, we implemented different augmentation techniques. The images contained a black background which did not have any useful information for the model. So to get rid of it, we manually cropped the rectangular images into square images by putting the retinal fundus images in the center. We separated the images based on their classes from the labels that were in the CSV file using the Shutil module. The images that contained multiple diseases were copied and given a sample to each disease-containing class. We took the ten classes with the highest number of images for the classification. The classes are consecutively DR, MH, ODC, DN, TSLN, MYA, ARMD, BRVD, ODP and ODE. We calculated the mean and standard deviation for our data. We used those values to normalize the images. Then we resized the images into a 224x224 shape as both the VGG19 and ResNet50 models require this input size. On the training set, we did augmentation which included 90-degree clockwise rotation, 90-degree counterclockwise rotation, 180-degree rotation, vertical flip, and horizontal flip. The augmentation increased the training images from 1400 to 9800 displayed in figure 4.5. Lastly, we used the images in the models for the classification tasks.

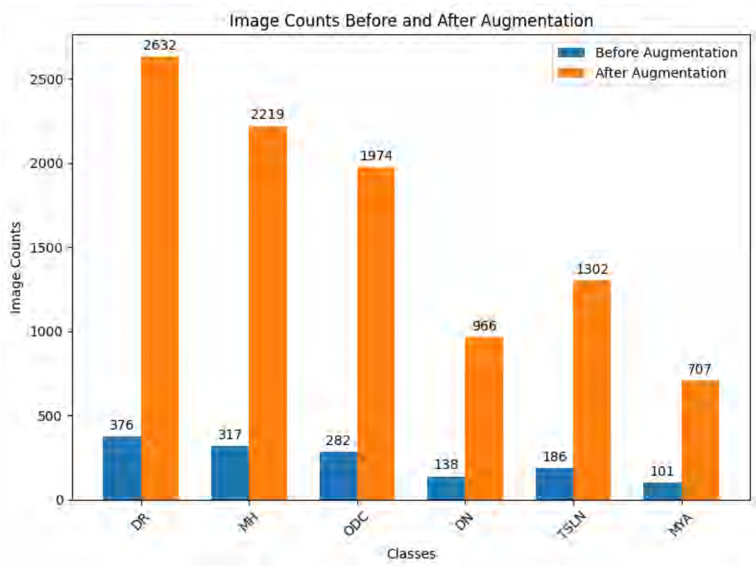


Figure 4.5: Visual before and after augmentation

Chapter 5

Implementation

5.1 Proposed Algorithm

5.1.1 Few-Shot Approach

In the conducted study, the research focused on the application of meta-learning as a pivotal approach within few-shot learning to address the challenges associated with limited labeled data, particularly evident in domains such as the medical field in Bangladesh. Conventional machine learning algorithms often necessitate substantial labeled data for effective generalization to unseen examples, making few-shot learning techniques crucial in scenarios where data collection is impractical.

Meta-learning, also referred to as learning to learn, emerged as a widely adopted approach in few-shot learning. The methodology involved training a model on various tasks to enable it to swiftly adapt to new tasks with a limited number of examples. The research employed a support set and a query set for the training of the meta-learning model.

For example, the training process utilized a 3-way, 2-shot configuration. In the figure 5.1 term "3-way" denoted that the model was trained on three distinct classes, representing different diseases in this particular case. Meanwhile, "2-shot" indicated that only two labeled images per class were provided in the support set. The primary objective was not to acquaint the model with the specific characteristics of individual classes, such as the similarities within the ODC or TSLN classes, but rather to equip it with the ability to differentiate between classes.

The query set consisted of labeled images from the same classes as the support set, allowing the model to focus on learning discriminative features between classes rather than within a single class. During testing in figure 5.2, the model was presented with a support set comprising labeled images, successfully identifying the class of query images based on its acquired ability to discern features prioritized during the training phase. Notably, the model demonstrated its capacity to adapt to varying support sets, even when the classes differed from those encountered during the training period. This approach proves effective in scenarios where the model needs to distinguish between different classes despite changes in the composition of labeled data.

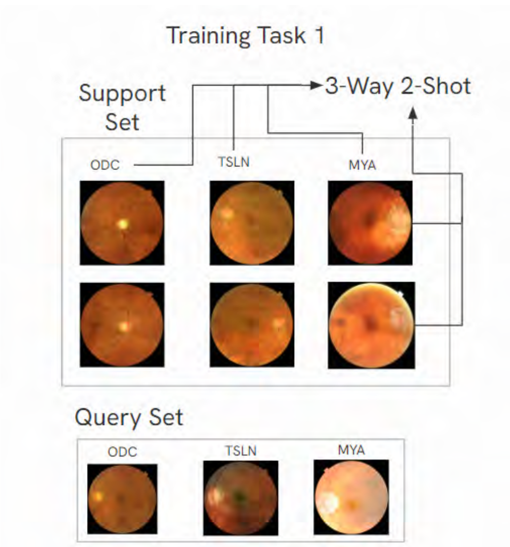


Figure 5.1: Fewshot Training

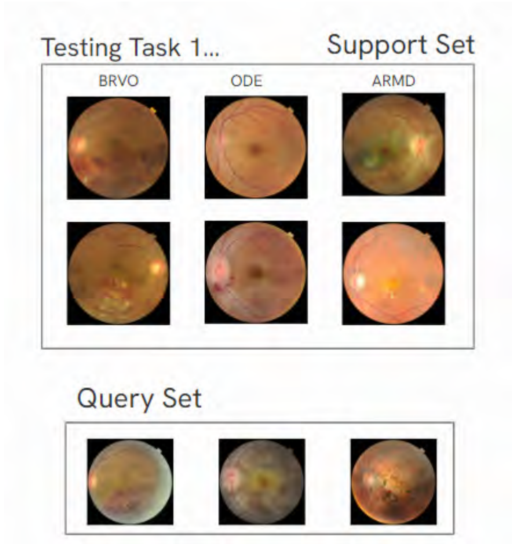


Figure 5.2: Fewshot Test

5.1.2 Prototypical Network

Prototypical networking is a concept of few-shot learning that includes prototype-based representation[27]. It falls under the criteria of meta-learning[18]. In prototypical networking, each class is represented by a prototype similar to the the figure in 5.3 that is usually the mean of the feature vectors of its examples in the embedding space. Embedding space in machine learning dictates how to represent or map data into a new space that makes it easier to understand the relationships among the pieces of the dataset and facilitate them accordingly.

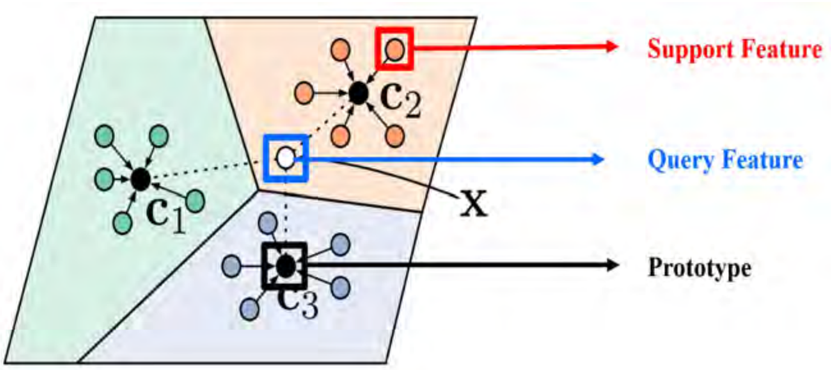


Figure 5.3: Prototypical Network

The concept of prototypical networking is based on the usage of prototypes or an M-dimensional representation $c_k \in R^M$. The prototype is computed through the mean of the feature vectors of its examples in the embedding function $f_\phi : R^D \rightarrow R^M$. It serves the purpose of an average representative in the training phase. This embedding function is learned in the training phase so that it can make accurate predictions, enabling better classification results with fewer examples per class[21].

$$c_k = \frac{1}{|S_k|} \sum_{(x_i,\, y_i) \in S_k} f_\phi\left(x_i\right) \tag{5.1}$$

After that, a distance metric called Euclidean distance is used in the embedding space to determine the similarity between the test example and the prototypes of different classes. The class that comes the closest is assigned for the prediction purpose. Prototypical networks are well fit for few-shot learning scenarios where each class does not have enough number of examples available for learning[21].

$$p_\phi(y = k \mid \mathbf{x}) = \frac{exp\left(-d\left(f_\phi\left(\mathbf{x}\right),\ c_k\right)\right)}{\Sigma_{k'}\ exp\left(-d\left(f_\phi\left(\mathbf{x}\right),\ c_{k'}\right)\right)} \tag{5.2}$$

The model gets trained to generalize well with better prediction to new classes with limited labeled data. In the phase of training, prototypical networks use episodic training so that it can simulate the few-shot learning method by building episodes or tasks. Each episode is combined of two parts called support set and query set. The support set is made of labeled examples for a few classes. These examples are the known ones. The query set is a combination of the unlabeled examples for the same classes[21].

$$S = \{(x_i, y_i)\}_{i=1}^{N \cdot K} \tag{5.3}$$

$$Q = \{(x_j, y_j)\}_{j=1}^{N \cdot Q} \tag{5.4}$$

Then classification happens based on the proximity of a query example to the prototypes of different classes in the embedding space. The class that gives the result closest to the prototype is predicted as the outcome. Probabilities are computed in the classification step to dictate the probability of a query example belonging to each class. Classification uses Euclidean distance for calculating the likelihood[12].

$$p(y = k|x) = \frac{e^{-d(\phi(x), c_k)}}{\sum_{j=1}^{N} e^{-d(\phi(x), c_j)}} \tag{5.5}$$

After that, the cross entropy loss calculation is done based on the model's predictions and how well it matches the true labels of the query set[12].

$$\text{Average Cross Entropy Loss} = -\frac{1}{|S|} \sum_{(x,y)\in S} \log(p(y|x)) \tag{5.6}$$

This error calculation is used to update the parameters of the model in the training phase for better accuracy rate.

5.1.3 Experiment

For our experiment, we first used inceptionV3 on our dataset without few-shot learning. As we used transfer learning, the inceptionV3 model was pre-trained on an image net dataset. We unfreezed the last few layers as fine-tuning. Additionally, we added some dense layers with ReLU activation. Furthermore, we have used different hyperparameters such as beta 1, beta 2, and epsilon. After adding all these to our model we trained the model for the general 4-class classification task on our dataset. The training accuracy was around 68% but the test accuracy was around 60%. As inceptionV3 was not performing well with our dataset, we did not proceed to train it for our few-shot learning model. We used VGG19 and ResNet50 for feature extraction and used the feature vectors in the prototypical network similar to the work done in this paper[30]. Among these two, VGG19 showed significantly better results. The model was trained on six classes and those were DN, DR, ODC, TLSN, MYA, and MH. In each training episode, a support set was created from training data and a query set was created from the validation data. The hyperparameters were set using trial and error methods.

Adam optimizer was used with a learning rate of 0.0001. Higher than that caused the model to overfit. For ResNet50, we used learning rate of 0.001. Changing this value caused the loss to increase. Beta1 and Beta2 parameters were set to 0.9 and 0.99 consecutively. L2 regularization was implemented while training to reduce overfitting. The regularization with the coefficient of 0.001 was incorporated into the loss function for better generalization. Here, we did a 2-way 10-shot classification which showed the best results. We also experimented with 3-way classification using a different number of shots. We also tried variations of training episodes to get the optimal result. The model performed better when in each epoch 100 training tasks with 5 query images were given. In each training task, 2 classes were taken randomly with 10 support images each. The V100 GPU from Google Colab Pro was used to experiment. Both the ResNet50 and VGG19 were run for 100 epochs which took around 2 hours each. For the test, both the support and query images were taken from the test data. The model was tested on 4 unseen classes. Those were ARMD, BRVO, ODE, and ODP. The number of test episodes was 3000.

Chapter 6

Result Analysis

6.1 Overview of Results

In our paper, we have employed VGG19, ResNet50, and Inception v3 to classify four classes: ARMD, BRVO, ODE, and ODP. Among these mentioned models, VGG19 has performed better than the rest with a test accuracy of 81.64%. The dataset classes were imbalanced which led to a slightly overfitting model.

Our proposed model has fewer parameters and it works faster in terms of training compared to other approaches. The training showed satisfactory results in the table6.1. The Few-Shot learning works on n-way k-shot. Here we tried 2-way and 3-way with 5 and 10 support images. The 2-way led to much better results compared to the 3-way classification which can be seen in table 6.2. It was easier for the model to predict when fewer options were given. We also noticed that increasing the support set made it easier for the model to give more accurate predictions. There was a significant improvement in test results when the support images were increased from 5 images to 10 images. In each epoch, we used 100 training tasks, and increasing this number caused the model to under-fit. Lowering this number reduced accuracy. If we look at the graphs, the training and validation accuracy curves are very close. The same thing can be seen for the losses. The curves do not fluctuate excessively. This means that the test results will also be similar to the training results. The classification report for VGG19 in figure 6.3 shows how well the model performs.

Pre-Trained Model	Average Training Accuracy	Average Training Loss	Average Validation Accuracy	Average Validation Loss
VGG19	83.96%	0.363	83.77%	0.371
ResNet50	80.65%	1.639	82.20%	0.467
InceptionV3	68.58%	0.954	62.35%	1.24

Table 6.1: Average Training & Validation Accuracy & Loss

Pre-Trained Model	2-Way 10-Shot	2-Way 5-Shot	3-Way 10-Shot	3-Way 5-Shot
VGG19	81.66%	77.56%	74.77%	68.58%
ResNet-50	76.67%	71.46%	65.98%	60.08%

Table 6.2: Test Accuracy

	Precision	Recall	F1 Score
Class1	81%	80%	81%
Class2	80%	81%	81%
Accuracy		81%	

Table 6.3: VGG19 Classification Report

The 2-way 10-shot on VGG19 achieved the highest test accuracy of 81.66% (6.5). The average training and validation accuracy was 83.96% and 83.77% consecutively shown in figure 6.2. The losses for the training and validation were 0.363 and 0.371 (6.1).

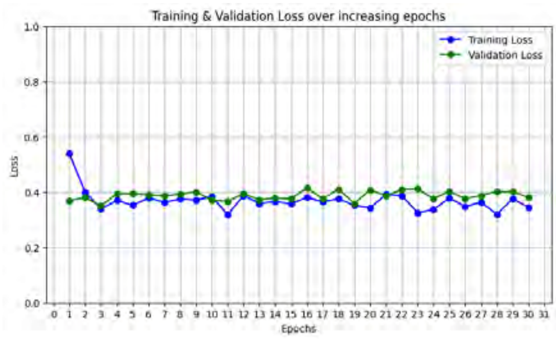


Figure 6.1: VGG19 Training Loss

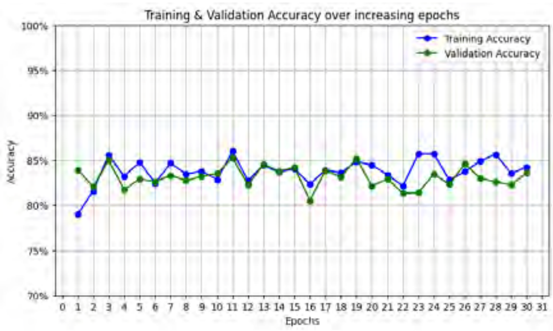


Figure 6.2: VGG19 Training Accuracy

Although the training and validation accuracy of the ResNet50 was promising, the loss count was high. The average training accuracy was 80.65% and the average validation accuracy was 82.20.%. The model was under-fitting as per the figure 6.3 and 6.4. The average test accuracy for ResNet50 was 76.67% with an average loss of 0.570. For ResNet50, precision, recall, and f1 score was 77%.

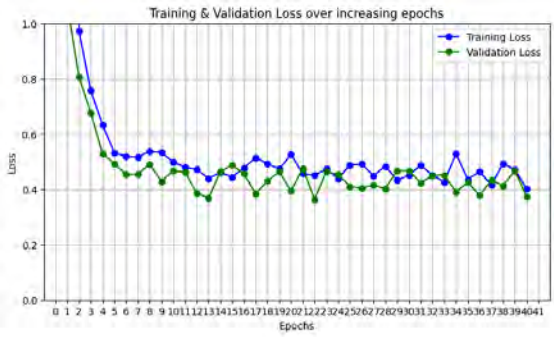


Figure 6.3: ResNet50 Training Loss

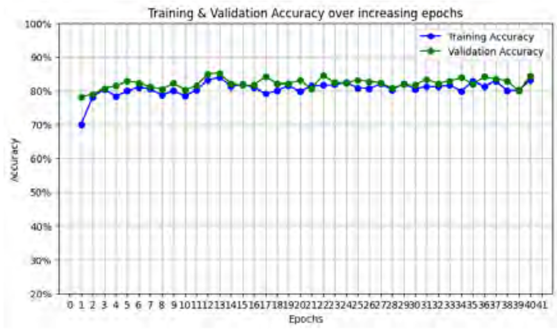


Figure 6.4: ResNet50 Training Accuracy

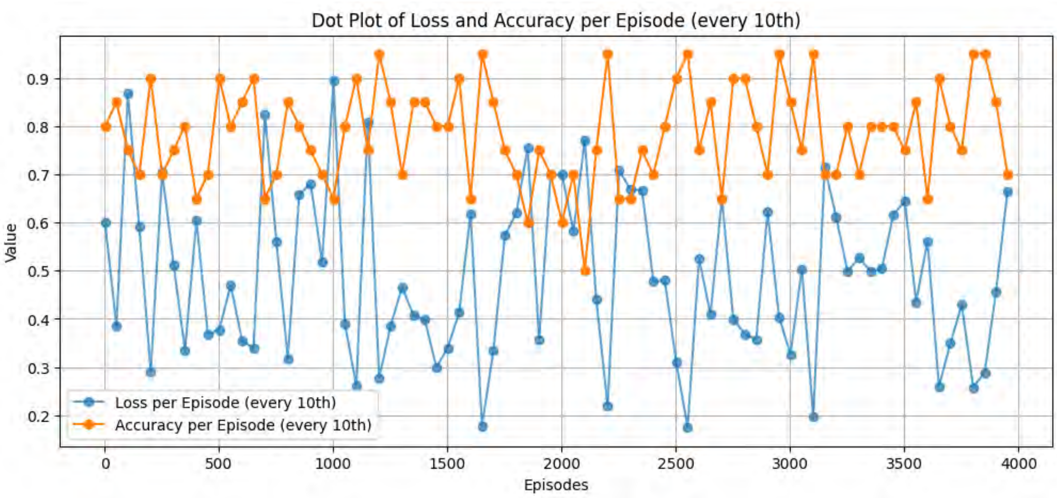


Figure 6.5: Dot Graph of Test Loss and Accuracy of VGG19

From the 4 classes of the test set, when we used a 2-way 10-shot, the confusion matrix shows the efficiency of the model by predicting a large number of true positive and false negative classes properly. For further details, the ROC curve is shown in Figure 6.6 and 6.7 along with the confusion matrix in figure 6.8 and 6.9.

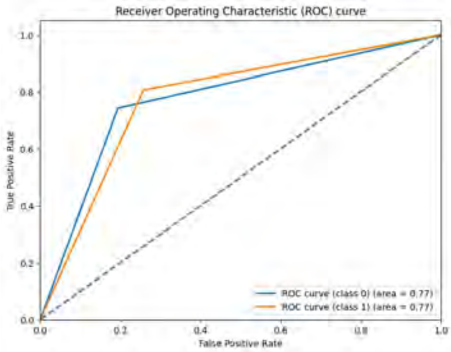


Figure 6.6: VGG19 ROC Curve

Figure 6.7: VGG19 Confusion Matrix

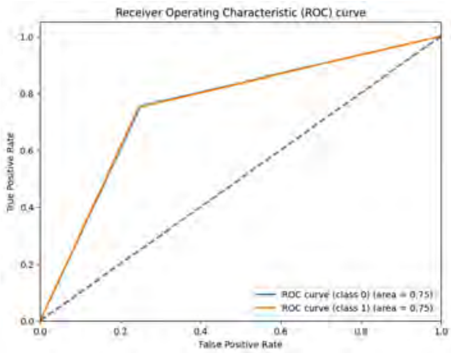


Figure 6.8: ResNet50 ROC Curve

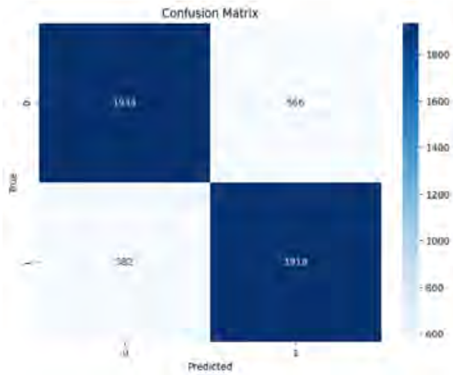


Figure 6.9: ResNet50 Confusion Matrix

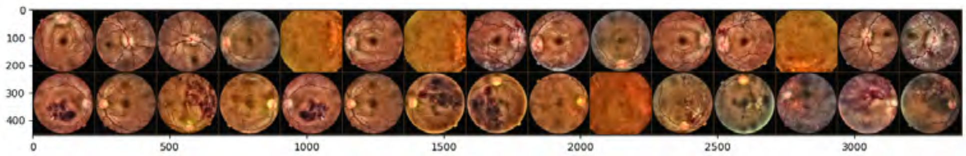


Figure 6.10: Prediction Example

Class labels sampled: ['ODE' 'BRVO'] Loss: 0.364, Accuracy: 0.80, y hat: tensor([[0, 0, 0, 0, 0], [1, 0, 1, 1, 0]]). This is the predicted result shown for the figure 6.10. This was predicted on 2-way 10-shot classification with VGG-19 using 5 query images from each class. The two classes were ODE and BRVO. The ODE was labeled as 0 and BRVO as 1. The predicted labels showed that all 5 images from the ODE class were predicted correctly as all of them had predicted labels of 0. Among the 5 images from the BRVO class, 3 were predicted correctly with the predicted label of 1. Out of the 10 images, 8 were correctly predicted resulting in an 80% accuracy.

Chapter 7

Conclusion And Future Work

7.1 Final Discussion

Every year the number of retinal diseases is increasing rapidly. As a result, AI has been working to develop multiple automated models that can detect retinal diseases efficiently. As the dataset in this regard is mostly hard to find, it has become a challenge to acquire accurate results in the initial phase of retinal disease detection. In the conducted study, diverse models such as VGG19, InceptionV3, and ResNet50 have been used. Subsequently, for enhancing the dataset we intend to use unique image enhancement techniques and multiple machine learning algorithms for better success rate. In this research, image classification has been successfully incorporated with Few-Shot learning to build up an automated model for retinal disease classification purposes. Moreover, the prototypical network has been used to train our model for the classification and it has given a promising accuracy rate. Image classification on unseen classes has been done with a limited amount of data. A noteworthy fact about this research work is that fewer parameters and lower computational costs led the model to be faster in training. Soon, we intend to enhance the performance of our model by exploring new algorithms and techniques.

7.2 Future Work

In our paper, we have addressed one approach to retinal disease detection. For future improvement, we can employ multiple approaches to enhance accuracy and efficiency. Customizing the layers can have a significant impact on the output for the feature extractor model. Adding or removing layers can help detect complex patterns and features. Furthermore, ensemble learning can be used to create better and more accurate predictions. Different models have different features and areas of strength. The idea is to combine them to get the best result. On the contrary, MobileNet can show far better results and uses separable convolutional layers to reduce the parameters leading to lesser computational complexity. MobileNet is an efficient neural network architecture used when there is a lack of resources. As the goal of the few-shot learning technique is to discriminate or find similarities between classes, increasing the number of classes during training can be beneficial for the model to predict more accurately during testing. The only limitation we had in the dataset was that we had examples of a single image containing multiple diseases. If a query image has multiple diseases from the support set classes, the model may give the wrong prediction as the prototypical network returns a single predicted label for each image.

Bibliography

[1] Ali Mohammad Alqudah. “AOCT-NET: a convolutional network automated classification of multiclass retinal diseases using spectral-domain optical coherence tomography images”. In: *Medical & biological engineering & computing* 58 (2020), pp. 41–53.

[2] Monika Bansal et al. “Transfer learning for image classification using VGG19: Caltech-101 image data set”. In: *Journal of ambient intelligence and humanized computing* (2021), pp. 1–12.

[3] Philippe Burlina et al. “Low-shot deep learning of diabetic retinopathy with potential applications to address artificial intelligence bias in retinal diagnostics and rare ophthalmic diseases”. In: *JAMA ophthalmology* 138.10 (2020), pp. 1070–1077.

[4] A.S. Choudhary. *Meta-Learning: Structure, Advantages & Examples*. 2022. URL: https://www.analyticsvidhya.com/blog/2022/09/meta-learning-structure-advantages-examples/.

[5] Malena Daich Varela et al. “Artificial intelligence in retinal disease: clinical application, challenges, and future directions”. In: *Graefe’s Archive for Clinical and Experimental Ophthalmology* (2023), pp. 1–15.

[6] Nasr Gharaibeh et al. *An effective image processing method for detection of diabetic retinopathy diseases from retinal fundus images*. 2018.

[7] Anik Ghosh, ABM Aowlad Hossain, and SM Taslim Uddin Raju. “Classification of diabetic retinopathy using few-shot transfer learning from imbalanced

data". In: *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*. Vol. 1. IEEE. 2021, pp. 78–83.

[8] Z-K Han et al. *A few-shot learning-based eye diseases screening method.* 2022.

[9] Kaiming He et al. "Deep Residual Learning for Image Recognition". In: (2016), pp. 770–778. DOI: 10.1109/CVPR.2016.90.

[10] D. Jude Hemanth, Omer Deperlioglu, and Utku Kose. "An enhanced diabetic retinopathy detection and classification approach using deep convolutional neural network". In: *Neural Computing and Applications* (Jan. 2019). DOI: 10.1007/s00521-018-03974-0.

[11] Mohammad Hossin and Md Nasir Sulaiman. "A review on evaluation metrics for data classification evaluations". In: *International journal of data mining & knowledge management process* 5.2 (2015), p. 1.

[12] Steinar Laenen and Luca Bertinetto. "On episodes, prototypical networks, and few-shot learning". In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 24581–24592.

[13] Feng Li et al. "Deep learning-based automated detection of retinal diseases using optical coherence tomography images". In: *Biomedical optics express* 10.12 (2019), pp. 6204–6226.

[14] Feng Li et al. "Fully automated detection of retinal disorders by image-based deep learning". In: *Graefe's Archive for Clinical and Experimental Ophthalmology* 257 (2019), pp. 495–505.

[15] Xiang Li et al. "Lightweight attention convolutional neural network for retinal vessel image segmentation". In: *IEEE Transactions on Industrial Informatics* 17.3 (2020), pp. 1958–1967.

[16] Vladimir Lyashenko. *Understanding Few-Shot Learning in Computer Vision: What You Need to Know.* neptune.ai, Dec. 2020. URL: https://neptune.ai/blog/understanding-few-shot-learning-in-computer-vision.

[17] Siwei Mai et al. "Few-shot transfer learning for hereditary retinal diseases recognition". In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24*. Springer. 2021, pp. 97–107.

[18] M Murugappan et al. "A novel few-shot classification framework for diabetic retinopathy detection and grading". In: *Measurement* 200 (2022), p. 111485.

[19] Raffaele Nuzzi et al. *The impact of artificial intelligence and deep learning in eye diseases: a review*. 2021.

[20] Samiksha Pachade et al. "Retinal Fundus Multi-disease Image Dataset (RFMiD)". In: (2020). DOI: 10.21227/s3g7-st65. URL: https://dx.doi.org/10.21227/s3g7-st65.

[21] Archit Parnami and Minwoo Lee. *Learning from Few Examples: A Summary of Approaches to Few-Shot Learning*. 2022. arXiv: 2203.04291 [cs.LG].

[22] Peng Peng and Jiugen Wang. "How to fine-tune deep neural networks in few-shot learning?" In: *arXiv preprint arXiv:2012.00204* (2020).

[23] Juan-Manuel Perez-Rua et al. "Incremental few-shot object detection". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, pp. 13846–13855.

[24] Tanzila Saba et al. "Automatic detection of papilledema through fundus retinal images using deep learning". In: *Microscopy Research and Technique* (July 2021). DOI: 10.1002/jemt.23865.

[25] Rubina Sarki et al. "Automated detection of mild and multi-class diabetic eye diseases using deep learning". In: *Health Information Science and Systems* 8.1 (2020), p. 32.

[26] Agung Setiawan et al. "Color retinal image enhancement using CLAHE". In: June 2013, pp. 1–3. ISBN: 978-1-4799-0143-2. DOI: 10.1109/ICTSS.2013.6588092.

[27] Jake Snell, Kevin Swersky, and Richard S. Zemel. *Prototypical Networks for Few-shot Learning*. 2017. arXiv: 1703.05175 `[cs.LG]`.

[28] Malliga Subramanian et al. "Diagnosis of retinal diseases based on Bayesian optimization deep learning network using optical coherence tomography images". In: *Computational Intelligence and Neuroscience* 2022 (2022).

[29] Christian Szegedy et al. "Rethinking the Inception Architecture for Computer Vision". In: (2016), pp. 2818–2826. DOI: 10.1109/CVPR.2016.308.

[30] Rodney Tsiko. *Eye Disease Classification - Few-Shot Learning*. https://github.com/cv21rgt/Eye-Disease-Classification-Few-Shot-Learning. 2022.

[31] Lilit Yolyan. *Review on Few Shot Object Detection*. Medium, Dec. 2021. URL: https://towardsdatascience.com/review-on-few-shot-object-detection-185965e0e6a2.

[32] Tae Keun Yoo, Joon Yul Choi, and Hong Kyu Kim. "Feasibility study to improve deep learning in OCT diagnosis of rare retinal diseases with few-shot classification". In: *Medical & Biological Engineering & Computing* 59 (2021), pp. 401–415.