Revolutionizing Ocular Diagnosis: An Enhanced CNN based approach for Multi-class Classification of Optical Coherence Tomography Images

Anas Khalid Awan

Registration No: 178-FET/MSEE/F22

Supervisor:

Dr. Zeeshan Aslam Khan

Co Supervisor:

Dr. Khizer Mehmood

Department of Electrical and Computer Engineering

Faculty of Engineering and Technology

International Islamic University Islamabad

2024

DISSERTATION

A dissertation submitted to the Department of Electrical Engineering, International Islamic University Islamabad as a partial fulfillment of the requirements for the award of the degree.

Department of Electrical and Computer Engineering

Faculty of Engineering and Technology

International Islamic University Islamabad

2024

DEDICATION

I dedicate research work to my beloved parents, respected teachers, siblings and all those who prayed for my success.

"Alhamdulillah for everything, we can never thank Allah enough for the countless bounties He blessed us with"

CERTIFICATE OF APPROVAL

Title of Thesis: Revolutionizing Ocular Diagnosis: An Enhanced CNN based approach for Multi-class Classification of Optical Coherence Tomography Images

Name of Student: Anas Khalid Awan

Registration No: 178/MSEE/FET/F22

Accepted by the Faculty of Engineering and Technology, International Islamic University Islamabad in partial fulfillment of the requirement of the MS Degree in Electrical Engineering.

Viva voice committee:

Dean FET, IIU Islamabad

Assistant Professor, Chairman DEE, FET, IIUI

External Examiner

Internal Examiner

Supervisor:

Dr. Zeshan Aslam Khan

DECLARATION

I certify that research work titled “Revolutionizing Ocular Diagnosis: An Enhanced CNN based approach for Multi-class Classification of Optical Coherence Tomography Images” has been completed by me and it has not been done before and presented anywhere for evaluation. Furthermore, I have properly acknowledged the material taken from related sources.

ACKNOWLEDGEMENT

First of all, I would like to thank ALLAH (SWT) for giving me the great family, supportive teachers and co-operative friends. I couldn't finish my Research-work without His help.

This work is completed with the help of two people. I would like to express my gratitude to my supervisor Dr. Zeshan Aslam Khan for his continuous support, motivation, strength and valuable guidance. I would also pay my sincere gratitude to my co-supervisor Dr. Khizer Mehmood for his sincere devotion, valuable time, suggestions throughout the work.

In the last, I am very thankful to my beloved parents for their continuous support, unconditional love. They have always given me the moral and spiritual support and motivation to achieve my goals.

ABSTRACT

Image classification of medical images for diagnostic purposes plays a significant role in early and accurate disease detection in various scenarios such as brain tumor classification, kidney stones detection, breast cancer detection, colorectal cancer classification, detection of certain lungs conditions such as pneumonia, lung cancer, COVID-19 and post-COVID-19 pneumonia etc., detection and classification of skin cancer and various other ailments. In the realm of ophthalmic health, conventionally, the surgeon or physician diagnoses the condition by analyzing the optical coherence tomography scans (Retinal OCT Scans). To automate and revolutionize the diagnostic process, an AI based efficient, fast, accurate, robust, reliable and interpretable system is required to treat the patients with due care. To address the challenges related to the AI based ophthalmic diagnosis tools, researchers have proposed various CNN based solutions in which various trade-offs are observed such as architectural complexity for improved accuracy, compromised feature extraction mechanisms and neglecting the integration of explainability techniques to limit memory usage and lack of various benchmark evaluation metrics that give information about various aspects of model's performance. In this research work, two architectures of CNN are designed, beginning with an explainable "DA-CNN" model, having enhanced feature extraction mechanism that is based on latest proposed techniques like separable dynamic convolution and channel split dual attention. The second "CF-CNN" model is designed using dynamic convolution and static convolution attention and it is also integrated with explainability techniques. The proposed DA-CNN and CF-CNN models classify the OCT scan images in four separate categories i.e., Normal, DME, Drusen and CNV, accurately through enhanced feature extraction mechanisms, quickly through latest optimizers and

also provide the explanations for their decisions by highlighting the prominent features that were crucial in the decision-making process through explainable artificial intelligence (XAI) techniques like GradCam and LIME. The suggested model architectures surpass various standard competitors regarding accuracy, M.A.E., dice-coefficient, AUC-ROC, sensitivity and specificity on benchmark OCT Kermany 2018 dataset. The proposed models achieve the optimal accuracy of 97.4 percent for 'DA-CNN' and 96.68 percent for 'CF-CNN' in relatively lesser number of iterations as compared to standard counterparts. Moreover, the 'DA-CNN' model is compiled by exploiting four different optimizers i.e., Adam, RMSprop, Nadam and Adafactor at four different instances. The case study of various evaluation metrics based on these optimizers is also presented in this thesis.

List of Figures

List of Tables

Table of Contents

CHAPTER 1

INTRODUCTION

1.1. Introduction

This chapter provides an overview of applications of Machine Learning and Artificial Intelligence techniques in the field of healthcare, highlighting their inevitability, implications and practical applications. Moreover, this chapter also elaborates how different CNN based image classifiers are revolutionizing the diagnostic systems for various ailments such as brain tumor classification, kidney stones detection, breast cancer detection, colorectal cancer classification, detection of certain lungs conditions such as pneumonia, lung cancer, COVID-19 and post COVID-19 pneumonia etc., detection and classification of skin cancer and various other conditions. Moreover, it presents the background and motivation behind the research work which includes the basic information of retinal conditions in the scope of the study and some statistics about the affecters of the three ocular conditions i.e. Choroidal Neovascularization, Diabetic Macular Edema and Drusen that are in the scope of this research work.

1.2. Inspiration and Background

The biomedical sciences have seen a computer science revolution in the last few decades, particularly in the areas of diagnosis and disease detection[1]. Artificial intelligence (AI) has completely changed the process of diagnosing diseases and anatomizing bodies by doing classification tasks that previously required a great deal of human effort. The medical industry is eager to embrace artificial intelligence approaches because of the recent widespread expansion in the usage of AI-based solutions, which function with very few errors, almost no accidents, and very few misdiagnosis [2]. The application areas of artificial intelligence in healthcare are shown in Figure 1.

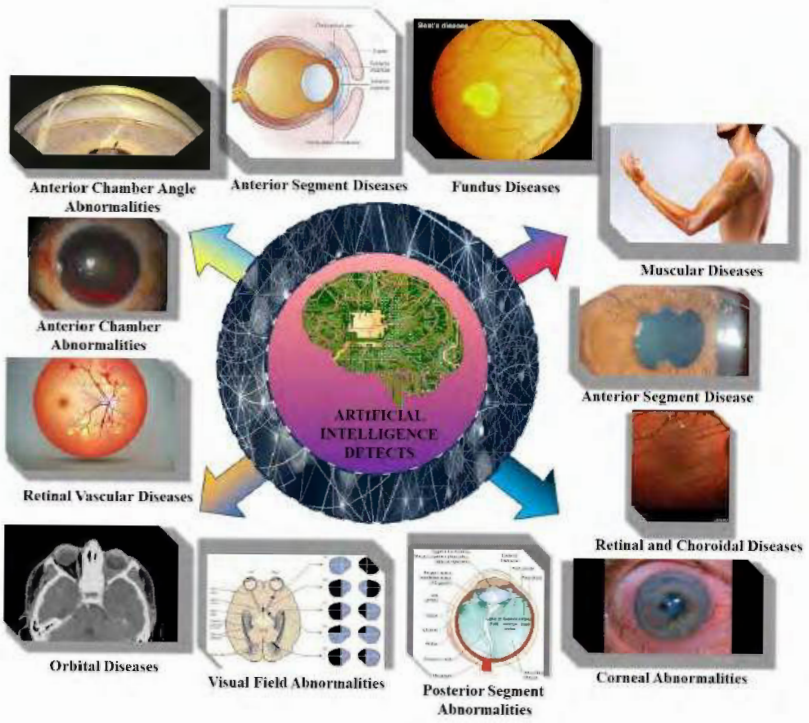


Figure 1: Applications of AI in healthcare

Using medical images to diagnose the condition, which if done manually can lead to human error, several AI and deep learning networks and models are helpful in the diagnosis and prognosis of numerous diseases, including brain tumors, ocular conditions, kidney stone detection, lung cancer, and breast cancer[3]. Deep learning techniques are utilized to interpret medical images that are obtained via various techniques such as computed tomography scans, X-Ray scans, magnetic rays imaging scans and optical coherence tomography scans. for the purpose of performing various tasks, such as segmentation, classification, and prediction, all of which can be completed accurately and without human intervention.

In the rapidly progressing world of medical diagnostics particularly medical imaging, exploiting latest technology is important to make diagnosis better and quicker [4] As far as eye care is concerned, where being accurate is super important for efficient patient care, an optical coherence tomography (OCT) method has been developed [5].This tool helps doctors see structural details better than ever but various eye conditions are very complex, making it hard to identify them quickly and correctly [2].

The identification and supervision of retinal illnesses are significantly impacted by

artificial intelligence. The technique makes it easier for ophthalmologists to focus on patient care by demanding precise and accurate ocular layer identification and extraction. In this work, the benefits of AI have been utilized to identify and classify an eye condition; however, because of the retina's intricate anatomy, it takes a while for the specialist to appropriately assess the state of affairs[6]. The retina is responsible for providing the brain with images and light and is located inside the posterior wall of the eye.

When light concentrates on the retina as opposed to some other place, Normal vision is noted. Those with normal vision are able to see objects both up close and far away. The retinal layers that are affected by some condition ultimately lead to macular degeneration, myopia and loss of vision. Eye conditions play pivotal roles in the field of ophthalmic health, presenting unique challenges and consequences for vision [7] Conditions like Choroidal Neovascularization (CNV) [8], Diabetic Macular Edema (DME) [9], and Drusen [10] stand out due to their distinct characteristics and impacts on visual well-being.

In the case of CNV as shown in Figure 2, abnormal blood vessels growing beneath the retina pose a significant threat, often resulting in leakage and harm to nearby tissues. Indicators of CNV encompass vision distortion, blurred or wavy vision, and, in severe

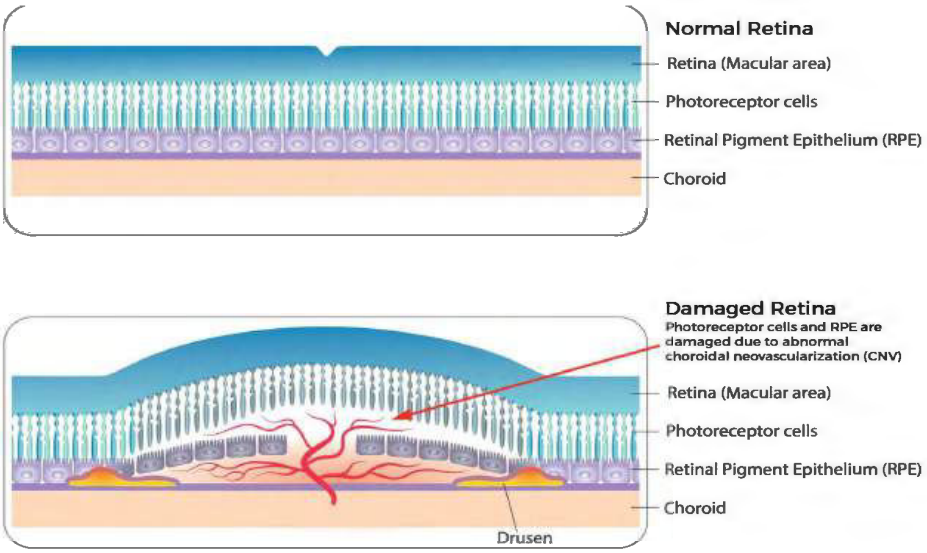


Figure 2: Pictorial representation of Choroidal Neovascularization condition[11]

instances, an abrupt loss of eyesight in the center. While Age-related Macular Degeneration (AMD) is a common culprit, CNV can also stem from various other retinal disorders [8]. DME, as shown in Figure 3, a complication of diabetic retinopathy, involves the buildup of fluid in the macula, the area of the retina in the center that is essential for detailed vision. Manifestations include blurred or distorted central vision, making tasks like reading, facial recognition, and detailed activities challenging.

Long-term diabetes, particularly when not effectively managed, heightens the risk of developing DME [9]. Drusen, little yellow or white spots under the retina, as shown in Figure 4, often go with AMD. This is common in people over 60 years old and can be linked to Age-related macular degeneration that is an eye condition (AMD). On the other hand, soft drusen linked to AMD can cause you to lose your vision. The role of drusen in AMD progression involves the degradation of the macula's light-sensitive cells [10]. The OCT scans of all three conditions i.e. CNV, DME and Drusen in comparison with the normal retinal scan is shown in Figure 5.

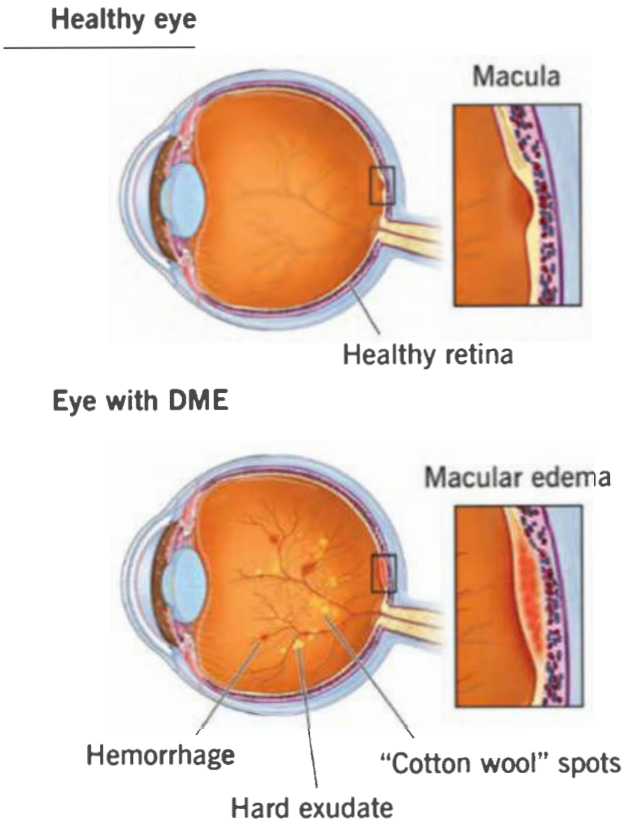


Figure 3: Pictorial representation of Diabetic Macular Edema condition[12]

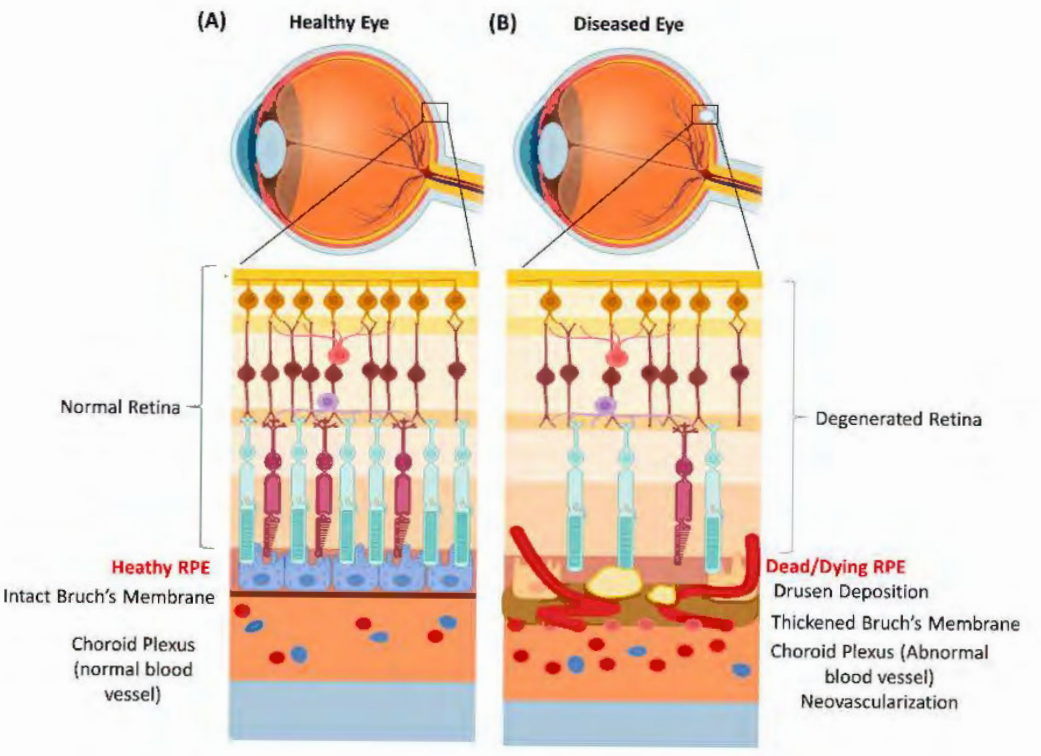


Figure 4: Pictorial Representation of Drusen condition[13]

In persons between the ages of 43 and 86, the prevalence of CNV linked to age-related macular degeneration (ARMD) was 1.2% in the Wisconsin Beaver Dam Study[14]. In the US and Europe, myopia is the second most typical cause of CNV. It is expected that 5–10% of myopes experience CNV, with 60–75 percent of them being sub foveal. In 0.1 percent of endemic area residents, disciform scars secondary to CNV from suspected ocular histoplasmosis syndrome (POHS) were seen.

The formation of CNV is uncommon in multiple evanescent white dot syndrome (MEWDS). Estimates of CNV in multifocal choroiditis patients range from 25 to 40 percent. Thirty-three percent of patients with punctate inner choroidopathy (PIC) experience CNV. Fifty percent of individuals have sub foveal acuities, meaning their visual acuities range from 20/80 to 20/200.

Five percent of patients with birdshot chorioretinopathy experience CNV. Almost all choroidal ruptures experience CNV during the healing phase, with the majority involute spontaneously. 15% to 30% of patients may experience a recurrence of CNV, which could result in a serious or hemorrhagic retinal detachment and accompanying vision loss. One estimate estimates the prevalence of optic nerve drusen

between 1 and 2 percent in the US population. People of all sexes are equally affected by them. The percentage is estimated to be between 1.8 and 2.4 percent by another source. Compared to other racial groups, white persons are more likely to develop optic disc drusen. Thirty-three percent of patients with punctate inner choroidopathy (PIC) experience CNV.

Out of the 1038 individuals over 40 who were diagnosed with diabetes mellitus and examined for this research, 55 of them had DME, resulting in a weighted prevalence of 3.8 percent (95% confidence interval: 2.7 to 4.9 percent) or around 746,000 people in the US population aged 40 or above in 2010. Age or gender did not appear to have any bearing on the occurrence of DME. In comparison to non-Hispanic Whites, non-Hispanic Blacks were more likely to have DME using multivariable logistic regression (OR 2.64; 95 percent CI, 1.19–5.84; P=.02). DME prevalence was similarly correlated with higher hemoglobin A1c levels (OR 1.47; 95 percent CI, 1.26–1.71 for each 1 percent; P<.0001) and longer diabetes duration (OR 8.51; 95 percent CI, 3.70–19.54 for ≥10 vs <10 years; P<.0001).

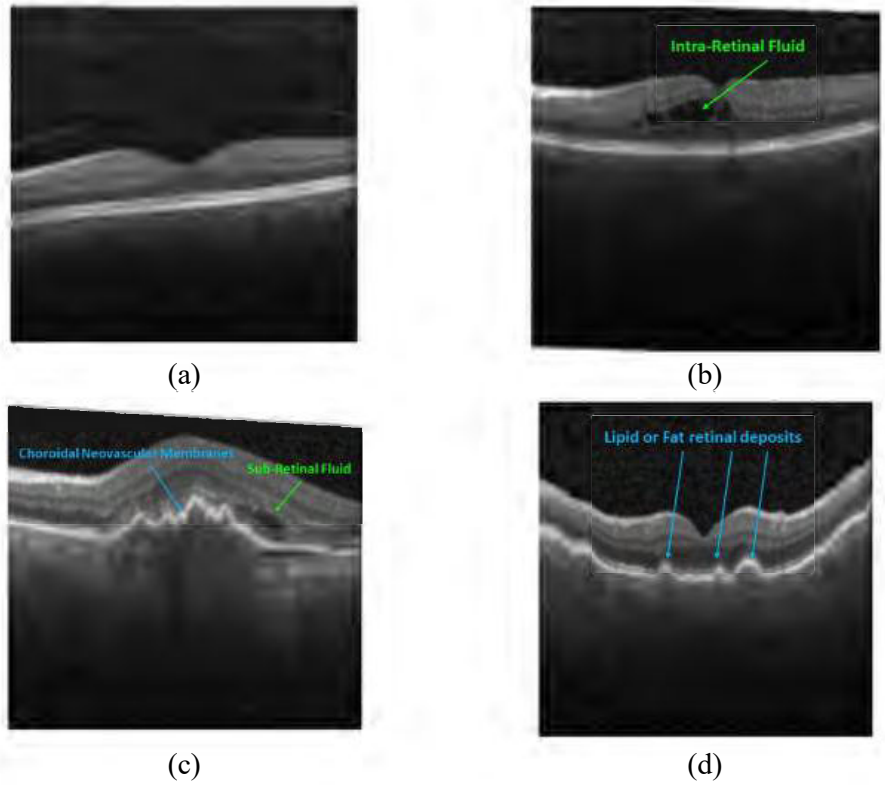


Figure 5: OCT Scans of Ocular Ailments (a) Normal (b) DME (c) CNV (d) Drusen

1.3. Problem Statement

The classification of ocular diseases has benefited substantially from advancements in deep learning, as evidenced by the popularity of numerous convolutional neural network (CNN)-based architectures, such as Google's Inception models (V1 and V3), ResNet-50, AlexNet, and VGG-16, which have become popular benchmark models. These architectures are useful, but they still have a lot of serious problems.

First, the lack of interpretability of many of the present CNN models makes it hard for practitioners to understand and trust the predictions they provide. Slow classification speed is another problem that prevents timely diagnosis and treatment of eye diseases in clinical settings. Moreover, the complex topologies of these models may result in increased computation costs and need high-performance technology, both of which are often available in healthcare environments. Low classification accuracy is another ongoing issue that compromises these models' trustworthiness in real-world scenarios. The limitations such as low accuracy, slow classification speed, lack of model's interpretability and architectural complexity will be addressed in this research work through enhanced CNN models.

1.4. Goals and Objectives

The prime objectives of the research work are stated as follows:

- Design and implementation of two enhanced CNN architectures to accurately classify the OCT images in 4 distinct labels i.e., DME, Normal, Drusen, and CNV.
- Integration of latest proposed features such as dynamic convolution, separable dynamic convolution, convolutional attention and channel split dual attention mechanisms for enhanced feature extraction and dimensionality reduction process, hence enabling the model to learn the essential features effectively.
- Focus on enhanced accuracy along with classification speed, hence avoiding the tradeoffs that is commonly observed in complex architectures.

- Integration of model's interpretability through Explainable Artificial Intelligence techniques such as Grad Cam and LIME, hence providing transparency in decisions and overcoming the black-box nature of CNN models.
- Training and evaluating the model that surpasses its standard counterparts in terms of accuracy, sensitivity, specificity, dice-coefficient, AUC-ROC, mean absolute error and other benchmark evaluation metrics.

1.5. Contributions

1. Design and implementation of novel CNN architectures to accurately classify the OCT images in 4 distinct labels i.e., DME, Normal, Drusen, and CNV.
2. Integration of latest proposed features such as separable dynamic convolution and channel split dual attention for enhanced feature extraction and dimensionality reduction process, hence enabling the model to learn the essential features effectively.
3. Focus on enhanced accuracy along with classification speed, hence avoiding the tradeoffs that is commonly observed in complex architectures.
4. The interpretability of the model is integrated through Explainable Artificial Intelligence technique Grad Cam, hence providing transparency in decisions and overcoming the black-box nature of CNN models.
5. Training and evaluating the model that surpasses its standard counterparts in terms of accuracy, sensitivity, specificity, dice-coefficient, AUC-ROC, mean absolute error and other benchmark evaluation metrics.
6. Achieved optimal accuracy of 97.42% in relatively much lesser number of iterations hence showcasing the efficiency of the proposed DA-CNN architecture and 96.68% for CF-CNN architecture.

1.6. Thesis Organization

The chapter-wise organization of the research work is presented below.

Chapter 1: gives a conceptual summary of the entire thesis, including research gaps, statements, and definitions that explicitly outline the objectives of the study, as well as the background and reasons for the identification of significant issues and the formulation of the research topic.

Chapter 2: discusses the benefits and drawbacks of previously proposed techniques in the literature to give a detailed overview of the work completed thus far.

Chapter 3: describes the research methodology with proposed models by elaborating the architectures, simulation environment, dataset description, data preprocessing and optimization algorithms employed in this research work.

Chapter 4: includes hyper-parameters selection details. Moreover, it provides complete simulation results in terms of tables and learning curves for a detailed comparison of the proposed models along with the case studies on different optimization techniques. This chapter also includes the details on integrating explainable artificial intelligence techniques with the proposed CNN models.

Chapter 5: presents the discussion on the results along with comparison with state-of-the-art models and conclusions drawn from the research work along with future research directions for the possible extension of a current study.

1.7. Summary

This chapter introduces the scope of the study mentioning the advancements and exploitation of AI based diagnostic systems in medicine and healthcare particularly, classification of medical images through deep learning architectures followed by inspiration and background, the goals and objectives of the research, the problem statement, research contributions and ending on the chapter-wise organization of the thesis.

CHAPTER 2

LITERATURE REVIEW

2.1. Introduction

Medical imaging methods have completely changed the healthcare industry by enabling specialists to explore deeper within the human body and try to cure ailments in ways never possible before which are shown in Figure 6. This is particularly significant when it comes to the identification of eye disorders, particularly in relation to medical scan analysis, which is critical to comprehending and treating various ocular illnesses. This has long been the responsibility of the licensed physicians who carefully review the scans and diagnose conditions pertaining to the eyes.

It is important to emphasize that in ophthalmology, accurate diagnosis is the cornerstone of treatment; without it, there could never be a successful therapy. However, it is a process that is fundamentally dependent on humans and may be limited by time constraints, weariness, and individual knowledge gaps. The prevalence of technology in today's world makes it imperative to look for a few ways to automate and streamline this diagnostic process in order to provide high-quality medical care quickly and accurately.

It has also been discovered that artificial intelligence (AI) is a highly powerful tool for solving these issues. The diagnosis of eye conditions can be altered by integrated AI, especially CNNs. Researchers need to strike a compromise between the three factors of accuracy, speed, and interpretability because several CNN designs are recommended to increase diagnostic accuracy. Because of their intricate designs, the majority of the current models slow down diagnostic processing, and interpretability based on XAI approaches is frequently disregarded.

In this regard, a novel CNN-based model is presented in a ground-breaking study that aims to simplify the multi-class classification of optical coherence tomography samples into four different classes: Drusen, Normal, CNV, and DME. The interpretability using XAI approaches is one of the main improvements proposed in this work. Unlike many

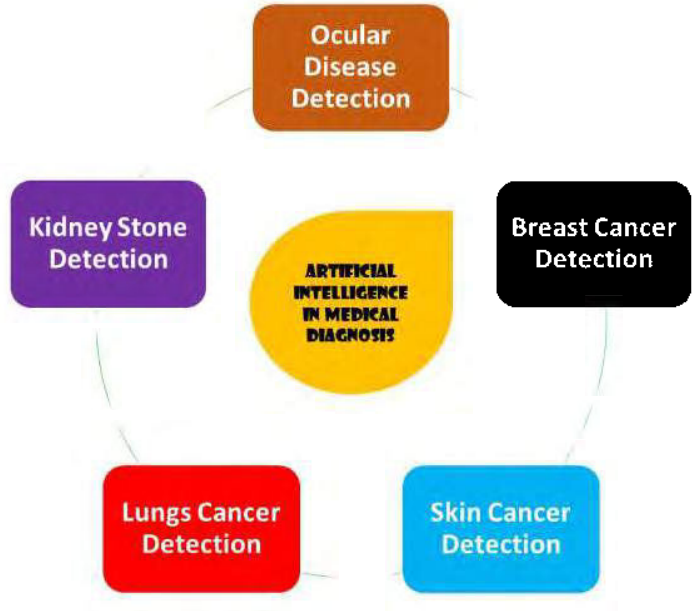


Figure 6: AI in medical imaging

other models that only place the onus of interpretation in a black box, our model goes one step further and provides justifications for classifications. The model can identify certain regions on the OCT images using XAI techniques, and these regions are in charge of the final classification. This openness fosters confidence in the model's judgments and grants medical practitioners a unique authority that they can verify and comprehend through the diagnostic results.

The application of AI techniques for image classification in medical diagnosis is one of the most significant advancements in medicine. Image pre-processing, which includes quantization, sampling, and segmentation; mass data training for neural networks; statistical analysis; and imaging processing are all necessary for the automated detection of retinal diseases. Presently, scientists primarily focus on the in-depth improvement of disease categorization and description linked to a decrease in time processing and memory needs. Additionally, they are capable of accurately segmenting ocular layers using sufficiently simple computations.

2.2. Literature Review

Deep learning techniques are being exploited in a lot of diverse biomedical applications to comprehend and analyze different illnesses. CNN was suggested by Perdomo et al. [15] as a means of identifying DME from OCT pictures. They presented the OCT-NET architecture, a 12-layer structure that uses median filters to highlight essential information in the picture layers.

A unique CNN design was presented by Jeffrey et al. [16] and is capable of recommending referrals for a range of retinal conditions. The work uses a 3 D U-Net model to extract a segmentation plot that successfully distinguishes different retinal tissues and structures from the provided retinal picture. After segmentation, the discovered retinal diseases are analyzed using a classification network to provide recommendations for referrals.

Fang et al. [17] suggested an abrasion-conscious CNN for OCT scan categorization that used data from the lesion regions. This improvement produced better categorization results, which are encouraging. Alqudah et al. [18] created a hybrid artificial intelligence system for efficient feature extraction in the multi-class classification of eye retina illnesses. Advanced OCT Network is used to take the highlights out of the pictures (AOCTNet). ANN is used to classify the OCT pictures after feature extraction.

Tuncer et al. [19] offer a This work uses a hybrid CNN model to categorize normal, CNV, and DME images using OCT photographs. The CNN-SVM model is created for OCT image categorization. The suggested method for processing the OCT images does not call for any extra feature extraction or noise filtering. For efficient classification, an SVM classifier is utilized in lieu of the SoftMax layer. SVM is utilized in the last pooling layer of CNN to classify the attributes of OCT images. Paima et al.'s convolutional neural network was built on the architecture known as the feature pyramid network [20]. This technique accurately diagnoses both wet AMD (CNV) and normal, dry AMD (drusen). In this method, the model is trained using a single CNN, negating the necessity for preprocessing the input data.

Akinniyi et al. [21] created a categorization network that was comprised of multiple stages for the classification of OCT images. Scale-adaptive neural networks

are established to create inputs for extraction of features at multiple scales. An enhanced feature-driven pyramidal network model is used to get the multi scale characteristics, with DenseNet serving as the backbone. A DL model with ResNet architecture comprising of 101 layers that was pretrained using the ImageNet dataset was proposed by Lu et al. [22]. In order to distinguish serious detachment of the macular layer, macular hole, epiretinal membrane, and cystoid macular edema from normal OCT pictures, four binary classifiers were trained independently.

A transfer learning (TL)-based CNN method was presented by Kermany et al. [23] for categorization of OCT images. The approach identified characteristics of both healthy and diseased OCT pictures, and the findings demonstrated that it could accurately categorize images of DME and AMD. Saleh et al. classified several retinal illnesses using a blend of DL and ML techniques in Refs. [24], [25], and [26]. A few studies that used deep learning frameworks to segment the IRF have also been published. A kernel regression-driven technique for determining the locations of fluid and retinal layers in OCT scans was presented by Chiu et al. [27].

A DL-based technique for multi-retinal fluid segmentation was introduced by Lu et al. [28]. The system utilized a complete CNN in conjunction with the segmentation outcomes produced by a graph-cut algorithm to distinguish between retinal fluid pixels. A method for segmenting and classifying retinal fluids through U-Net architecture was presented by Tennakoon et al. [29].

Considering the model's forecasts, this pipeline positively segmented retinal liquids at the voxel threshold. Ref. [29] proposed an IRF segmentation technique that made use of an 18-convolutional-layer altered iteration of the U-Net model. The findings suggested that in order to get adequate performance, training must cover a wide range of diseases. Upadhyay et al. created the coherent convolutional neural network (CCNN) for 4 class retinal disease categorization [30]. Accurate detection of OCT images, including DME, CNV, Normal, and Drusen is achieved by the established CCNN model.

This model accurately detects the erratic trends of retinal illness for every category. This technique effectively maintains consistency amid the network's input and output.

In this research, the network layer of the network is effectively integrated to accomplish the basic picture feature processing. The categorization process uses a five-layer CNN. The method that is being discussed incorporates the batch-normalization layer with every activity-layer to generate cohesive behavior.

Ţălu et al. [31] and Schmidt- rfurth et al. [32] said that OCT, which is divided into SD-OCT and TD-OCT categories, is a high-resolution imaging technology. The high-resolution cross-sectional and volumetric views of the retina are provided by the SD-OCT results.

A two-dimensional image of the specified internal retinal structure is produced by TD-OCT. Because TD-OCT only analyses the macula's thickness, it is ineffectual. In contrast, SD-OCT allows for the monitoring and evaluation of a variety of distinguishing characteristics. According to the study, OCT is a helpful method for evaluating, tracking, and analyzing the various stages of AMD. Furthermore, drusen's structure could be examined using a variety of features. Srinivasan et al. [33] suggested a support vector machine (SVM) model and histogram of oriented gradients model to successfully achieve the categorization of DME and dry AMD using optical CT scans. The inner retinal layers were not segmented in their suggested methodology. There were 45 volumetric scans in the SD-OCT datasets: 15 AMD, 15 DME, and 15 normal. The system detected all of AMD cases, all of DME patients, and 86.67 percent of normal cases with the highest specificity and flawless sensitivity. When it comes to high-resolution 3D tomography imaging employing optical coherence tomography (OCT) for retinal diagnosis, ophthalmologists' manual classification of OCT pictures is still subjective and time-consuming.

Addressing this challenge, the study [34]presents an automatic method for classifying retinal OCT scans using an IFCNN. By fusing data from current and previous layers of the convolutional neural network (CNN), the IFCNN leverages features from various scales within the CNN. This method enables a more thorough use of characteristics, resulting in precise classification of OCT images. The proposed IFCNN method outperforms conventional CNNs and known OCT classification

algorithms, as demonstrated by experimental findings on real retinal OCT and musculoskeletal radiograph datasets.

In [35], Resnet-50 [36] based transfer learning was used to develop a deep CNN, which was evaluated on a dataset labelled for CNV, DME, and drusen. According to reports, this model has a 96.1 percent accuracy rate. However, because there were so many parameters in the transfer learning-trained network, the system's complexity increased. Real-time deployment may not be appropriate for this kind of complex network. In order to categorize healthy retina, Drusen, Diabetic Macular Edema and Choroidal Neovascularization, a layer-guided CNN model was suggested in [37] which was 89.9 percent accurate.

More learnable parameters result from using two networks—one for classification and the other for layer segmentation—than from using only one. Deep learning was suggested as a quick and automated way to classify wet and dry AMD in [38] where just the wet and dry AMD were categorized using the same data set. A unique technique was developed to classify DME, two phases of drusen and Choroidal Neovascularization from healthy scan images in [39]. This presented a classification method-based CNN variant (MDFF). On the test set, average values of 99.6% for sensitivity, 99.6% for specificity, and 99.6% for accuracy were attained; nonetheless, the network's complexity resulted from its multi-scale structure and numerous learnable parameters.

A strategy for classifying AMDs based on deep learning processes was suggested in [40]. For categorization, the normalized oct scan image was loaded in a neural network. With the Inception v3 network, the reported accuracy was 96.93 percent. This type of heavy network might not be appropriate for instantaneous deployment because the pre-trained network contained a large number of learnable parameters. It was described how to automatically detect AMD and DME on OCT pictures. in [41]. A 97.1 percent accuracy was obtained using the AlexNet architecture. The enormous number of learnable parameters in the AlexNet made this technology unsuitable for real-time deployment as well.

The authors of [42] further tested retinal images with AMD grading. They employed a 21-layer CNN to do the binary classification task between healthy and AMD retinal images, and the result was a mean accuracy of 93%. The approach utilized in [43] tried to use dictionary learning and sparse coding to tackle the categorization challenge. This work shows that diabetic macular edema and drusen were involved in the binary classification, and over two datasets, a mean accuracy of 99% was recorded. So far, just a small number of photos were used to validate the algorithm. How well it would function on a sizable amount of untested data is unknown. In order to detect DME patients, a lot of prior methods have concentrated on the automatic analysis and classification of SD-OCT pictures. [44], [45].

This paper uses multiscale histograms of directed gradient descriptors as feature vectors for a support vector machine-based classifier, introducing a novel fully automated technique for the identification of retinal diseases via optical coherence tomography (OCT) imaging [46]. The classifier, trained and validated on spectral domain OCT datasets from 45 subjects, achieved impressive results, accurately identifying 100% of cases with dry age-related macular degeneration (AMD), 100% of cases with diabetic macular edema (DME), and 86.67% of cases for normal subjects. The algorithm presented in this paper emerges as a highly promising tool for remotely diagnosing ophthalmic diseases, providing a reliable and automated approach to identify and categorize retinal diseases with remarkable precision.

The automated classification of optical coherence tomography (OCT) images has become important in the noninvasive assessment of retinal eye disorders using OCT. A recent study [47] proposed a surrogate-assisted approach to classification using convolutional neural networks (CNNs). The topology involves image denoising, mask extraction, and generation of surrogate images for CNN training, achieving promising results with an AUC of 0.9783 in private dataset and 0.9856 in the duke database, highlighting its potential for automatic classification of retinal OCT images.

In [48], In order to distinguish between aberrant and healthy photos, researchers experimented with a pre-trained VGG-16 exploited for feature extraction. They found that the CNN achieved 93.5 percent accuracy and 81 percent specificity. Authors of

[49] used a combination of multi-scale CNN with a 98.66 percent claimed accuracy on 148 subjects in a private dataset and over 3000 photos from 45 subjects in a public dataset with disease grading of three classes. Even with the accuracy, their method is excessively complicated and laborious to run. Moreover, testing was done using a limited dataset. A technique for unsupervised learning was described in [50], intended to lessen the need for substantial amounts of training data in order to achieve correct categorization. A one-class SVM classifier was used after a multi-scale deep denoiser to denoise the images. The accuracy of this procedure was reported to be 81.4 percent, significantly lower than the accuracy of SOTA methods.

In [51], a feature extraction-based classification algorithm using 3000 photos from 45 people in a public dataset was experimented. The retinal OCT image was encoded using a multiscale Linear Configuration Pattern (LCP), and the optimal subset of features linked to the phenotypic was found using wrapper-based feature selection techniques. 99.3 percent accuracy was reported. Hand-crafted statistical features were combined with a Random Forest classifier in[52]. Following picture segmentation, ten features per image were obtained, and 15-fold cross-validation was applied to a complete dataset consisting of 177 AMD, 59 normal and 15 DME individuals. Ninety-six percent accuracy was reported.

In a recent work [53], model yielded a 100% accuracy rate. He employed an Inception v1 based model which was 230 layers deep in his network. A network with so many parameters is not appropriate for applications that require real-time processing. In [54], Perdomo proposed a CNN for automatic classification of normal and DME using sd-oct volumes and achieved the accuracy of 93%. In [55], Nugroho applied transfer learning approach and trained DenseNet and ResNet 50 achieving accuracy of 88 and 89 respectively. Lemaître et al., [56] tackled the issue of classifying SD-OCT data in order to automatically identify patients who are impacted by DME.

Tsanim et al. [57] Xception network, ResNet50, MobileNetV2, and Vanilla CNN were the four CNN models used to identify the illness categories from the retinal OCT scanned pictures. Feng et al., [58] used optical coherence tomography pictures to focus their investigation on a four-class retinal disease classification problem for the

identification of drusen, DME, CNV, and normal retina. They prepared a large collection of retinal OCT pictures and suggested a unique classification model for the automated detection of the majority of frequent blinding conditions. Improved ResNet50 served as the model's foundation. At the B-scan level, their method produced accuracy of 0.973, sensitivity of 0.963, specificity of 0.985, and AUC of 0.995. The prominent literature articles are presented in Table 1.

Table 1: Summarized literature review

Title	**Algorithm**	**Advantages**	**Limitations**	**DOP**
[38]	CNN	• Specific region emphasis • Innovative architecture • Improved accuracy	• Validation on limited datasets • Not applicable to real scenarios • Computational cost	2023
[39]	CNN	• Dual guidance network • Comprehensive evaluation Efficient testing	• Single label classification and lesion aggregation • Computationally complex • Less accurate • No interpretability	2023
[40]	Transformer network	• Clinical relevance • Efficiency improvement • Novel architecture	• Integration challenges • Accuracy improvement required • Interpretability • Complex architecture	2023
[37]	CNN ResNet50	TL approach utilized	• Data generalizability not integrated • Clinical applicability not in scope • Classification speed • Vital evaluation metrics ignored	2019

[29]	CNN VGG-16	Transfer Learning approach utilized	• Limited diversity in data • Accuracy not up to the mark and model is prone to misclassification • Interpretability not addressed	2017
[35]	Hybrid (Random Forest, KNN, SVM, Logistic Regression, Gradient boosting)	• Comprehensive investigation • Outperforming previous studies • Consideration of pre-processing steps • Classifies only DME condition	• Lack of external validation • Complexity of pre-processing • Limited performance metrics • Low accuracy • Classification speed • Computational complexity	2016
[31]	SVM	• Unsupervised anomaly identification • Scalability and reduced supervision • Qualitative analysis alignment • Novel machine learning approach • Performance	• Need for future validation • Limited generalized information • Interpretability not available • Accuracy of 94% can be improved • Complex architecture	2019
[18]	Transfer Learning (ANN)	• Transfer learning • General applicability • Transparent diagnosis	• Dataset specificity • Dependency on pre-trained weights • Accuracy concern • Number of iterations	2018

2.3. Summary

This chapter presents the extensive literature review done throughout the course of this research work providing insights on advantages limitations and research gaps upon which we have finalized our problem statement and carried out our research.

CHAPTER 3

PROPOSED METHODOLOGY

3.1. Introduction

The workflow of the proposed research will be included in this section, which covers the simulation environment, dataset description, proposed DA-CNN and CF-CNN models, optimization schemes and evaluation metrics.

3.2. Experimental Setup

In this research work, the CNN model utilized in order to classify optical coherence tomography pictures is designed and evaluated using a simulated environment. TensorFlow v2.14.0 with the Keras API in Python is used to implement the simulation framework. Google Collaboratory is used as a coding environment with an NVIDIA-SMI 525.105.17 Driver Version: Version 12.0 (CUDA) Graphics Processing Unit and running on a Lenovo 82TS with a 12th Gen Intel(R) Core (TM) i5-1235U 1.30 GHz and 8GB RAM.

3.3. Description of the Dataset

The most widely used dataset in computer vision for ophthalmology is the OCT 2017 Kermany dataset. It is regarded as one of the top databases for the analysis of OCT images. Together with a group of knowledgeable experts, Daniel Kermany assembled and released it initially. A wide range of OCT samples are included in the database, which is graphically divided into four groups: diabetic macular edema (DME), normal, Druse, and choroidal neovascularization (CNV).

As stated in the introduction chapter, these classes represent a variety of common eye disorders; as a result, the dataset is regarded as an essential tool for training and assessing machine learning models, which are subsequently used to diagnose and categorize ocular disorders to the fullest extent possible using data from OCT imaging.

Out of all the images that are included in the dataset, we have loaded 1500 images from each class i.e., Normal, CNV, DME, Drusen, so a total of 6000 training images, 8

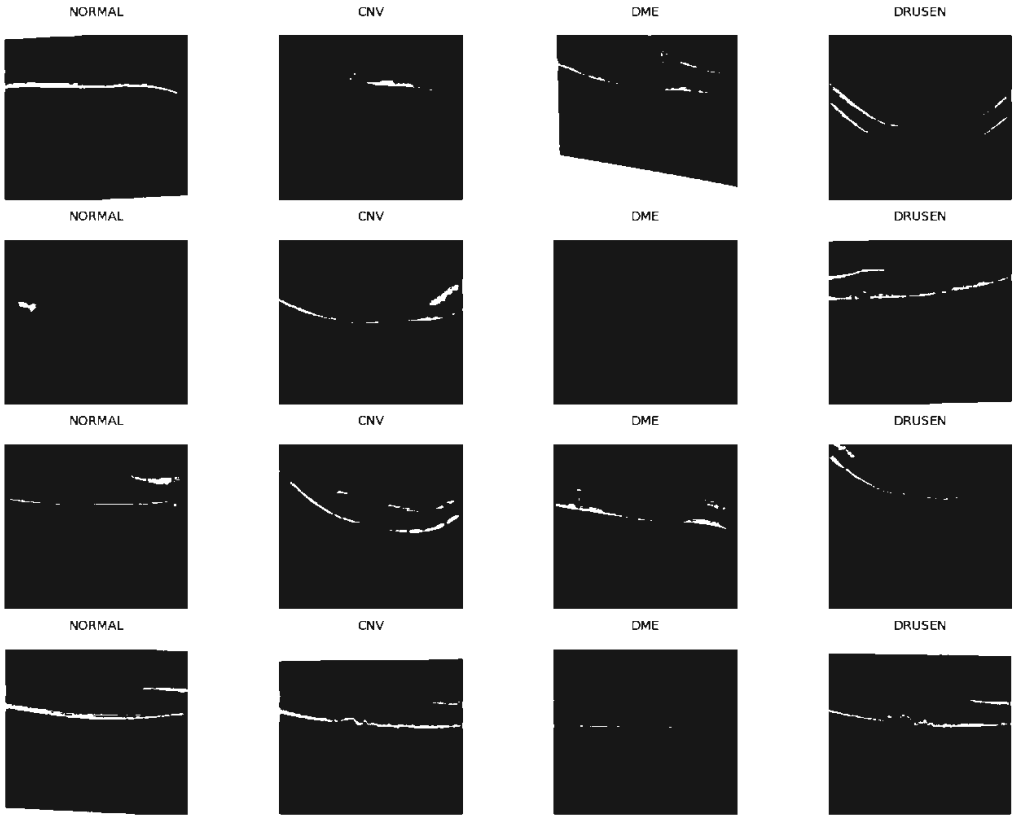


Figure 7: Sample images of the OCT Kermany 2018 Dataset at random

samples from each of the classes for validation, so a total of 32 images in validation dataset, 242 images from each class for testing hence 968 images in the test dataset. The dataset images that are obtained during the dataset analysis are presented in Figure 7**.**

3.4. Data Preprocessing

To prepare the data for training the CNN model, following approach is exploited.

3.4.1. Function to Load Images (get_data):

- This function loads a specific number of images i.e., 1500 from each class folder.
- It takes two parameters: **folder**, which is the directory containing class folders, and **num_images_per_class**, which specifies the number of images to load from each class.
- It initializes empty lists **X** for images and **y** for labels.

- It loops over each class ('NORMAL', 'CNV', 'DME', 'DRUSEN') and loads images from their respective folders.
- For each sample, it reads the image through OpenCV (**cv2.imread**) and resizes it to the specified **imageSize** using scikit-image (**skimage.transform.resize**).
- It converts the image to a NumPy array and appends it to the **X** list, while appending the label (encoded as an integer) to the **y** list.
- The function returns NumPy arrays **X** containing images and **y** containing corresponding labels.

3.4.2. Data Loading:

- It defines the directories for the training, testing, and validation sets (**train_dir**, **test_dir**, **validation_dir**).
- It sets the number of images to load per class (**num_images_per_class**).
- It calls the **get_data** function for each directory to load the images and labels for training, testing, and validation sets.

3.4.3. One-Hot Encoding:

- It one-hot encodes the labels using **to_categorical** function from Keras (**keras.utils.to_categorical**).
- With one-hot encoding, binary vectors representing integer class labels are created, with the exception of the index corresponding to the class label, which is set to 1. Each class is represented by a vector of all zeros.

$$y_{one-hot}[i] = \begin{cases} 1 & i=y \\ 0 & otherwise \end{cases} \quad (7)$$

  where i ranges from zero to N-1 representing the class indices.

- This is a common preprocessing step for categorical variables in classification tasks and helps the neural network to better understand the class relationships.

3.5. Proposed CF-CNN and DA- CNN Models

3.5.1. Dual Attentive Convolutional Neural Network or DA-CNN

The proposed DA-CNN model comprises of three convolutional blocks to begin with, five layers constituting the first two blocks and six layers constituting the third convolutional block. The first one is separable dynamic convolution layer, next up is the activation-layer that is governed by ReLU, then the batch normalization layer, then the max-pooling layer with pool size of 3*3 and stride of 2*2 and at the end, the dropout layer with rate of 20% to prevent overfitting.

This pattern is repeated for two blocks, the third one contains an additional channel split dual attention mechanism layer in between the activation and batch normalization layer. The convolutional filter size for first convolution block is 16 with the stride of 3 whereas the convolutional filter size for second and third convolutional block is 32 with the same stride. In separable dynamic convolution, (on the contrary to conventional static convolution in which a single set of filters is applied to the entire input feature map both spatially and across the channels), the convolutional operation is decomposed into spatial convolutions followed by channel wise convolutions, each with their own set of filters. Mathematically, the conventional convolution can be given as:

Let K denote the kernel filter and I denote the input feature map. The output feature map O is computed as follows:

$$O_{i,j}^{k} = \sum_{m,n,c} Q_{m,n,c} . I_{i+e,x+y}^{c}$$

where i,x are spatial indices, k is the channel index, m,n are kernel indices, c is the channel index of the input feature map, $Q_{m,n,c}$ is the kernel weight at position (m,n,c), $I_{i+e,x+y}^{c}$ is the input feature map value at $(i+e,x+y)$ and channel c.

Whereas the separable dynamic convolution can mathematically be expressed as:

Let K_s denote the spatial kernel and K_c denote the channel wise kernel, the output feature map O can be computed as follows:

$$O_{i,j}^{k} = \sum_{m,n} K_{s}^{m,n} (\sum_{c} K_{c}^{c} . I_{i+m,j+n}^{c})$$

where $K_s^{m,n}$ is the spatial kernel weight at (m,n), K_c^c is the channel-wise kernel weight for channel c and all other variables are defined as in conventional static convolution. The block diagram of separable dynamic convolution is shown in Figure 8.

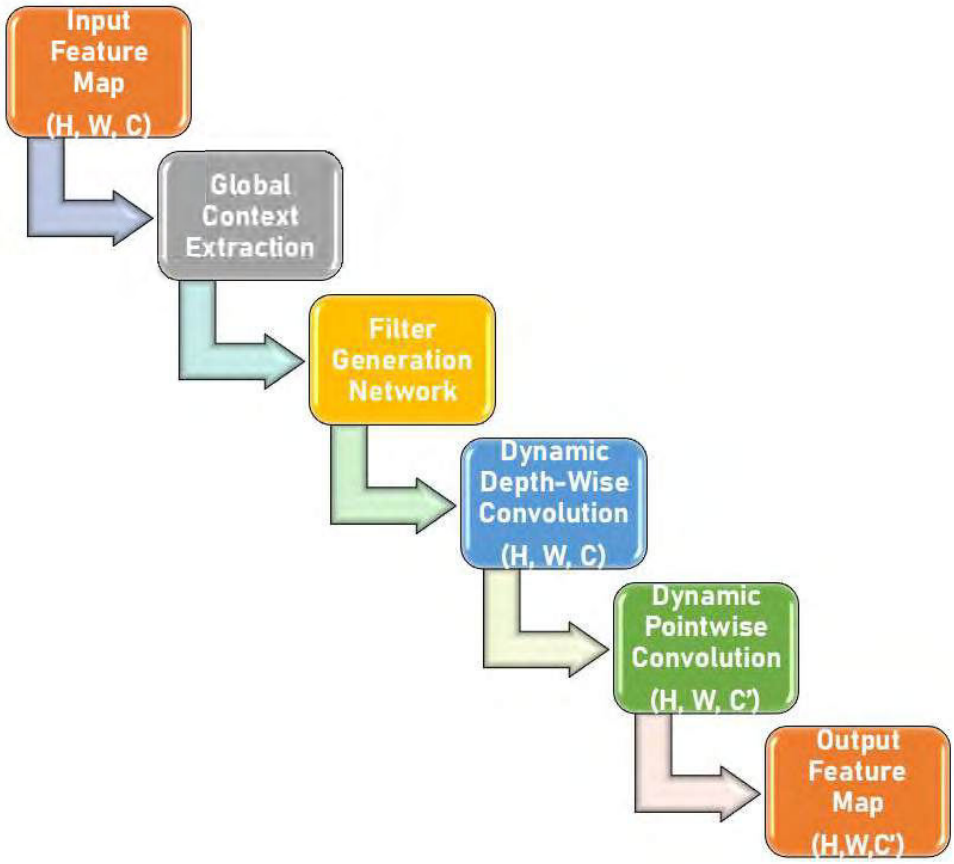


Figure 8: Block diagram of Separable Dynamic Convolution Operation

The key difference among the two convolution techniques lies in the decomposition of convolution operation. In the separable dynamic convolution, the convolutional operation is split into spatial and channel wise convolutions hence allowing for more efficient computations. By applying the separate filters for spatial and channel wise convolutions, the number of parameters to be learned is reduced, leading to computational savings and ultimately improved generalized performance of the model on the basis of enhanced feature extraction. Moreover, the separable dynamic convolution can capture both spatial and channel wise dependencies more effectively as compared to conventional static convolution.

Another important contribution of this research work is the incorporation of channel-split dual attention mechanism with convolutional filter size of 32 and stride of 3. The Channel Split Dual Attention Mechanism is an incredible technique that is exploited in Convolutional Neural Networks (CNNs) to vividly improve feature representation by concurrently recording channel-wise and spatial relationships in feature maps. It features the combination of spatial and channel attention mechanisms. First, it performs channel attention based on the attention weighted features and then spatial attention on those for spatial locations. Spatial attention is concerned with the materialization of relevant channel maps, while channel attention focused on convolutional channels highlights only the essential channels.

These attention maps are very specific for every image patch area. The feature maps are the same and these attention maps are integrated to create a final attention map that modifies the source feature maps. From the mathematical point of view, spatial attention is performed using a SoftMax function on the result of the linear transformation followed by ReLU activation, while channel attention refers to a sigmoid function on the result of an average or mean pooling on channel-wise element values.

A final map of attention is obtained by means of element-wise multiplications of two attention maps, namely, the spatial and channel attention maps. This is introduced into CNN architectures; it is where this mechanism comes in to play by boosting model efficiency in tasks including object detection, segmentation, and image classification through the ability to focus on relevant spatial regions and channels, hence the models

produce more discriminative feature representation and get better at generalization. Mathematically, the channel-split dual attention can be expressed as:

3.5.1.1. Spatial Attention

$$A_s = soft\max(W_s.\mathrm{Re}\,LU(W_f.X_{i,j}))$$

where A_s is the spatial attention map, W_s *and* W_f are learnable parameters (weights) of the spatial attention mechanism. $X_{i,j}$ displays the feature map at spatial location *(i,j)*, ReLU is the rectified linear unit activation function, softmax computes the softmax function along the spatial dimensions hence yielding attention weights A_s for each spatial location.

3.5.1.2. Channel Attention

$$A_c = sigmoid(W_c.avgpool(X)) \qquad (4)$$

where A_c is the channel attention map, W_c is the learnable parameter (weight) of the channel attention mechanism, *avgpool* is the average pooling operation performed along the spatial dimensions to obtain global channel-wise statistics, *sigmoid* computes the sigmoid function to normalize the attention weights A_c between zero and one.

3.5.1.3. Final Attention

$$A = A_s \otimes A_c \qquad (5)$$

where A is the final attention map, $\otimes$ denotes the element wise multiplication or Hadamard product, A_s is spatial attention map and A_c is the channel attention map.

3.5.1.4. Enhanced Feature Representation

$$X' = A \otimes X$$

where X' represents the enhanced feature maps after application of attention mechanism and X is the original feature map.

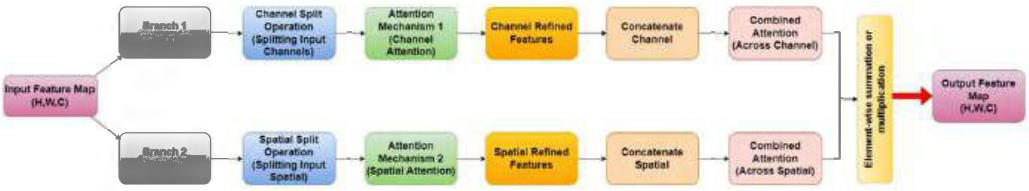


Figure 9: Block diagram of Channel Split Dual Attention technique

In a nutshell the channel split dual attention leads to more discriminative feature representation and improved performance by permitting the network to emphasis on prominent spatial areas and channels within the feature maps. The block diagram of channel split dual attention is presented in Figure 9.

To continue the model architecture, followed by the convolutional blocks is a flattening layer to convert the data into 1-D array. The FC layer with 128 neurons, ReLU activation, batch-normalization and drop-out layer with rate of 50% to prevent overfitting. The last layer of the proposed CNN model is dense layer having 4 units, representing the ocular condition whether Normal, CNV, DME or Drusen. The final classification layer has SoftMax activation. The model summary is presented in the table. Figure 10 displays the overall block diagram of the suggested DA-CNN model. Table 2 presents the detailed model architecture along with the parameters that is obtained in Python environment using model.summary() function.

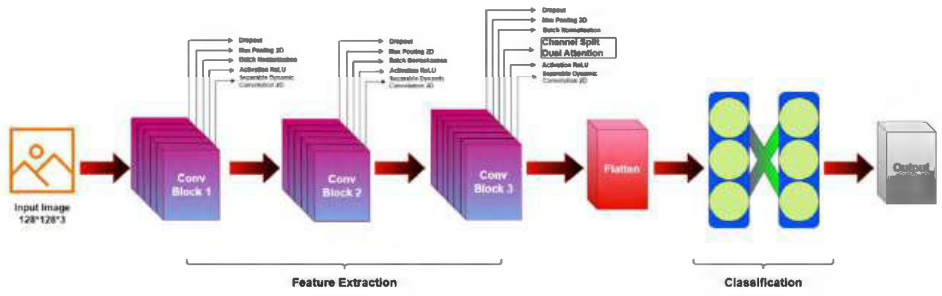


Figure 10: Block diagram of DA-CNN model architecture

Table 2: Summary of the layers of the DA-CNN Model

Sr.	Layer	Output shape	Params#
1	Separable Dynamic-Convolution 2D	None, 126,126,16	91
2	Activation	None, 126,126,16	0
3	Batch-Normalization	None, 126,126,16	64
4	Max-Pooling 2D	None, 62,62,16	0
5	Drop-out	None, 62,62,16	0
6	Separable Dynamic-Convolution 2D	None, 60,60,32	688
7	Activation	None, 60,60,32	0
8	Batch-Normalization	None, 60,60,32	128
9	Max-Pooling 2D	None, 29,29,32	0
10	Drop-out	None, 29,29,32	0
11	Separable Dynamic-Convolution 2D	None, 27,27,32	1344
12	Activation	None, 27,27,32	0
13	Channel Split-Dual Attention	None, 27,27,64	18496
14	Batch-Normalization	None, 27,27,64	256
15	Max-Pooling 2D	None, 13,13,64	0
16	Drop-out	None, 13,13,64	0
17	Flatten	None, 10816	0
18	Dense	None, 128	1384576
19	Activation	None, 128	0
20	Batch-Normalization	None, 128	512
21	Drop-out	None, 128	0
22	Dense (Eye-Condition)	None, 4	516

Total params: 1406671 (5.37 MB), Trainable params: 1406191 (5.36 MB)

3.5.2. Channel Focused Convolutional Neural Network of CF-CNN

The CF-CNN model is implemented using the Keras Sequential API and comprises a series of layers for a CNN with convolutional layers that is intended for image classification. The model begins with a DynamicConv2D layer with 16 filters and a (3, 3) filter size, a Rectified Linear Unit (ReLU) activation function came next, an attention

layer with 16 filters and a (1, 1) kernel size, batch normalization, and max-pooling with a (2, 2) pool size. Next, the same convolution blocks are repeated two times as Conv2 and Conv3 having the same structure as first convolutional block. The key point in Conv1 and Conv2 units is the convolution attention mechanism that is employed for the enhanced feature extraction and hence better accuracy. Following the convolutional blocks containing activation, batch normalization and dropout layers, flatten layer is incorporated to convert the output of last layer in 1D array. Next to flatten layer are the Fully connected layers containing 128 neurons for the first layer having ReLU activation and dropout rate of 20 percent to avoid overfitting. Lastly, the model concludes with a final FC layer encompassing 4 units, utilizing a SoftMax activation function to represent the various output classes for eye conditions. The model structure has properly schemed dynamic convolutional layers, activation functions, attention mechanisms, and regular CNN components like batch normalization and max-pooling. Dropout is applied only in the FC layers to enhance generalization. The last layer produces a SoftMax output for classification into four eye condition classes.

3.5.2.1. Dynamic Convolution:

Dynamic convolution layers represent a remarkable development over traditional convolution layers in Convolutional Neural Networks. In standard convolution, the kernel remains fixed during inference, meaning that a single set of weights is applied uniformly across the entire input feature map. While effective, this static nature limits the adaptability of the model, as the same filter processes all inputs, regardless of their local context or variations.

Contrary to this, dynamic convolution incorporates adaptability by allowing the convolutional filters to change dynamically based on the input features or some external conditioning. This adaptability is achieved through a mechanism where multiple sets of filters are learned, and a dynamic aggregation or selection process determines which filters to apply at each spatial location. Mathematically, if X represents the input feature map and $\{W_1, W_2, \ldots, W_k\}$ are the possible filter sets, dynamic convolution computes the output as a weighted sum of these filters:

$z = \sum_{a=1}^{s} \varphi_a(u) * W_i$

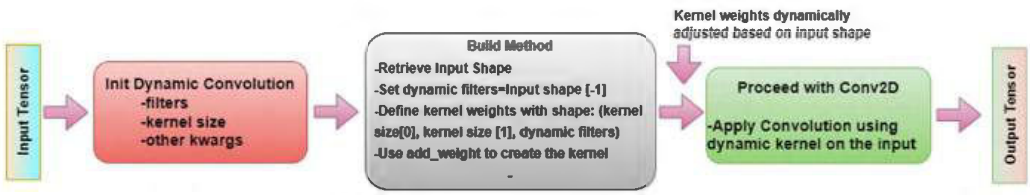


Figure 11: Block diagram of Dynamic Convolution Operation

Here, $\varphi_a(u)$ are the dynamically computed coefficients that depend on the input u, and * denotes the convolution operation. The coefficients $\varphi_a(u)$ are often obtained using an additional lightweight network, such as a small fully connected layer or a SoftMax function, ensuring that the filter selection is both data-driven and context-sensitive.

This dynamic approach enables the network to adapt its behaviour to different regions of the input. By allowing the network to effectively "choose" the most relevant filters on the fly, dynamic convolution layers can lead to more efficient and expressive models compared to their static counterparts, often resulting in improved generalization and robustness across varied datasets. Figure 11 presents the block diagram of dynamic convolution operation

3.5.2.2. Channel Split Dual Attention Mechanism:

The Channel Split Dual Attention (CSDA) technique is an advanced mechanism designed to enhance the representational capacity of convolutional neural networks (CNNs) by focusing on important features within input tensors. It splits the input tensor into two pathways: one emphasizes global average features, while the other highlights the maximum values. Each pathway generates an attention map, which is then used to re-weight the original features through element-wise multiplication, enhancing relevant information and suppressing less important data.

These re-weighted features are then concatenated and passed to subsequent layers, resulting in more informative and discriminative feature representations. The CSDA technique is particularly beneficial in tasks requiring fine-grained feature detection, such as medical image analysis and fine-grained object recognition, and can be easily integrated into existing CNN architectures with minimal computational overhead. By

selectively emphasizing important channels and spatial regions, the CSDA technique improves the performance of deep learning models in various applications. The block diagram of the phenomenon is represented in Figure 12. The overall architectural diagram of the proposed CF-CNN model is shown in Figure 13. Table 3 presents the architectural summary of the layers along with filter sizes and parameter numbers.

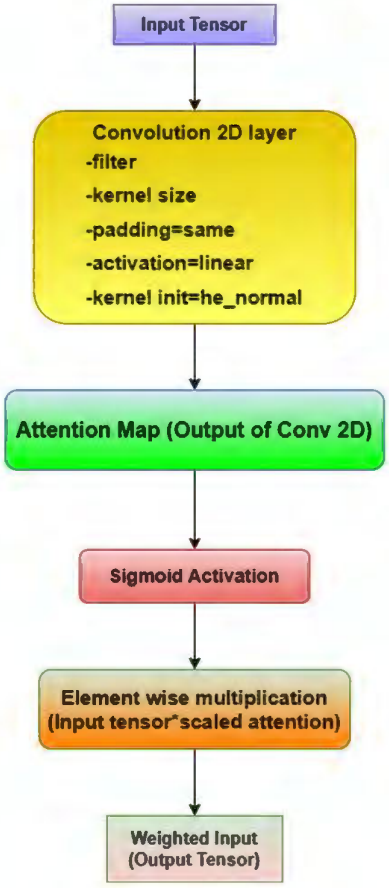


Figure 12: Flowchart of Channel Split Dual Attention technique

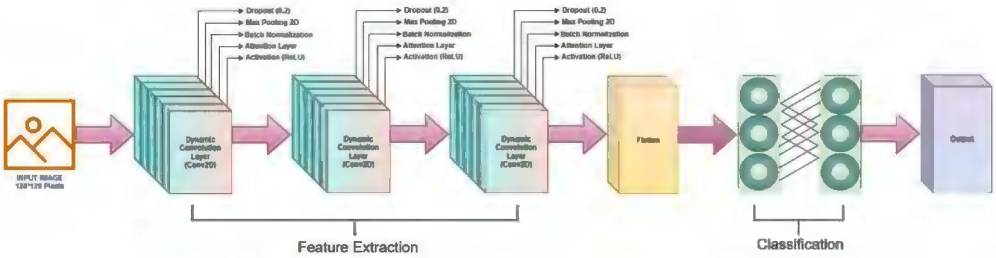


Figure 13: Block diagram of CF-CNN architecture

Table 3: Architectural summary of CF-CNN model

Sr.	Layer type	Output Shape	Param #
1	dynamic_conv2d (Dynamic Convolution 2D)	None, 128,128,16	448
2	activation (Activation)	None, 128,128,16	0
3	conv_attention_layer (ConvAttentionLayer)	None, 128,128,16	272
4	batch_normalization (Batch Normalization)	None, 128,128,16	64
5	max_pooling2d (Max Pooling 2D)	None,64,64,16	0
6	dynamic_conv2d_1 (Dynamic Convolution 2D)	None,64,64,16	2320
7	activation_2 (Activation)	None,64,64,16	0
8	conv_attention_layer_1 (ConvAttentionLayer)	None,64,64,16	272
9	batch_normalization_1 (Batch Normalization)	None,64,64,16	64
10	max_pooling2d_1 (Max Pooling 2D)	None, 32,32,16	0
11	dynamic_conv2d_2 (Dynamic Convolution 2D)	None, 32,32,16	2320
12	activation_4 (Activation)	None, 32,32,16	0
13	conv_attention_layer_2 (ConvAttentionLayer)	None, 32,32,16	272
14	batch_normalization_2 (Batch Normalization)	None, 32,32,16	64
15	max_pooling2d_2 (Max Pooling 2D)	None, 16,16,16	0
16	flatten (Flatten)	None, 4096	0
17	dense (Dense)	None, 128	524416
18	activation_6	None, 128	0
19	batch_normalization_3 (Batch Normalization)	None, 128	512
20	dropout (Dropout)	None, 128	0
21	Eye_Condition (Dense)	None, 4	516

Total params: 531540 (2.03 MB) Trainable: 531188 (2.03 MB) Non-trainable: 352 (1.38 KB)

3.5.3. Optimizers

Optimizers, that are applied while compiling a machine learning or deep learning model, are of significant importance when training the neural networks as they iteratively modify the model parameters and reduce the cost function during the

training-process. The primary objective of optimizers is to help the algorithm for optimization to determine the ideal combination of parameters to attain the optimal results.

During the process of training, the model is presented with the input labelled data in different batch sizes and ultimately predictions are made. Subsequently, the alteration amongst the goal values and the forecasted values is measured by the loss function or the cost function. The optimizer determines the direction and magnitudes towards the steepest ascent by calculating the gradient of this loss function regarding each parameter. By adjusting the parameters in the direction opposite to these gradients, the optimizer seeks to advance the model's efficiency by minimizing the loss function.

Including the advantages and limitation, each optimizer exploits various strategies to update the parameters. As in case of Stochastic Gradient Descent, it updates the parameters that are directly proportional to the negative gradient and learning rate. The detailed description of all the optimizers that are in the scope of this study is presented in the following sections.

3.5.3.1. Adafactor

Adafactor is a variation of the AdaGrad optimizer that adapts learning rates according to several factors by performing the following:

a. Adaptive Learning Rate Scaling

Adafactor scales the learning-rate differently for each parameter based on the historical gradients of that parameter. It uses a two-dimensional parameter scale matrix, which helps in scaling the learning rates differently for different parameters.

b. Adaptive Gradient Clipping

It clips the per-parameter gradients based on the statistics of the past gradients.

c. Initialization

Initialize the scale matrix G as an identity matrix of size (d,d) where d is the amount of parameters

Initialize the exponential moving average of squared gradients *v* as a zero matrix of the same size as *G*.

d. Per-Parameter Gradient Scaling

- For each parameter w_i compute its gradient g_i.
- Update the squared gradients exponential moving average v:

$$v_i \leftarrow \beta_2 . v_i + (1 - \beta_2) . g_i^2$$

Update the scale matrix *G*

$$G_{ii} \leftarrow \max(G_{ii}, v_i)$$

e. Compute the effective learning rate

The parameter's effective learning rate can be computed as follows:

$$lr_i = \frac{lr.scale}{\sqrt{v_i}}$$

where *lr* is the base learning rate and scale is the scalar value for numerical stability, typically set to “ $\min(1.0, \sqrt{1.0 / scale\_parameter})$

f. Update Parameters

Each parameter w_i is updated using the corresponding effective learning rate lr_i

$$w_i \leftarrow w_i - lr_i . g_i$$

Optionally, clip gradients before updating the parameters to prevent large gradient updates that might lead to unstable training. This whole process repeats for each training iteration.

β_2 refers to the squared gradients of the moving average exponential decay rate; *lr* corresponds to the learning rate; and the scale parameter is for numerical stability.

By automatically adjusting the learning rates for various parameters according to their past gradients, this technique helps to increase convergence and stability during training.

3.5.3.2. Root Mean Square Propagation (RMSprop)

An optimization approach called Root Mean Square Propagation, or RMSprop, modifies the learning rate for each parameter according on the strength of its gradients. The following are the RMSprop optimizer's mathematical expressions:

a. Initialization

- Learning rate η: Step size for updating the parameters
- Decay rate ρ: Decay rate for the moving average of squared gradients
- Epsilon ε: Small constant for numeric stability
- Initialization step $t = 0$
- Initialize cache $= 0$ (Initial accumulated squared gradients)

b. Update Rule

The update strategy for the RMSprop optimizer can be written as follows given a loss function *L* and its gradient in terms of the model parameters:

$$cache_t = \rho.cache_{t-1} + (1-\rho).g_t^2$$

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{cache_t + \varepsilon}}.g_t$$

Here, *t* represents the time step, θ_t represents the models parameters at that particular time step *t*, moving average of the squared gradients is represented by $cache_t$, ρ represents the decay rate, η the learning rate, ε is considered as a small constant to prevent division by zero.

c. Adaptation of learning rate

Based on the size of each parameter's gradients, RMSprop modifies the learning rate for each one separately. Larger gradient values will have a lower effective learning rate, whereas smaller gradient parameters would have a higher effective learning rate. This aids in preventing oscillations or divergence during training due to an excessively high learning rate. A small constant 'epsilon' is added to the denominator, particularly when the cumulative squared gradients are minimal, to prevent division by zero and provide numerical stability.

RMSprop is a highly helpful optimizer for training neural networks since it adapts the learning rates to stabilize the process. It is a widely applied optimization method that serves as the basis for other algorithms, such as Adam.

3.5.3.3. Adam

The Adam optimizer is a method that integrates the concepts of RMSprop and momentum-based optimization. The following are the Adam optimizer's mathematical expressions:

a. Initialization

- Learning rate η : Step size for updating the parameters
- β_1 : exponential decay rate for first moment
- β_2 : exponential decay rate for second moment
- ε : Small constant for numeric stability
- *t=0* Initialization step

- Initialize $m_0 = 0$, The first initial moment vector
- Initialize $\nu_0 = 0$, The second initial moment vector

b. Update Rule

Assuming a loss function *L* and its gradient concerning the model parameters, the Adam optimizer's update rule may be articulated in this way:

$$m_t = \beta_1 . m_{t-1} + (1 - \beta_1) . g_t$$

$$\nu_t = \beta_2 . \nu_{t-1} + (1 - \beta_2) . g_t^{\,2}$$

14

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}, \hat{u}_t = \frac{u_t}{1 - \beta_2^t}, \psi_{t+1} = \psi_t - \eta . \frac{\hat{m}_t}{\sqrt{\hat{u}_t} + \varepsilon}$$

where *t*, as usual represents the time step, ψ_t are model parameters at time step *t*, g_t is the loss function gradient with respect to the model parameters at time step t, m_t is the first moment estimate (mean) of gradients, u_t is the second moment estimate (uncentered estimate) of the gradients, $\hat{m}_t$ and $\hat{u}_t$ are bias-corrected estimates of m_t and ν_t to account for their initialization at zero, η is the learning rate.

c. Bias correction

The Adam optimizer applies bias correction to the moment estimates to account for their initialization at zero. This correction is necessary particularly at the beginning of training when *t* is small. Adam can also incorporate optional L2 regularization (weight decay) by adding $\lambda . \theta_t$ to the gradient update, where λ is the regularization parameter.

Because Adam is adaptable, it can handle a variety of data formats and model topologies with ease, which makes it a popular choice in practice. Adam converges more quickly and is more resilient to noisy gradients thanks to the combination of momentum and RMSprop.

3.5.3.4. Nadam

Nadam is an advancement of the Adam that incorporates Nesterov momentum into the adaptive learning rate scheme. The mathematical expressions of Nadam optimizer can be given as follows:

a. Initialization

- Learning rate η : Step size for updating the parameters
- β_1 : exponential decay rate for first moment
- β_2 : exponential decay rate for second moment
- ε : Small constant for numeric stability
- *t=0* Initialization step
- Initialize $m_0 = 0$, The first initial moment vector
- Initialize $v_0 = 0$, The second initial moment vector

b. Update Rule

Given a loss function *L* and its gradient with respect to model parameters g_t, the update rule for Nadam optimizer is same as Adam and can be formulated as follows:

$$m_t = \beta_1 . m_{t-1} + (1 - \beta_1) . g_t$$

16

$$v_t = \beta_2 . v_{t-1} + (1 - \beta_2) . g_t^2$$

17

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}, \hat{u}_t = \frac{u_t}{1 - \beta_2^t}, \psi_{t+1} = \psi_t - \eta . \frac{\hat{m}_t}{\sqrt{\hat{u}_t} + \varepsilon}$$

where t is time step, ψ_t are model parameters at time step t, g_t is the loss function gradient with respect to the model parameters at time step t, m_t is the first moment estimate (mean) of gradients, u_t is the second moment estimate (uncentered estimate) of the gradients, $\hat{m}_t$ and $\hat{u}_t$ are bias-corrected estimates of m_t and v_t to account for their initialization at zero, η is the learning rate.

c. Nesterov Momentum

Nadam incorporates Nesterov momentum into the update rule by updating the parameters in two steps. First, it calculates the gradient at the estimated next position

$$\theta_t - \eta . \frac{\hat{m}_t + (1-\beta_t).g_t}{\sqrt{\hat{v}_t} + \varepsilon}$$

then it applies the gradient descent step using this estimated gradient.

Nadam is effective for training neural networks, combining the advantages of adaptive learning rates (as in Adam) with the improved convergence properties of Nesterov momentum. It is widely used in practice for optimizing deep learning models.

3.5.4. Evaluation Metrics

During the evaluation phase of our proposed DA-CNN model, a wide range of evaluation measures are employed to attain a complete comprehension of the model's functionality and efficacy. The accuracy and other factors that effectively convey the performance of the model are included in these assessments. Higher the accuracy, better the model's performance. The performance metrics employed are presented in this section.

3.5.4.1. Accuracy

The primary criterion for evaluating the model's overall soundness is its accuracy. It is considered by taking the total quantity of predictions and dividing it by the right number of predictions.

$$Accurcy = \frac{P_c}{P_t} \times 100$$

Where P_c represents the number of accurate predictions and P_t represents the number of total predictions.

3.5.4.2. Mean Absolute Error

MAE is the most crucial metric for assessing how well ML and DL models are working. Essentially, it informs us of the degree to which the predicted value differs from the true or anticipated value. The positive difference among the true and predicted value at each data point is utilized in the computation. The model performs better when the MAE value is lower. Mathematically, it can be shown as:

$$MAE = \frac{1}{z} \sum_{i=1}^{z} | k_i - \hat{k}_i |$$

Here, z shows the number of total data points, k_i shows the measured value for i^{th} data point, $\hat{k}_i$ shows the predicted value for i^{th} data point.

3.5.4.3. Dice Coefficient (DSC)

Dice coefficients are frequently utilized performance statistic in computer vision and medical image processing that evaluates the level of resemblance between two sets and the accuracy or total overlap of segmentation algorithms. This metric yields a value that represents the extent to which segmentations of objects within an image truly overlap. On the dice coefficient scale, which goes from 0 to 1, a higher score indicates greater overlap or agreement between the two segmentations, and vice versa. The dice coefficient can be calculated mathematically using the following formula:

$$DiceCoeff = \frac{2 \times | L \cap M |}{| L | + | M |}$$

where L denotes the entire set of pixels that the predicted segmentation classifies as belonging to the object, M represents the whole set of pixels that the true segmentation, which is used as a reference, and $|L \cap M|$ denotes the total number of pixels that overlap between the two sets.

3.5.4.4. Sensitivity

In the context of CNN models, sensitivity is an evaluation metric that is employed to get a picture of model's performance in the classification tasks. It can also be termed as true positive rate or the recall. In essence, sensitivity indicates how well the model detects positive cases among all of the real positive examples present in the dataset. It is especially crucial in situations like anomaly detection or medical diagnostics where accurately recognizing positive cases is essential. This is basically the number of actual positive cases that are correctly predicted by the model.

Mathematically, it can be expressed as:

$$Sety = \frac{Tx}{Tx + Uy}$$

Tx represents the true positives and *Uy* represents the false negatives.

3.5.4.5. Specificity

On the contrary to sensitivity, the specificity measures the true negative cases that the model accurately identifies. The model's specificity indicates how well it detects negative cases among all of the real negative cases in the dataset. It is especially significant in applications related to security or quality control, where accurately recognizing negative cases is essential. Mathematically, it can be expressed as:

$$Spty = \frac{Ty}{Ux + Ty}$$

Ty represents true negatives and *Ux* represents false positives.

3.6. Summary

This chapter presents the detailed methodology which has been followed throughout the course of this research work. This include all the technicalities and details on the model architectures, dataset description, optimization techniques and hyperparameter tuning process along with theoretical and mathematical frameworks of the modalities and optimizers followed by the evaluation metrics that have been employed to get a clear understanding on the performance of the proposed models.

CHAPTER 4

SIMULATIONS AND ANALYSES

4.1. Introduction

This section presents various subsections regarding the results obtained during the course of simulation and a case study based on three different optimization strategies as well as the evaluation of the proposed models on various evaluation metrics.

4.2. Simulations and Results

This section refers to the outcomes that are attained after the comprehensive experimentation in the environment mentioned in section 3. The extensive pre-experimentation leads us to define some optimal hyperparameters such as 0.001 learning rate, 32-sample batch size, verbose equal to 2 and the training process continued for 100 epochs. The loss monitored was categorical cross-entropy and five benchmark evaluation metrics such as accuracy, mean absolute error, dice-coefficient, sensitivity and specificity were monitored epoch wise.

4.3. Learning behavior of DA-CNN model

4.3.1. Case-1: Adam

In the first case study, the suggested CNN architecture has been compiled by the Adam optimizer that combines the ideas of momentum-based optimization and RMSprop (Root Mean Square Propagation) into a single algorithm as discussed in detail in section 5.3. The optimal hyperparameter stood out at batch-size of 32, learning-rate of 0.001 for 100 iterations. The model's accuracy came out to be 91.11% as the result of implementing the enhanced feature extraction techniques based on separable dynamic convolution and channel split dual attention. The model has demonstrated the bias of 0.05 and variance of 0.10. Moreover, the generalized accuracy stands out to be 0.91, test loss of 0.28, test MAE of 0.08, test dice coefficient of 0.84, test sensitivity of 0.90

and test specificity of 0.97, AUR-ROC of 0.991. The epoch wise trend of evaluation metrics on training and validation sets is presented in Figure 14. Furthermore, the resulted confusion matrix is displayed in Figure 15.

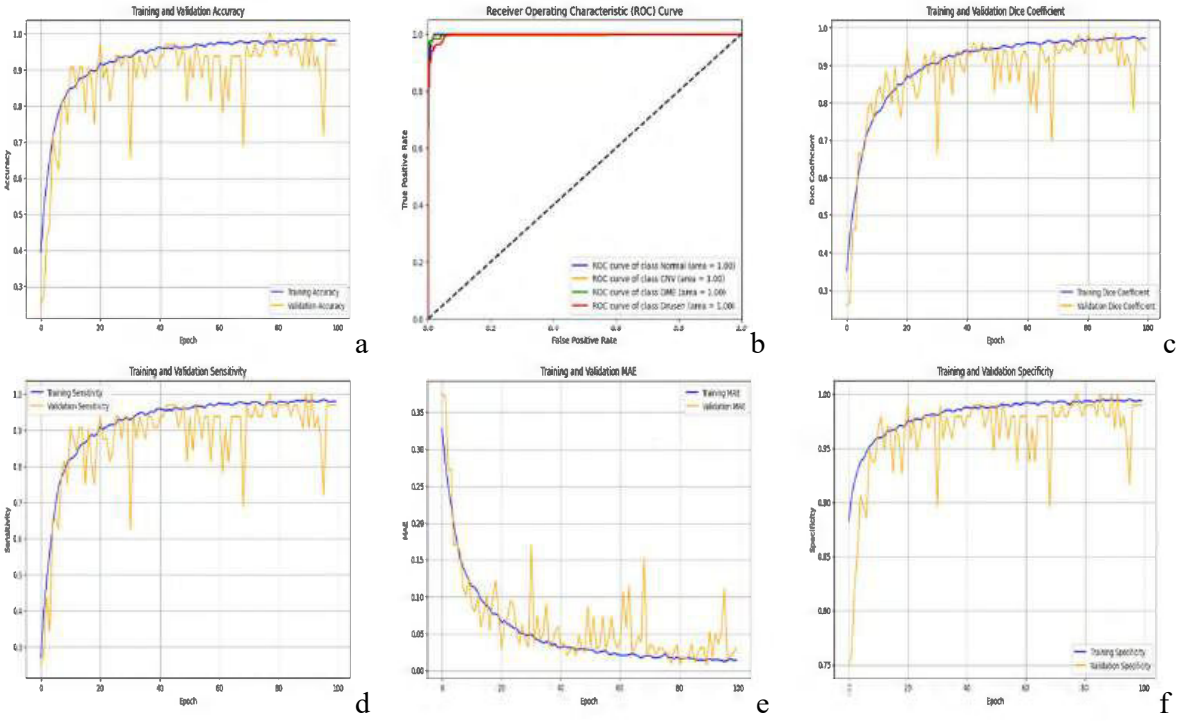


Figure 14: Final AUC-ROC Curve of the model(b); Epoch wise trend of evaluation metrics (a) Accuracy (c) Dice coefficient (d) Sensitivity (e) Mean Absolute Error (f) Specificity

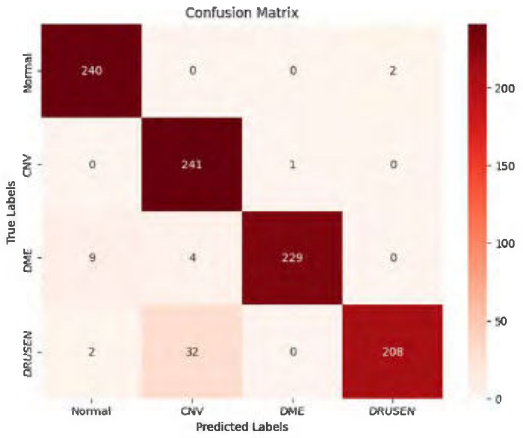


Figure 15: Confusion matrix presenting the complete picture of predictive capabilities of the model

4.3.2. Case-2: Adafactor

In the second case study, the proposed DA-CNN model has been compiled using the Adafactor optimizer which scales the learning-rate differently for each parameter based on the past gradients of that parameter as discussed in detail in section 5.1. The optimal hyperparameter stood out at batch-size of 32, learning-rate of 0.001 for 100 iterations. The model's accuracy came out to be 91.01% as the result of implementing the enhanced feature extraction techniques based on separable dynamic convolution and channel split dual attention. The model has demonstrated the bias of 0.36 and variance of 0.21. Moreover, the generalized accuracy comes out to be 0.91, test loss of 0.28, test MAE of 0.08, test dice coefficient of 0.85, test sensitivity of 0.90 and test specificity of 0.97, AUC-ROC of 0.991. The epoch wise trend of evaluation metrics on training and validation sets along with confusion matrix is presented in Figure 16**.** Furthermore, the computed confusion matrix is displayed in Figure 17.

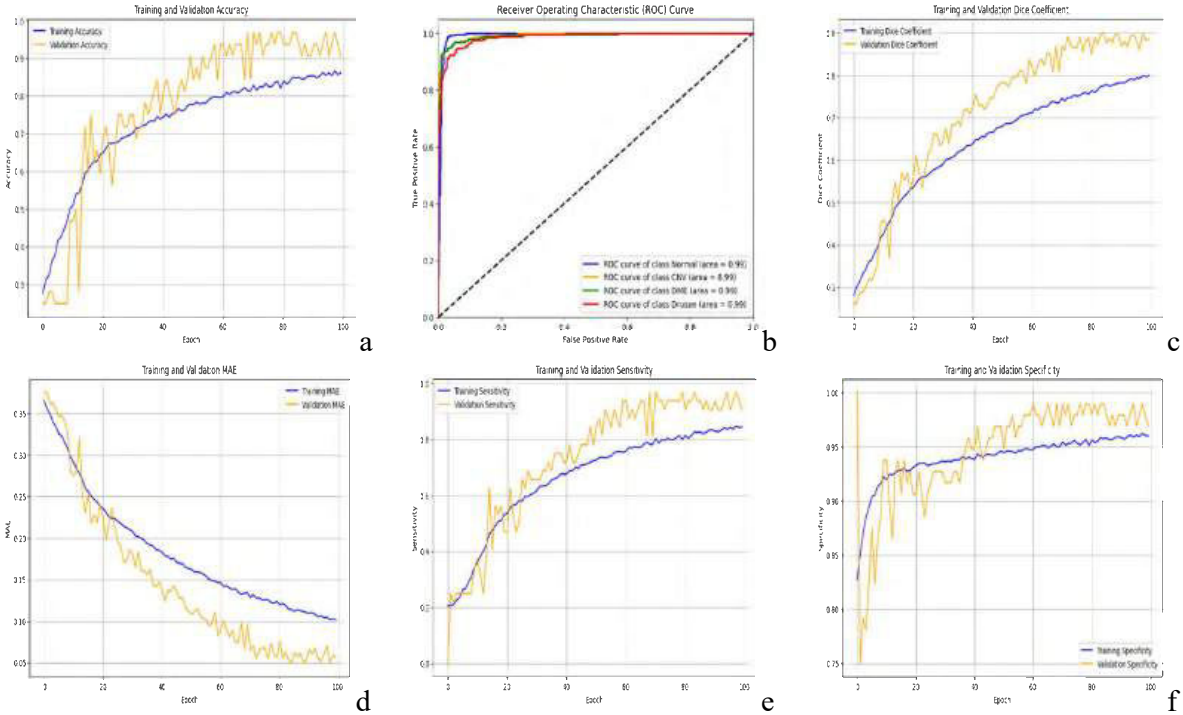


Figure 16: Final AUC-ROC Curve of the model(b); Epoch wise trend of evaluation metrics (a) Accuracy (c) Dice coefficient (d) Mean Absolute Error (e) Sensitivity (f) Specificity

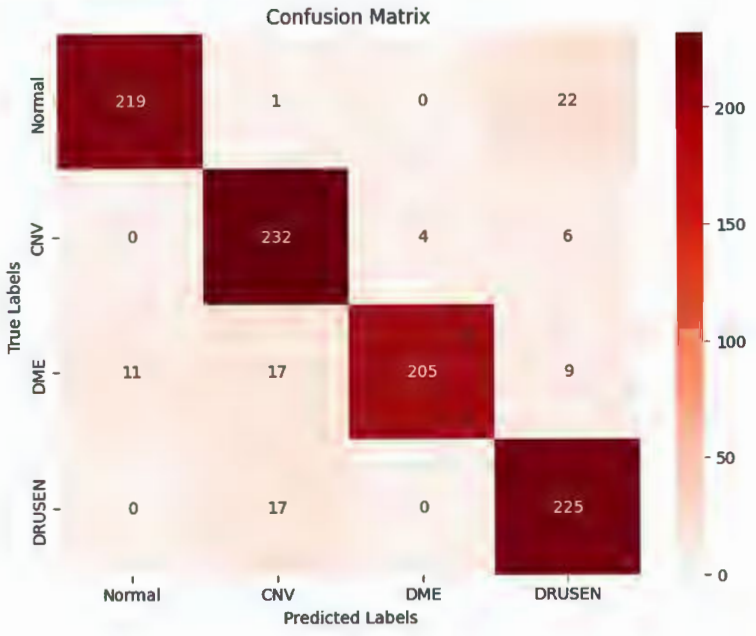


Figure 17: Confusion matrix presenting the complete picture of predictive capabilities of the model

4.3.3. Case-3: Nadam

In the third case study, the proposed DA-CNN method has been compiled using the Nadam optimizer that is an extension of the Adam optimizer that incorporates Nesterov momentum into the adaptive learning rate scheme, as discussed in detail in section 5.4. The optimal hyperparameter stood out at batch-size of 32, learning-rate of 0.001 for 100 iterations. The model's accuracy came out to be astonishing 97.4% which is the ultimate result of implementing the enhanced feature extraction techniques based on separable dynamic convolution and channel split dual attention. The model has demonstrated the bias of 0.06 and variance of 0.01. Moreover, the generalized accuracy comes out to be 0.974, test loss of 0.08, test MAE of 0.02, test dice coefficient of 0.97, test sensitivity of 0.98 and test specificity of 0.99, AUC-ROC of 0.999. The epoch wise trend of evaluation metrics on training and validation sets along with confusion matrix is presented in Figure 18**.** Furthermore, the calculated confusion matrix is displayed in Figure 19.

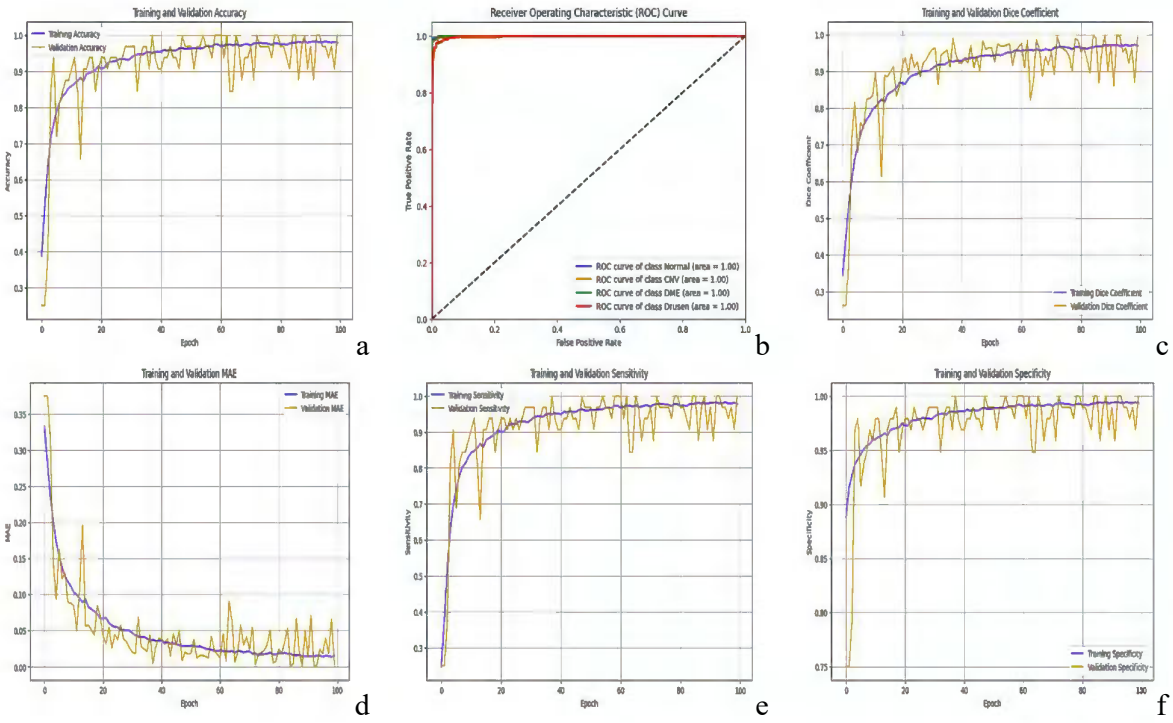


Figure 18: Final AUC-ROC Curve of the model(b); Epoch wise trend of evaluation metrics (a) Accuracy (c) Dice coefficient (d) Mean Absolute Error (e) Sensitivity (f) Specificity

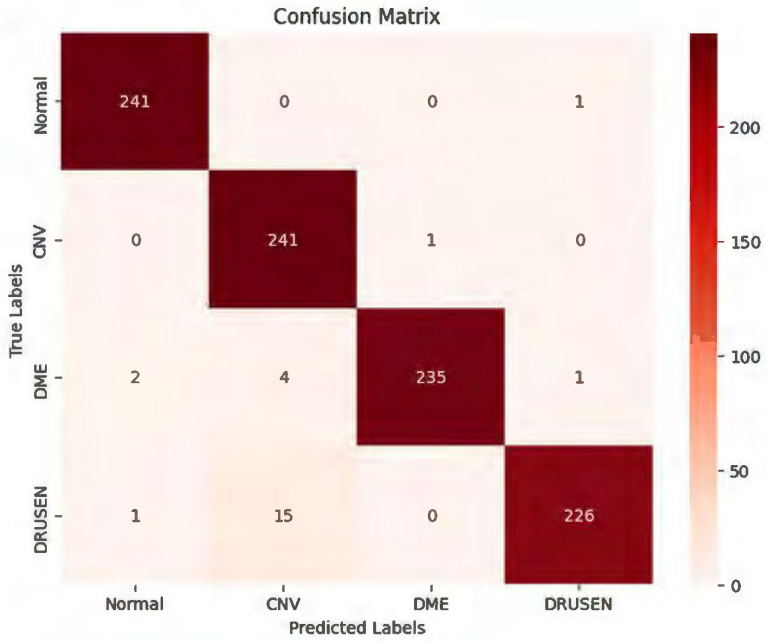


Figure 19: Confusion matrix presenting the complete picture of predictive capabilities of the model

4.3.4. Case-4: RMSprop

In the fourth case study, the proposed DA-CNN model has been compiled using the RMSprop optimizer. The optimal hyperparameter stood out at batch-size of 32, learning-rate of 0.001 for 100 iterations. The model's accuracy came out to be astonishing 89% due to implementing the enhanced feature extraction techniques based on separable dynamic convolution and channel split dual attention. The model has demonstrated the bias of approximately 0.10 and variance of 0.10. Moreover, the generalized accuracy comes out to be 0.89, test loss of 0.37, test MAE of 0.06, test dice coefficient of 0.88, test sensitivity of 0.89 and test specificity of 0.96, AUC-ROC of 0.992. The epoch wise trend of evaluation metrics on training and validation sets along with confusion matrix is presented in Figure 20. Additionally, the resulted confusion matrix is displayed in Figure 21.

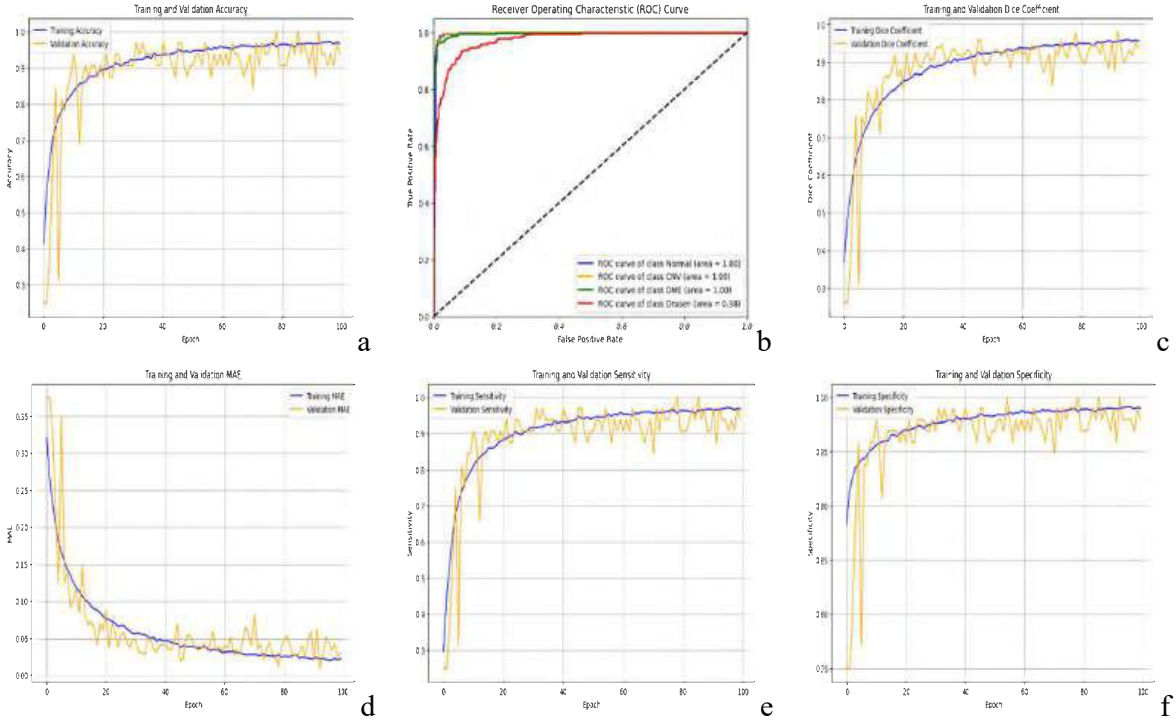


Figure 20: Final AUC-ROC Curve of the model(b); Epoch wise trend of evaluation metrics (a) Accuracy (c) Dice coefficient (d) Mean Absolute Error (e) Sensitivity (f) Specificity

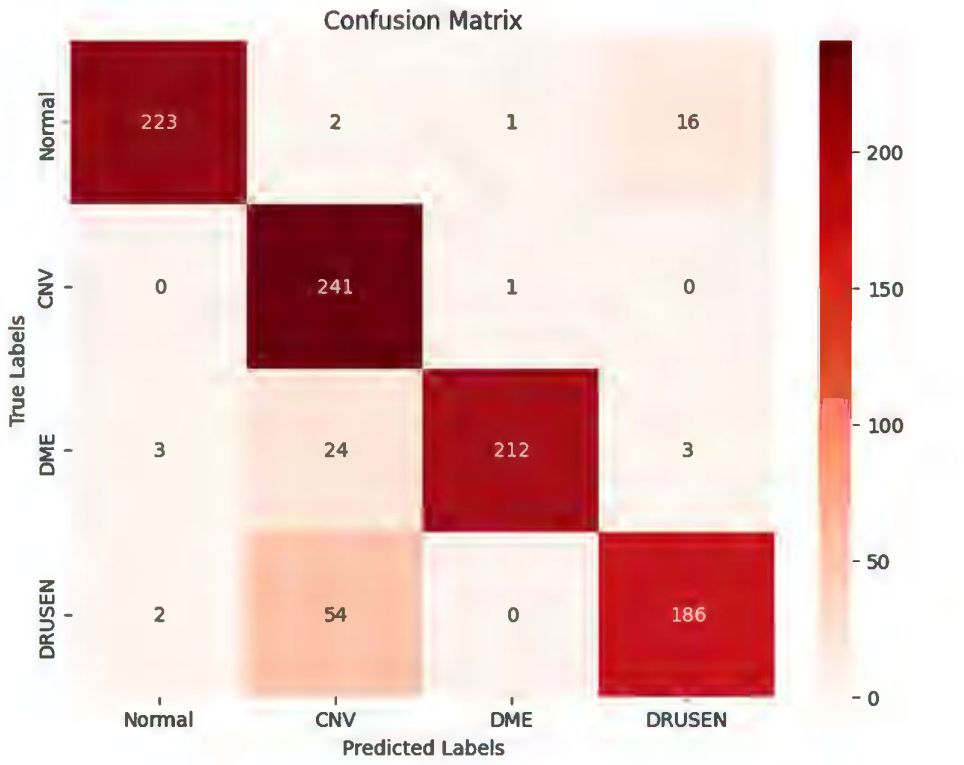


Figure 21: Confusion matrix presenting the complete picture of predictive capabilities of the model

4.4. Learning Behavior of CF-CNN Model

In this research work, the training process of the proposed CF-CNN model goes on for 100 epochs which is significantly low as compared to various state-of-the-art models discussed in 'Related Work' section (avg_epochs=350). The overall training performance is assessed by the training and validation loss, training and validation accuracy. Moreover, the key evaluation metrics such as mean absolute error (MAE) and the dice coefficient (DSC) are also calculated at each epoch to get a better understanding of training process and ultimately the model's performance. Figure 22 presents the train and validation loss whereas figure 10 shows the train and validation accuracy. We see a gradual decline in the loss epoch wise. Figure 23 presents the model's accuracy for the train and validation data. The figure exhibits minimal overfitting and a notable increase in accuracy. Particularly noteworthy is the significant increase in accuracy at epoch 15, which corresponds to a turning point in the model's learning trajectory.

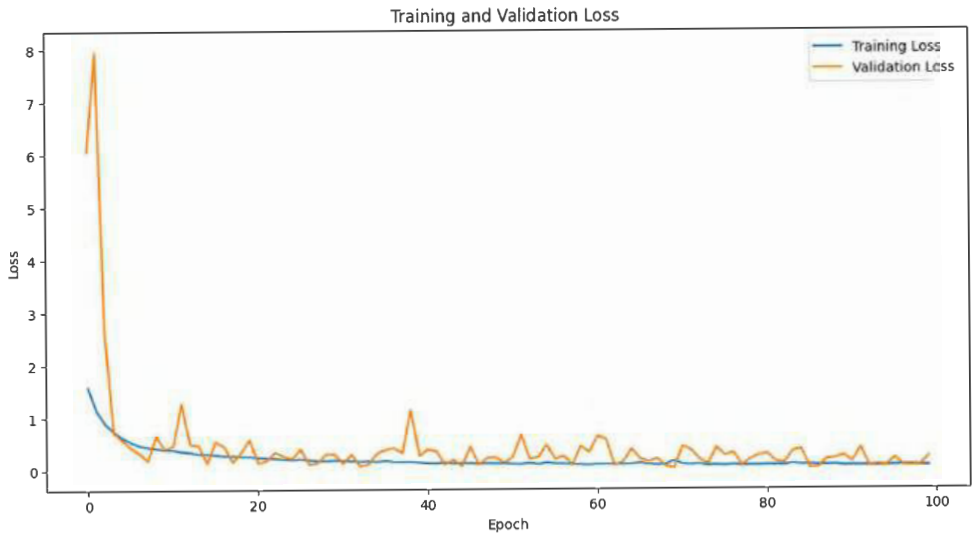


Figure 22: Training and Validation Loss

Figure 23: Training and Validation Accuracy

Figure 24 shows the plot of training and validation Dice coefficient metric over the epoch during the training phase. Overall, every key performance metric, including values of dice coefficient with peaks or plateaus, shows the segmentation of the model and identifies the critical features well throughout the training procedure.

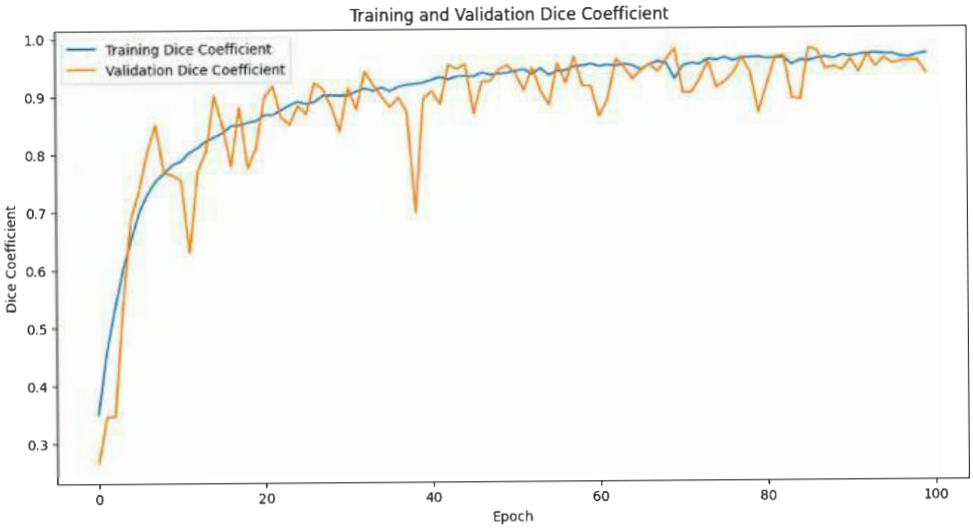


Figure 24: Training and Validation Dice Coefficient

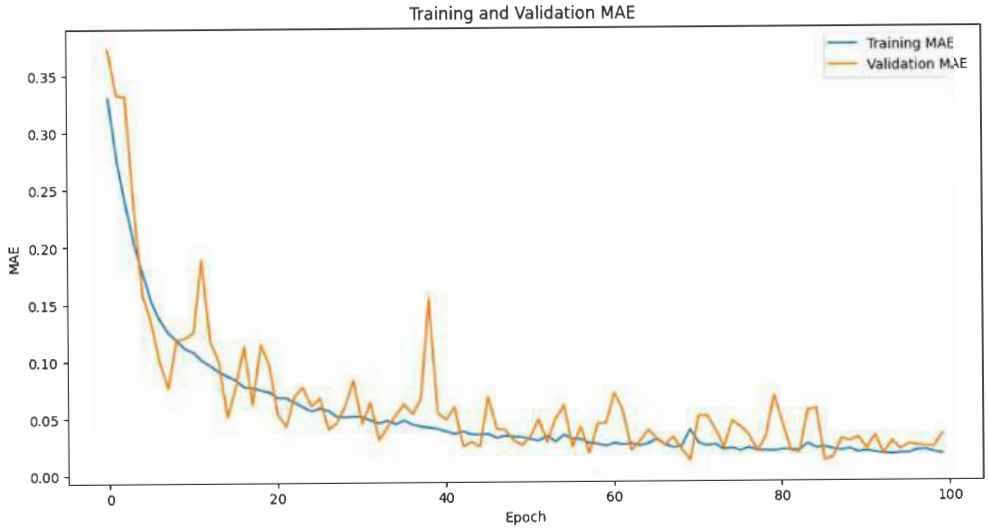


Figure 25: Training and Validation Mean Absolute Error

Figure 25 extends the analysis by presenting the Mean Absolute Error (MAE) across epochs. The model's ability to minimize average absolute discrepancies between predicted and true values is demonstrated by the steady drop in MAE. This pattern demonstrates how the model's accuracy increased during training. These figures' analysis points to a balanced model that performs well in segmentation tasks and skillfully minimizes both error measures. We investigate further possible relationships and trade-offs between accuracy, Mean Absolute Error (MAE), and Dice coefficient, offering valuable information on the complex functionality of our suggested design.

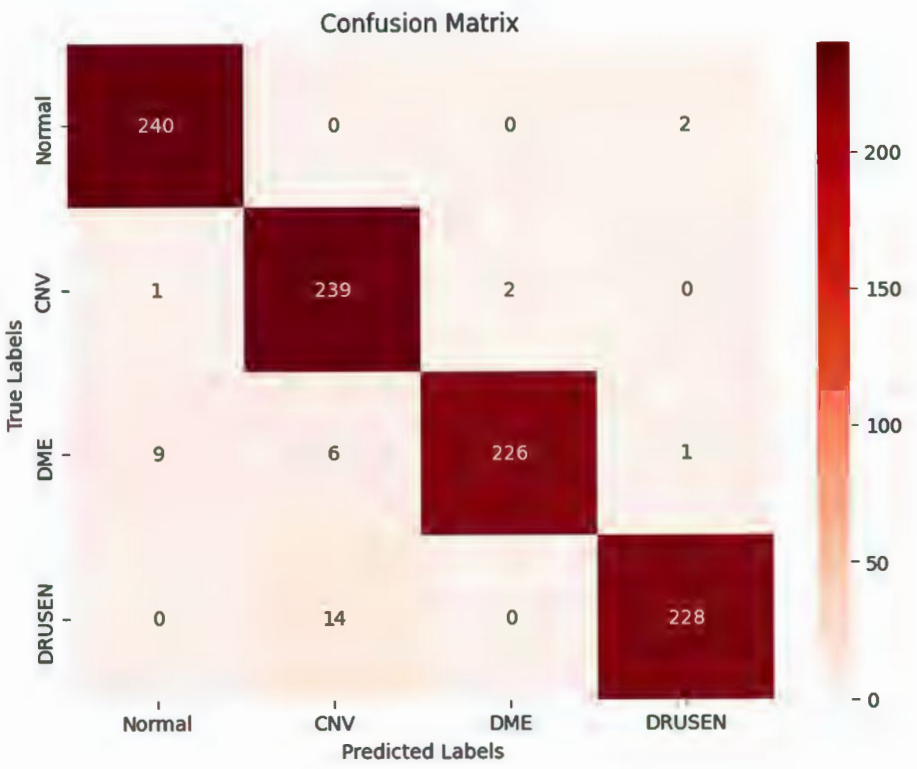


Figure 26: Confusion Matrix

4.5. Integrating the interpretability of the model using XAI

This section discusses the incorporation of explainable Artificial Intelligence (XAI) techniques specifically Grad-CAM (Gradient-weighted Class Activation Mapping) and LIME (Local Interpretable Model-agnostic Explanations) into our proposed CF-CNN model design with a focus on increasing interpretability of the model. The concept of XAI contributes to research particularly in certain fields of study such as medical imaging because the transparency of decision-making systems makes an impact on acceptance and confidence. A tool – XAI augments the cognitive abilities of humanity by showing how complex the decisions models have to make are. The introduction presents a detailed on XAI approaches and leads to the point that it is necessary to improve models' transparency. There are very important to get rid of the intrinsic ambiguity of complex neural networks.

It becomes clear that the last Conv Attention layer is essential to maintaining the model's interpretability. All of the crucial features needed for precise classification are retrieved at this layer, providing a clear understanding of the methods used to make the model's conclusions intelligible to end users.

Grad-CAM distinguishes itself in particular by emphasizing important areas in medical pictures, giving end users insight into the particular characteristics that affect the model's forecasts. Furthermore, LIME provides insights at the local level to supplement Grad-CAM with its model-agnostic methodology. The detected patterns, revelations, or anomalies revealed by these interpretability techniques are the main topics of discussion. To complement and illustrate our findings, we have provided visualizations such as Grad-CAM heatmaps and LIME perturbation explanations. Various highlight the interpretative value of these XAI techniques in improving our comprehension of the model's decision-making processes.

4.5.1. XAI techniques incorporated with CF-CNN model

Following figures shows the interpretability of CF-CNN model. In Figure 27, it can be seen quite clearly that the true label of image is DRUSEN, which is condition in which lipid and protein deposits occur under the retina, which are highlighted in the heatmap and then superimposed on to the original image so that the interpretability is made. The model has learned this feature and made the decision based on the selected region in the heatmap. Figure 28 shows the model’s interpretability through XAI technique GradCAM on sample image of Diabetic Macular Edema. As we know that DME is manifested as retinal thickening caused by the accumulation of intraretinal fluid primarily in the inner and outer plexiform layers. It can be seen quite clearly on heat map that the feature that made the model to classify this image as DME in the intraretinal fluid that has been accumulated in the layer of retina. Hence, the XAI feature of the model is performing and it can be interpreted that the model has learned the features quite well.

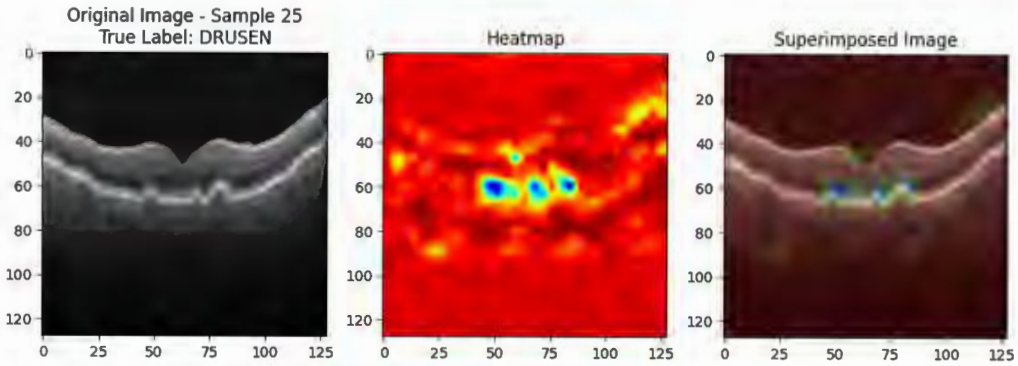


Figure 27: Model's interpretability through XAI on sample image of Drusen

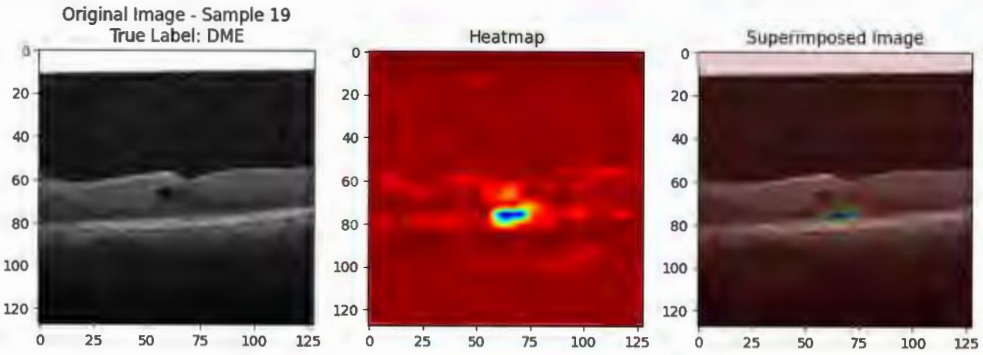


Figure 28: Model's Interpretability through XAI on sample image of DME

As we know from the introduction section that CNV is a component of the exudative age-related macular degeneration (AMD) spectrum, which is characterized by aberrant vessel growth over the Bruch's membrane from the choroidal vasculature to the neurosensory retina. Figure 29 presents the XAI's interpretation of sample image of choroidal neovascularization (CNV) condition which demonstrates through the heatmap clearly, the unnatural development of blood vessels beneath the retina. Similarly, the XAI technique LIME marks the boundaries in the region where the problem lies, giving the treating physician or surgeon a better understanding of the model's prediction, which addresses the purpose of integrating the XAI techniques in precision medicine, precisely the artificial intelligence systems in the field of medical imaging. LIME interpretations are shown in Figure 30.

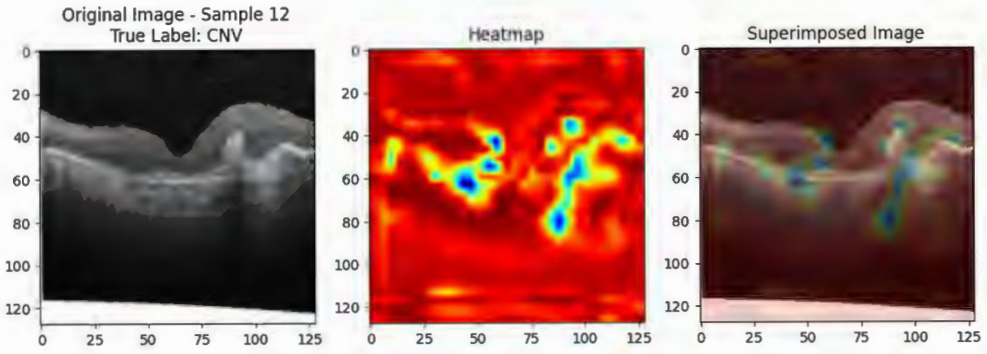


Figure 29: Model's interpretability through XAI on sample image of CNV

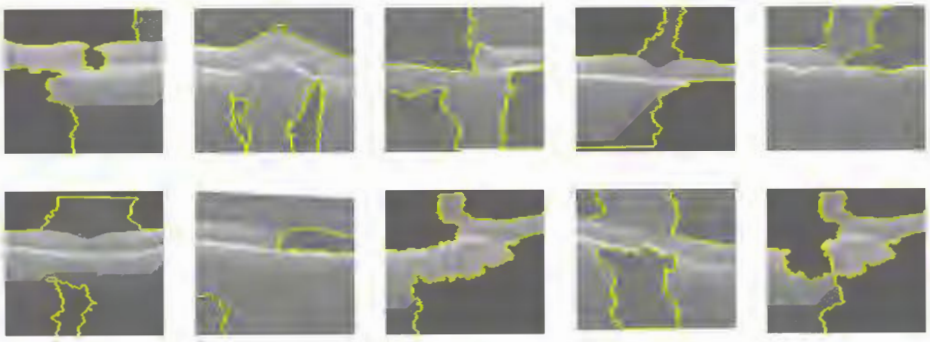


Figure 30: LIME interpretation of the proposed CF-CNN model

4.5.2. XAI techniques incorporated with DA-CNN model

To complement and illustrate our findings, we have provided visualizations based on Grad-CAM heatmaps. To implement Grad-CAM (Gradient-weighted Class Activation Mapping) for explainable AI (XAI) technique, the following stages are involved in Grad-CAM explanations: 1. Loading the trained model. 2. Getting the output of the last convolutional layer and the predicted class score. 3. Computing the gradient of the predicted class score with respect to the output feature map. 4. Computing the importance weights by averaging the gradients. 5. Generating the heatmap by multiplying the importance weights with the output feature map.6. Visualizing the heatmap overlaid on the input image. For the visualizations, we will consider the model that is compiled using Nadam optimizer. The XAI interpretations are shown in Figure [31-34].

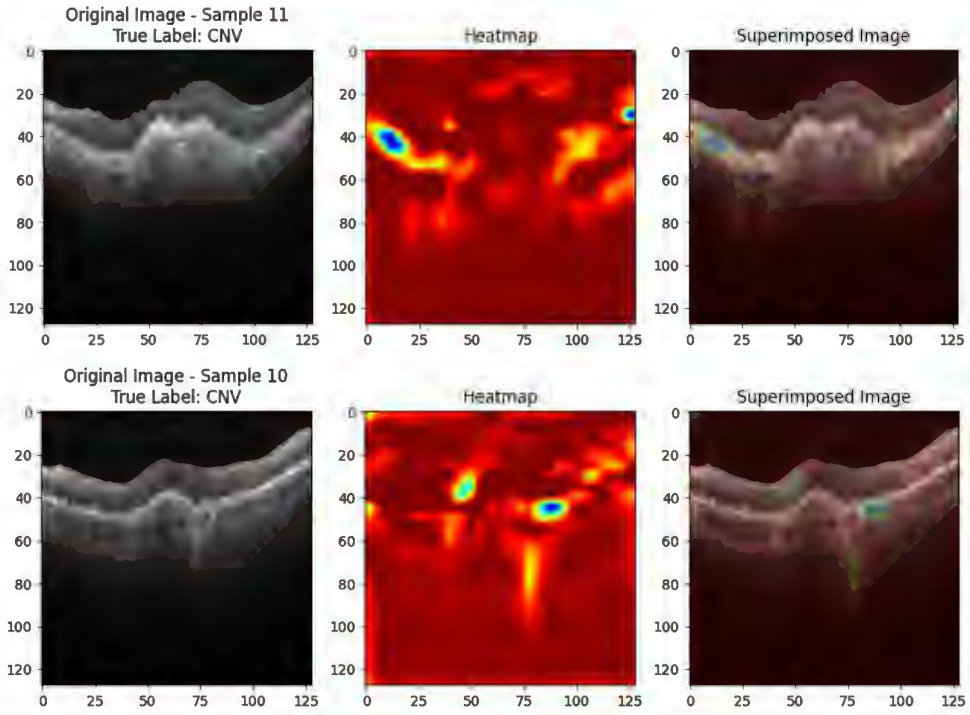
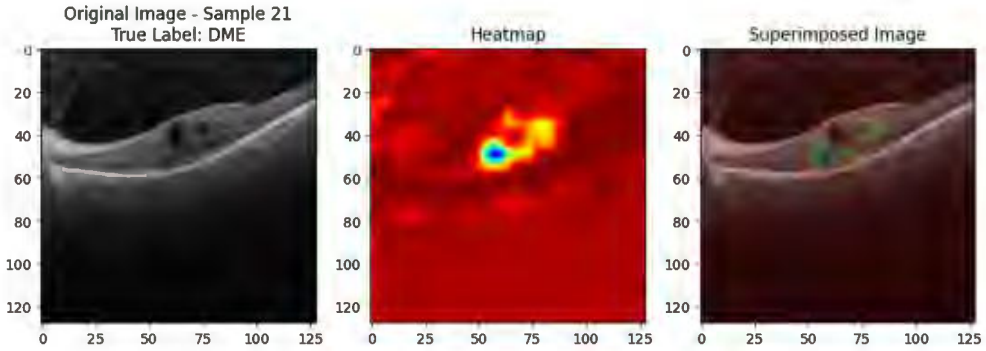


Figure 31: XAI interpretation of model on the class CNV

In Figure 31, it can be observed quite clearly that the model GradCAM technique highlights the regions where the problem lies. We are certain that the CNV condition arises due to irregular development of blood vessels beneath the retina. The OCT scan has the features that show the irregular evolution of blood vessels underneath the retina, hence captured by the model, learned and presented in the output. Heatmap is separately generated and then superimposed on to the original image pointing towards the actual spots where the blood vessels are grown.



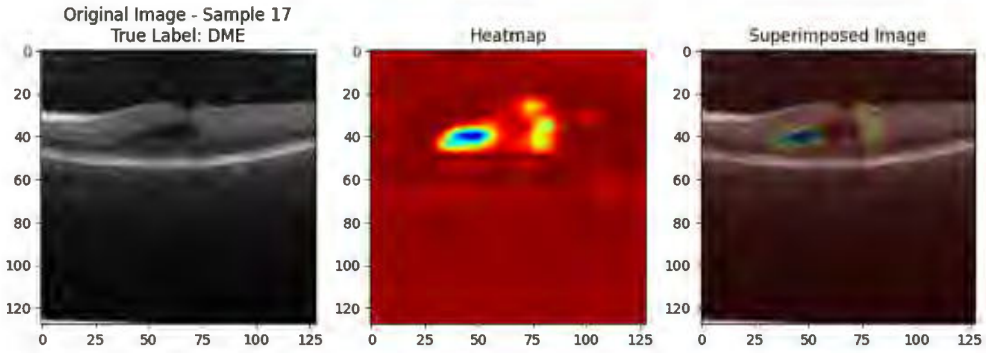


Figure 32: XAI interpretation of the model on the class DME

In Figure 32, the interpretations of GradCAM technique on DME class is presented. The region where the edema (accumulation of fluid) lies can be observed quite clearly which is the basic cause of the condition i.e., Diabetic Macular Edema. The OCT scan has the features that show the edema or the accumulation of sub retinal fluid underneath the retina, hence captured by the model, learned and presented in the output. Heatmap is separately generated and then superimposed on to the original image pointing towards the actual spots where the edema is located.

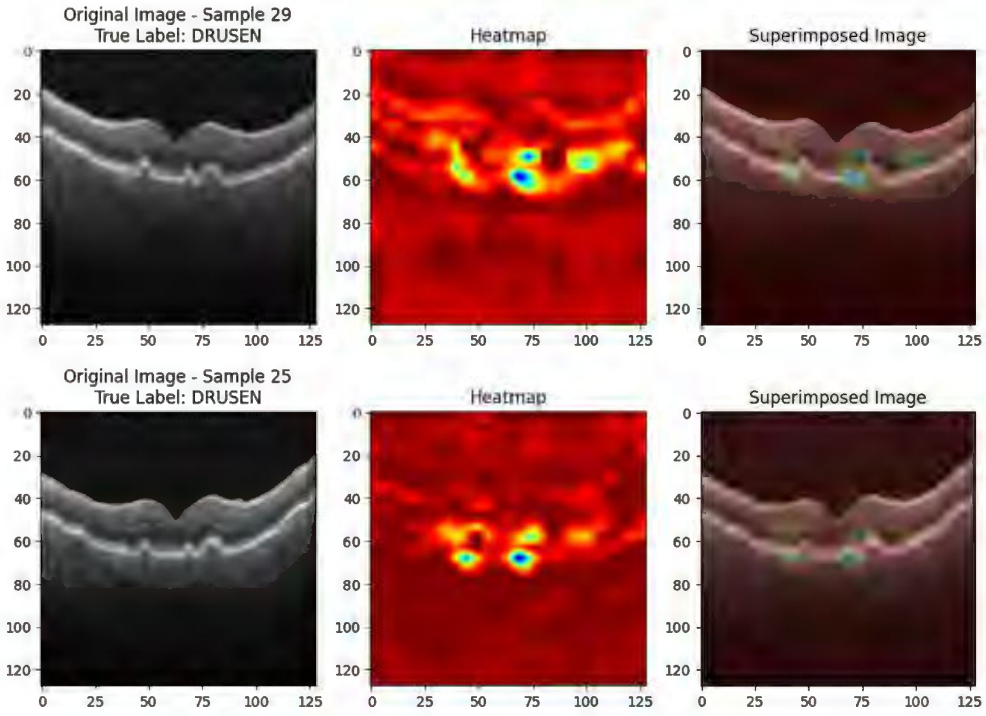


Figure 33: XAI interpretation of the model on the Drusen class

Drusen, we already know is the condition in which the lipids or fats are deposited underneath the retina hence resulting in AMD. In Figure 33, the interpretation of the model generated through XAI technique GradCAM is presented in which the retinal region where the lipid or fat deposits are located is highlighted. Ultimately, it assists the physician or the surgeon to trust the decision made by the CNN model hence eliminating any ambiguity. As the normal retinal layer must not contain any spot that is caused due to various conditions like CNV, DME, AMD or Drusen, the interpretation of the model that is presented in Figure 34 is clearly highlighting the whole retinal layer, demonstrating that there is no problem what so ever with the retina, hence the decision of the model to predict this image as Normal. Heatmap is separately generated again and then superimposed on to the original image pointing towards the whole retinal layer.

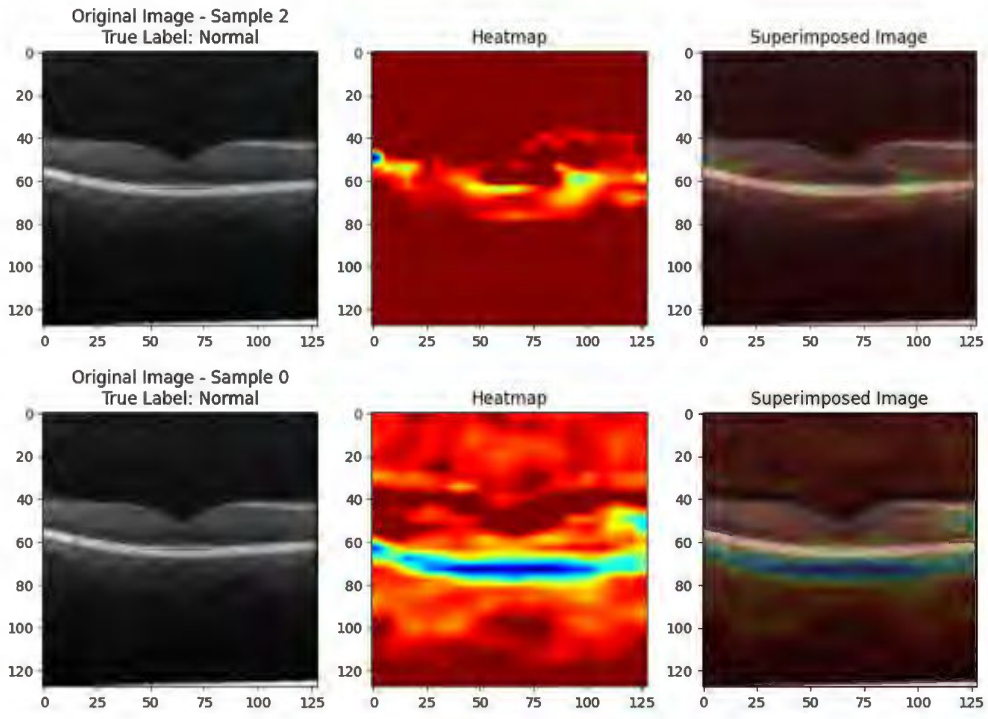
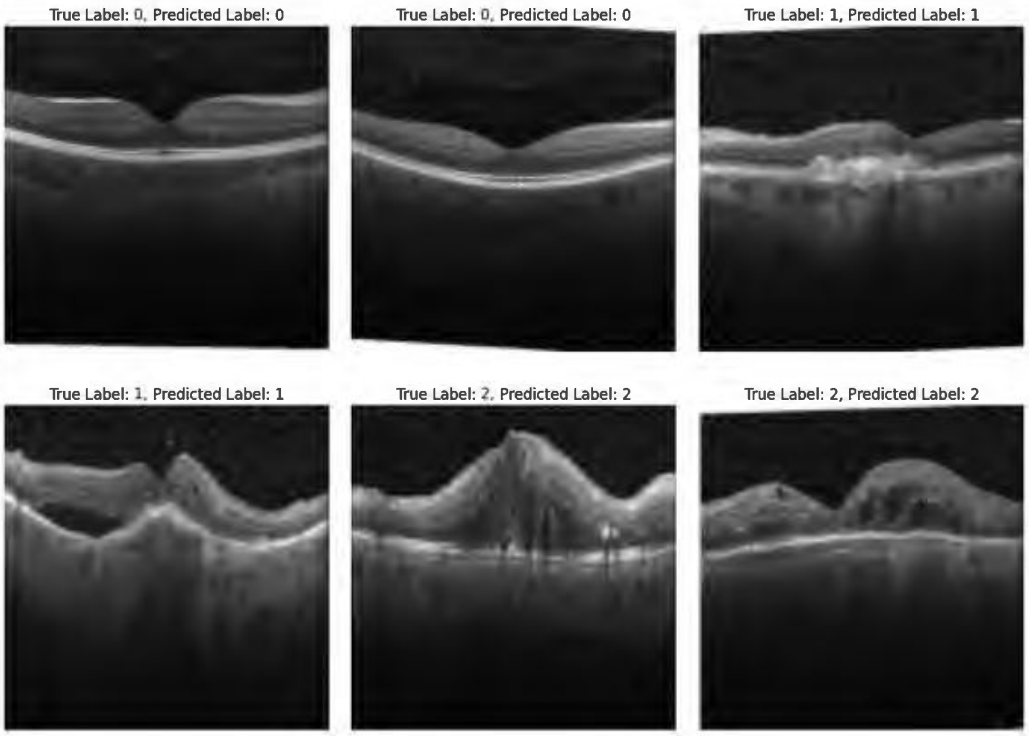


Figure 34: XAI interpretation of the model on the Normal class

4.6. Visualizing the predictions

This section analyses the predictive capacity of the proposed models, analyzing their performance based on important metrics and practical scenarios. The main objective is to get the understanding of the generalization ability of the models to real life data on which the model has not been trained and also to observe how much correctness is achieved by the model. In order to achieve accurate predictions in a variety of scenarios, this requires a thorough study that goes beyond simple statistics. This analysis will reveal the model's stability and consistency. The predictive capability of our models is tested in real-world settings in addition to numerical data. That is, testing the models with new and never-tried images to evaluate its adaptability for real application. Figure 35 demonstrate our models' predictions following training and validation.



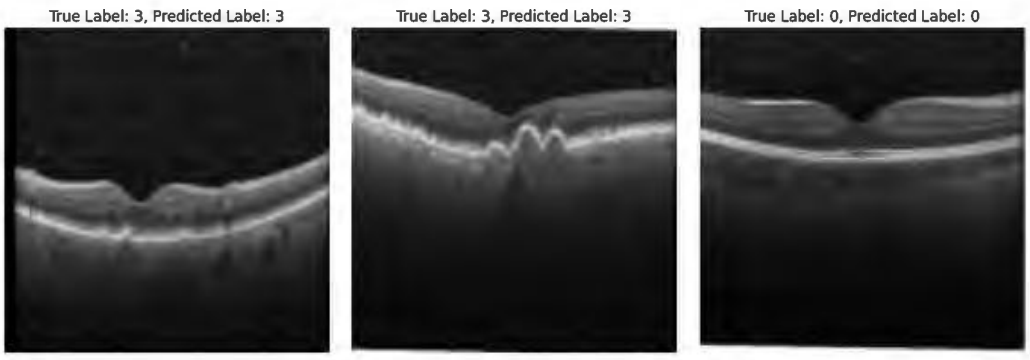


Figure 35: Generalized predictions obtained on test data

4.7. Summary

This chapter presents the detailed results along with the analyses of the results that have been obtained after extensive experimentation. Moreover, the visualizations of the predictions that the model has done are also presented alongside the visualizations of the Explainable Artificial Intelligence techniques GradCam and LIME.

CHAPTER 5

CONCLUSIONS AND FUTURE WORK

5.1. Introduction

In the concluding chapter of the research, we review the main findings of the study, stimulate a critical evaluation of the results obtained and describe the limitations that were observed throughout the research. The conclusions section gives an understanding of usefulness and significance of the proposed methodologies and the final section of the discussion offers an assessment of general significance of these findings. Furthermore, we suggest future research agendas pointing to areas of enhancement and exploration which the present study could build upon.

5.2. Discussions

By the analysis of the results presented in preceding chapter, it can be observed in terms of bias and variance that the DA-CNN model is trained with low bias and low variance with each optimizer, having slightly higher variance when compiled with generalized fractional steepest descent optimizer but not so high to consider it, hence all the optimizers result in best fit models.

There is no issue of underfitting or overfitting as per the analysis of bias and variance values. Now, to make the choice of the best performing model, we are presented with various other evaluation metrics. If we have a look at the model in terms of accuracy, the model shows the remarkable 97.4% accuracy when compiled with Nadam optimizer which is the best among all four, test loss of 0.08 which is lowest among all, MAE of 0.02 which is least among all, dice coefficient value of 0.97 which is best among all and shows 97% overlap hence the best segmentation between the sets, sensitivity of 0.98 which also is better than other 3 and shows the correct rate of predictions on positive cases, and specificity of 0.99 which is again the best among all four and shows the correct rate of predictions on negative cases and finally the AUC-ROC score of 0.999 which is the most among all the cases under observation.

In a nutshell, based on bias and variance, all the models are best fit with slight difference between the values of bias and variance but in the terms of other benchmark evaluation metrics that are exploited to get the comprehensive assessment of the model, the model stands out on the performance when it is compiled by the Nadam optimizer, hence we will consider the model compiled with Nadam as the go-to or first choice model. Moreover, the most can also be considered as scalable and robust to the inherent noise because the test data set that has been employed to get the results contains variations of noisy images as well and it performs equally good on unseen data. Table 4, enlists the detailed performance comparison of proposed DA-CNN model with specified metrics and Table 5 presents the classification report of CF-CNN model. The graphical illustration of the summarized evaluation metrics is revealed in Figure 36.

Table 4: Summarized performance metrics based of four optimizers on DA-CNN model

Optimizer	**Accuracy**	**Sensitivity**	**Specificity**	**AUCROC**	**Dice**	**MAE**	**Bias**	**Variance**
Adam	0.91	0.90	0.97	0.991	0.84	0.08	0.05	0.10
Adafactor	0.91	0.90	0.97	0.991	0.85	0.08	0.36	0.21
Nadam	0.97	0.98	0.99	0.999	0.97	0.02	0.06	0.01
RMSprop	0.89	0.89	0.96	0.992	0.88	0.06	0.10	0.10

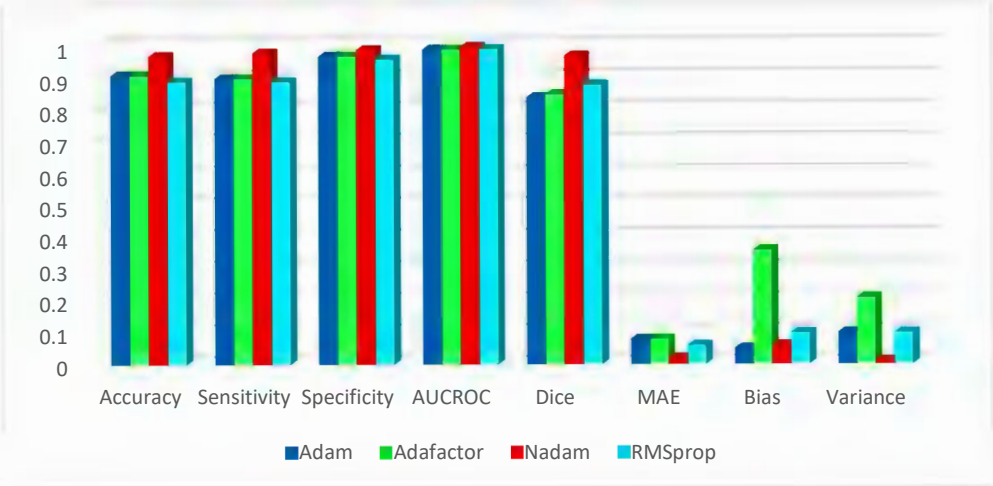


Figure 36: Graphical representation of the summary and comparison of performance metrics of DA-CNN model

Table 5: Classification report of CF-CNN model

Classification Report	**Precision**	**Recall**	**F1-Score**	**Support**
Normal	0.96	0.99	0.98	242
CNV	0.92	0.99	0.95	242
DME	0.99	0.93	0.96	242
Drusen	0.99	0.94	0.96	242
Accuracy			0.96	968
Macro avg	0.97	0.96	0.96	968
Weighted avg	0.97	0.96	0.96	968

5.3. Comparison with state-of-the-art models

5.3.1. Comparison of DA-CNN model

Table 6, shows the performance comparison of proposed DA-CNN model with benchmark methods on accuracy. The graphical comparison of CNN with state-of-the-art models can been seen in Figure 37. The remarkable gains are achieved in terms of accuracy, Mean Absolute Error and the Dice coefficient, Sensitivity and Specificity. From performance comparison tables and graphs, it is seen that proposed DA-CNN model has outperformed the counterparts in accurate and efficient classification of OCT images. This displays the efficiency of the proposed DA-CNN model with respect to accurate prediction of eye condition i.e., Normal, CNV, DME or Drusen.

Table 6: Performance Comparison of proposed DA-CNN model with existing benchmark models

Reference	**Model/Methodology**	**Classes**	**Accuracy**
[47]	Convolutional Neural Network (CNN)	2 and 3	91.77
[34]	IF-CNN	4	87.3
[46]	Transfer learning with Inception ResNet V2	4	86
[54]	CNN	2	93
[55]	DenseNet	4	88
[55]	ResNet50	4	89
Proposed	DA-CNN	4	97.4

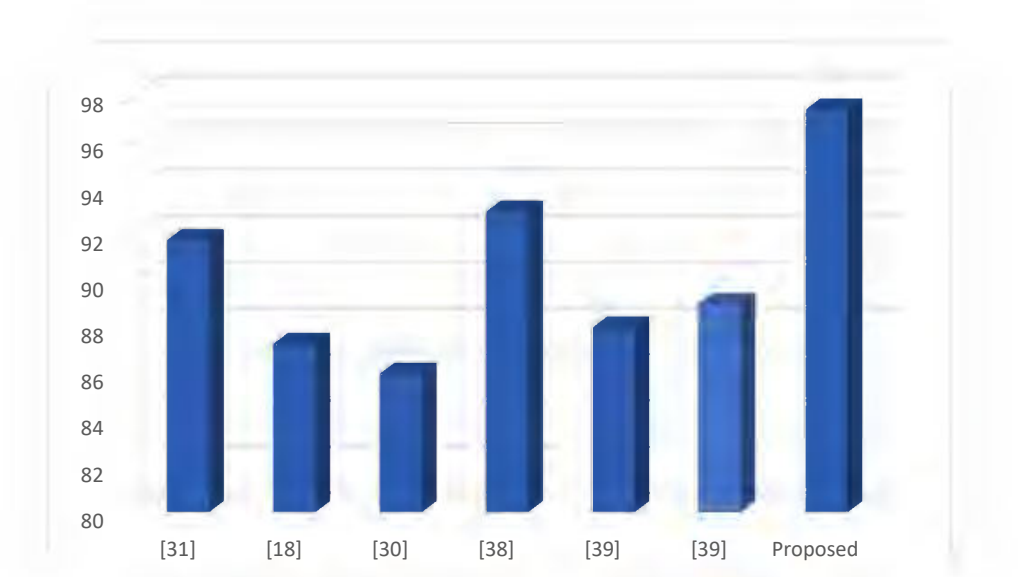


Figure 37: Graphical representation of comparison of proposed model with existing models

5.3.2. Comparison of CF-CNN model

Table 7 shows the performance comparison of proposed CF-CNN model with state-of-the-arts (SOTA) models on accuracy. The graphical comparison of CNN with state-of-the-art models can been seen in Figure 38. The remarkable gains are achieved in terms of accuracy, Mean Absolute Error and the Dice coefficient. From performance comparison tables and graphs, it is seen that proposed CF-CNN model has outperformed the counterparts in accurate and efficient classification of OCT images. This shows the effectiveness of the proposed CF-CNN model in terms of accurate prediction of eye condition i.e., Normal, CNV, DME or Drusen.

Table 7: Comparison of Accuracy of CF-CNN model with standard counterparts

Reference	**Model/Methodology**	**Classes**	**Accuracy**
[47]	Convolutional Neural Network (CNN)	2 and 3	91.77
[34]	IF-CNN	4	87.3
[46]	Transfer learning with Inception ResNet V2	4	86
[54]	CNN	2	93
[55]	DenseNet	4	88
[55]	ResNet50	4	89
Proposed	CF-CNN	4	96.88

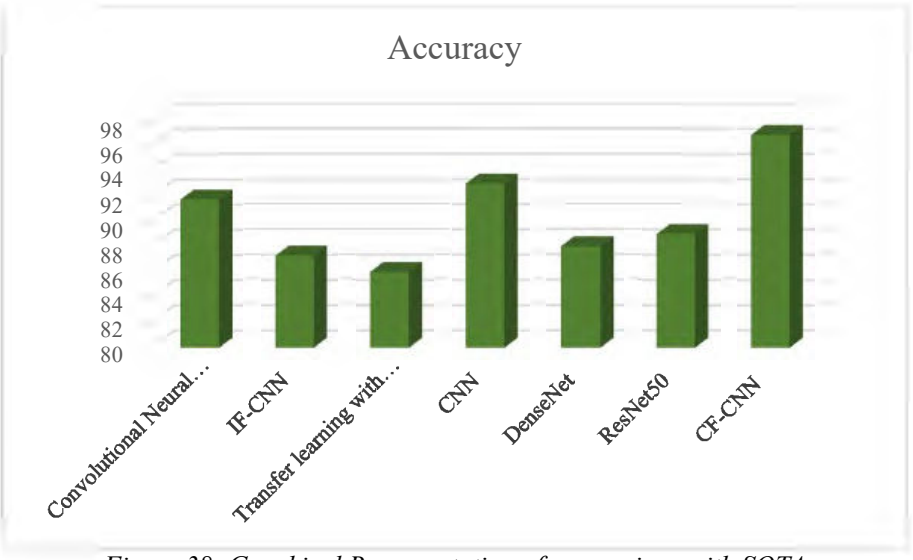


Figure 38: Graphical Representation of comparison with SOTA

5.4. Conclusions

- In this research work, two customized CNN based architectures to accurately distinguish the OCT scans into four different classes i.e., Drusen, CNV, Normal and DME are proposed.
- Latest features and techniques are exploited such as separable dynamic convolution, dynamic convolution, spatial attention and channel split dual attention for the enhanced feature extraction. This permits the model to acquire the features accurately through the OCT images which is a non-invasive method to get the image of internal structure of the eye.
- One of the models i.e. DA-CNN is then compiled using four different optimizers for error back propagation that are Adam, Adafactor, Nadam and RMS prop.
- By the comprehensive analysis of the results and comparison based on various evaluation metrics, we conclude that the DA-CNN model performs best when compiled with Nadam optimizer hence we consider that particular approach as best-fit.

- Interpretability of the model is integrated based on GradCAM XAI technique in which the model highlights the region or the features on the OCT images based on which the classification decision is matured.
- The model outperforms various state-of-the-arts on different evaluation metrics and achieves the optimal accuracy in quite a smaller number of iterations.

5.5. Future Work

As for future work, we have listed the performance improvement and the generalization of these CNN architectures as a promising direction.

- The model could be more generalizable if a larger and a more diverse dataset were collected from different population, aging, and diseases using different types of OCT images.
- The presence of such diversity should further add to the dataset's capacity to arrest the model's degeneration, creating a more fortified immunity in confronting a wider range of patients.
- Furthermore, using actual OCT images intensively processed with advanced data augmentation strategies that can be developed for OCT images, may further enhance its ability to detect even submicroscopic changes within the eye and enhance its diagnostic abilities even in the most complex cases.
- Another aspect of future work is that the current study can be extended by considering other forms of interpretability in addition to GradCAM, thereby gaining a better understanding of the predictions of the developed model.
- Other techniques, like LRP or SHAP, that provide more refined information that this paper provided may give the doctors a better non-simplified understanding of how the model works.
- Finally, integrating this CNN model into a real-time OCT image analysis platform or embedding it into portable diagnostic tools would be another step to democratizing diagnostic eye health assessment in point-of-care analyses.

5.6. Summary

This chapter presents the discussions on the results obtained and the comparison of the achieved values of the evaluation metrics with the standard counterparts and the models and methodologies mentioned in the literature along with the conclusions and the future directions of the research.

REFERENCES

[1] J. Yanase and E. Triantaphyllou, "A systematic survey of computer-aided diagnosis in medicine: Past and present developments," *Expert Syst Appl*, vol. 138, Dec. 2019, doi: 10.1016/J.ESWA.2019.112821.

[2] S. Kaur *et al.*, "Medical Diagnostic Systems Using Artificial Intelligence (AI) Algorithms: Principles and Perspectives," *IEEE Access*, vol. 8, pp. 228049–228069, 2020, doi: 10.1109/ACCESS.2020.3042273.

[3] M. Rezaei *et al.*, "Role of Artificial Intelligence in the Diagnosis and Treatment of Diseases," *Kindle*, vol. 3, no. 1, pp. 1–160, Jul. 2023, Accessed: Apr. 28, 2024. [Online]. Available: http://preferpub.org/index.php/kindle/article/view/Book23

[4] "Fundamentals of Medical Imaging - Paul Suetens - Google Books." Accessed: Jan. 06, 2024. [Online]. Available: https://books.google.com.pk/books?hl=en&lr=&id=U11EDgAAQBAJ&oi=fnd&pg=PA9&dq=medical+imaging&ots=vfYYP8OsTp&sig=zqz_Vv1AqR5-e0TnHf9QYRKlo8k&redir_esc=y#v=onepage&q=medical%20imaging&f=false

[5] D. Huang *et al.*, "Optical Coherence Tomography," *Science (1979)*, vol. 254, no. 5035, pp. 1178–1181, Nov. 1991, doi: 10.1126/SCIENCE.1957169.

[6] J. P. O. Li *et al.*, "Digital technology, tele-medicine and artificial intelligence in ophthalmology: A global perspective," *Prog Retin Eye Res*, vol. 82, p. 100900, May 2021, doi: 10.1016/J.PRETEYERES.2020.100900.

[7] A. Ashaye, A. J. Ajuwon, and C. Adeoti, "Perception of blindness and blinding eye conditions in rural communities.," *J Natl Med Assoc*, vol. 98, no. 6, p. 887, Jun. 2006, Accessed: Jan. 06, 2024. [Online]. Available: /pmc/articles/PMC2569369/?report=abstract

[8] H. E. Grossniklaus and W. R. Green, "Choroidal neovascularization," *Am J Ophthalmol*, vol. 137, no. 3, pp. 496–503, Mar. 2004, doi: 10.1016/J.AJO.2003.09.042.

[9] G. E. Lang, “Diabetic Macular Edema,” *Ophthalmologica*, vol. 227, no. Suppl. 1, pp. 21–29, Apr. 2012, doi: 10.1159/000337156.

[10] X. Zhang and S. Sivaprasad, “Drusen and pachydrusen: the definition, pathogenesis, and clinical significance,” *Eye 2020 35:1*, vol. 35, no. 1, pp. 121–133, Nov. 2020, doi: 10.1038/s41433-020-01265-4.

[11] N. J. Y. Yeo, E. J. J. Chan, and C. Cheung, “Choroidal neovascularization: Mechanisms of endothelial dysfunction,” *Front Pharmacol*, vol. 10, 2019, doi: 10.3389/FPHAR.2019.01363/FULL.

[12] “Diabetes-Related Macular Edema (DME): Symptoms & Treatment.” Accessed: Oct. 28, 2024. [Online]. Available: https://my.clevelandclinic.org/health/diseases/24733-diabetes-related-macular-edema

[13] “Schematic representation of the eye with structural and cellular... | Download Scientific Diagram.” Accessed: Oct. 28, 2024. [Online]. Available: https://www.researchgate.net/figure/Schematic-representation-of-the-eye-with-structural-and-cellular-organization-of-the_fig1_367350142

[14] “BDES Study - Klein Lab.” Accessed: Oct. 28, 2024. [Online]. Available: https://klein.ophth.wisc.edu/bdes-study/

[15] O. Perdomo, S. Otalora, F. A. Gonzalez, F. Meriaudeau, and H. Muller, “OCT-NET: A convolutional network for automatic classification of normal and diabetic macular edema using sd-oct volumes,” *Proceedings - International Symposium on Biomedical Imaging*, vol. 2018-April, pp. 1423–1426, May 2018, doi: 10.1109/ISBI.2018.8363839.

[16] J. De Fauw *et al.*, “Clinically applicable deep learning for diagnosis and referral in retinal disease,” *Nature Medicine 2018 24:9*, vol. 24, no. 9, pp. 1342–1350, Aug. 2018, doi: 10.1038/s41591-018-0107-6.

[17] L. Fang, C. Wang, S. Li, H. Rabbani, X. Chen, and Z. Liu, “Attention to lesion: Lesion-Aware convolutional neural network for retinal optical coherence tomography image classification,” *IEEE Trans Med Imaging*, vol. 38, no. 8, pp. 1959–1970, Aug. 2019, doi: 10.1109/TMI.2019.2898414.

[18] A. Alqudah, A. M. Alqudah, and M. Altantawi, “Artificial Intelligence Hybrid System for Enhancing Retinal Diseases Classification Using Automated Deep Features Extracted from OCT Images,” *International Journal of Intelligent Systems and Applications in Engineering*, vol. 9, no. 3, pp. 91–100, Sep. 2021, doi: 10.18201/ijisae.2021.236.

[19] S. A. Tuncer, A. Çinar, S. Arslan Tuncer, A. Çınar, and M. Fırat, “Hybrid CNN Based Computer-Aided Diagnosis System for Choroidal Neovascularization, Diabetic Macular Edema, Drusen Disease Detection from OCT Images”, doi: 10.18280/ts.380314.

[20] S. Sotoudeh-Paima, A. Jodeiri, F. Hajizadeh, and H. Soltanian-Zadeh, “Multi-scale convolutional neural network for automated AMD classification using retinal OCT images,” *Comput Biol Med*, vol. 144, p. 105368, May 2022, doi: 10.1016/J.COMPBIOMED.2022.105368.

[21] O. Akinniyi, M. M. Rahman, H. S. Sandhu, A. El-Baz, and F. Khalifa, “Multi-Stage Classification of Retinal OCT Using Multi-Scale Ensemble Deep Architecture,” *Bioengineering 2023, Vol. 10, Page 823*, vol. 10, no. 7, p. 823, Jul. 2023, doi: 10.3390/BIOENGINEERING10070823.

[22] W. Lu, Y. Tong, Y. Yu, Y. Xing, C. Chen, and Y. Shen, “Deep Learning-Based Automated Classification of Multi-Categorical Abnormalities From Optical Coherence Tomography Images,” *Transl Vis Sci Technol*, vol. 7, no. 6, pp. 41–41, Nov. 2018, doi: 10.1167/TVST.7.6.41.

[23] D. S. Kermany *et al.*, “Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning,” *Cell*, vol. 172, no. 5, pp. 1122-1131.e9, Feb. 2018, doi: 10.1016/J.CELL.2018.02.010.

[24] N. Saleh, M. Abdel Wahed, and A. M. Salaheldin, “Computer-aided diagnosis system for retinal disorder classification using optical coherence tomography images,” *Biomedizinische Technik*, vol. 67, no. 4, pp. 283–294, Aug. 2022, doi: 10.1515/BMT-2021-0330/MACHINEREADABLECITATION/RIS.

[25] N. Saleh, M. Abdel Wahed, and A. M. Salaheldin, “Transfer learning-based platform for detecting multi-classification retinal disorders using optical coherence tomography images,” *Int J Imaging Syst Technol*, vol. 32, no. 3, pp. 740–752, May 2022, doi: 10.1002/IMA.22673.

[26] A. M. Salaheldin, M. Abdel Wahed, and N. Saleh, “Machine Learning-Based Platform for Classification of Retinal Disorders Using Optical Coherence Tomography Images,” pp. 269–283, 2022, doi: 10.1007/978-981-19-1653-3_21.

[27] M. J. Allingham, J. A. Izatt, P. S. Mettu, S. J. Chiu, S. W. Cousins, and S. Farsiu, “Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema,” *Biomedical Optics Express, Vol. 6, Issue 4, pp. 1172-1194*, vol. 6, no. 4, pp. 1172–1194, Apr. 2015, doi: 10.1364/BOE.6.001172.

[28] D. Lu *et al.*, “Deep-learning based multiclass retinal fluid segmentation and detection in optical coherence tomography images using a fully convolutional neural network,” *Med Image Anal*, vol. 54, pp. 100–110, May 2019, doi: 10.1016/J.MEDIA.2019.02.011.

[29] R. Tennakoon, A. K. Gostar, R. Hoseinnezhad, and A. Bab-Hadiashar, “Retinal fluid segmentation in OCT images using adversarial loss based convolutional neural networks,” *Proceedings - International Symposium on Biomedical Imaging*, vol. 2018-April, pp. 1436–1440, May 2018, doi: 10.1109/ISBI.2018.8363842.

[30] P. K. Upadhyay, S. Rastogi, and K. V. Kumar, “Coherent convolution neural network based retinal disease detection using optical coherence tomographic images,” *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 10, pp. 9688–9695, Nov. 2022, doi: 10.1016/J.JKSUCI.2021.12.002.

[31] S.-D. Ţălu, Ş. Ţălu, S.-D. Ţălu, and Ş. Ţălu, “Use of OCT Imaging in the Diagnosis and Monitoring of Age Related Macular Degeneration,” *Age Related Macular Degeneration - The Recent Advances in Basic Research and Clinical Care*, Jan. 2012, doi: 10.5772/33410.

[32] U. Schmidt-Erfurth, S. Klimscha, S. M. Waldstein, and H. Bogunovićabstract, “A view of the current and future role of optical coherence tomography in the management of age-related macular degeneration,” 2017, doi: 10.1038/eye.2016.227.

[33] G. M. Comer *et al.*, “Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images,” *Biomedical Optics Express, Vol. 5, Issue 10, pp. 3568-3577*, vol. 5, no. 10, pp. 3568–3577, Oct. 2014, doi: 10.1364/BOE.5.003568.

[34] L. Fang, Y. Jin, L. Huang, S. Guo, G. Zhao, and X. Chen, "Iterative fusion convolutional neural networks for classification of optical coherence tomography images," *J Vis Commun Image Represent*, vol. 59, pp. 327–333, Feb. 2019, doi: 10.1016/J.JVCIR.2019.01.022.

[35] A. S. Daniel Kermany, M. Goldbaum, W. Cai, M. Anthony Lewis, H. Xia, and K. Zhang Correspondence, "Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning," *Cell*, vol. 172, pp. 1122-1131.e9, 2018, doi: 10.1016/j.cell.2018.02.010.

[36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," 2016. Accessed: Jan. 06, 2024. [Online]. Available: http://image-net.org/challenges/LSVRC/2015/

[37] L. Huang, X. He, L. Fang, H. Rabbani, and X. Chen, "Automatic Classification of Retinal Optical Coherence Tomography Images with Layer Guided Convolutional Neural Network," *IEEE Signal Process Lett*, vol. 26, no. 7, pp. 1026–1030, Jul. 2019, doi: 10.1109/LSP.2019.2917779.

[38] A. Serener and S. Serte, "Dry and wet age-related macular degeneration classification using OCT images and deep learning," *2019 Scientific Meeting on Electrical-Electronics and Biomedical Engineering and Computer Science, EBBT 2019*, Apr. 2019, doi: 10.1109/EBBT.2019.8741768.

[39] V. Das, S. Dandapat, and P. K. Bora, "Multi-scale deep feature fusion for automated classification of macular pathologies from OCT images," *Biomed Signal Process Control*, vol. 54, p. 101605, Sep. 2019, doi: 10.1016/J.BSPC.2019.101605.

[40] D. K. Hwang *et al.*, "Artificial intelligence-based decision-making for age-related macular degeneration," *Theranostics*, vol. 9, no. 1, pp. 232–245, 2019, doi: 10.7150/THNO.28447.

[41] S. Kaymak and A. Serener, "Automated age-related macular degeneration and diabetic macular edema detection on OCT images using deep learning," *Proceedings - 2018 IEEE 14th International Conference on Intelligent Computer Communication and Processing, ICCP 2018*, pp. 265–269, Oct. 2018, doi: 10.1109/ICCP.2018.8516635.

[42] C. S. Lee, D. M. Baughman, and A. Y. Lee, "Deep Learning Is Effective for Classifying Normal versus Age-Related Macular Degeneration OCT Images," *Ophthalmol Retina*, vol. 1, no. 4, pp. 322–327, Jul. 2017, doi: 10.1016/J.ORET.2016.12.009.

[43] Y. Sun, S. Li, and Z. Sun, "Fully automated macular pathology detection in retina optical coherence tomography images using sparse coding and dictionary learning," *https://doi.org/10.1117/1.JBO.22.1.016012*, vol. 22, no. 1, p. 016012, Jan. 2017, doi: 10.1117/1.JBO.22.1.016012.

[44] D. Sidibé *et al.*, "An anomaly detection approach for the identification of DME patients using spectral domain optical coherence tomography images," *Comput Methods Programs Biomed*, vol. 139, pp. 109–117, Feb. 2017, doi: 10.1016/J.CMPB.2016.11.001.

[45] O. Perdomo *et al.*, "Classification of diabetes-related retinal diseases using a deep learning approach in optical coherence tomography," *Comput Methods Programs Biomed*, vol. 178, pp. 181–189, Sep. 2019, doi: 10.1016/J.CMPB.2019.06.016.

[46] P. P. Srinivasan *et al.*, "Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images," *Biomed Opt Express*, vol. 5, no. 10, p. 3568, Oct. 2014, doi: 10.1364/BOE.5.003568.

[47] Y. Rong *et al.*, "Surrogate-assisted retinal OCT image classification based on convolutional neural networks," *IEEE J Biomed Health Inform*, vol. 23, no. 1, pp. 253–263, Jan. 2019, doi: 10.1109/JBHI.2018.2795545.

[48] M. Awais, H. Muller, T. B. Tang, and F. Meriaudeau, "Classification of SD-OCT images using a Deep learning approach," *Proceedings of the 2017 IEEE International Conference on Signal and Image Processing Applications, ICSIPA 2017*, pp. 489–492, 2017, doi: 10.1109/ICSIPA.2017.8120661.

[49] R. Rasti, H. Rabbani, A. Mehridehnavi, and F. Hajizadeh, "Macular OCT Classification Using a Multi-Scale Convolutional Neural Network Ensemble," *IEEE Trans Med Imaging*, vol. 37, no. 4, pp. 1024–1034, Apr. 2018, doi: 10.1109/TMI.2017.2780115.

[50] P. Seeböck *et al.*, “Unsupervised Identification of Disease Marker Candidates in Retinal OCT Imaging Data Index Terms-Unsupervised deep learning, anomaly detection, biomarker identification, optical coherence tomography,” *IEEE Trans Med Imaging*, vol. 38, no. 4, 2019, doi: 10.1109/TMI.2018.2877080.

[51] Y. Wang, R. Zhao, F. Zhou, Y. Zhang, and Z. Yao, “Machine learning based detection of age-related macular degeneration (AMD) and diabetic macular edema (DME) from optical coherence tomography (OCT) images,” *Biomedical Optics Express, Vol. 7, Issue 12, pp. 4928-4940*, vol. 7, no. 12, pp. 4928–4940, Dec. 2016, doi: 10.1364/BOE.7.004928.

[52] M. A. Hussain *et al.*, “Classification of healthy and diseased retina using SD-OCT imaging and Random Forest algorithm,” *PLoS One*, vol. 13, no. 6, p. e0198281, Jun. 2018, doi: 10.1371/JOURNAL.PONE.0198281.

[53] M. Haloi CTO, “Towards Ophthalmologist Level Accurate Deep Learning System for OCT Screening and Diagnosis”.

[54] O. Perdomo, S. Otalora, F. A. Gonzalez, F. Meriaudeau, and H. Muller, “OCT-NET: A convolutional network for automatic classification of normal and diabetic macular edema using sd-oct volumes,” *Proceedings - International Symposium on Biomedical Imaging*, vol. 2018-April, pp. 1423–1426, May 2018, doi: 10.1109/ISBI.2018.8363839.

[55] K. A. Nugroho, “A Comparison of Handcrafted and Deep Neural Network Feature Extraction for Classifying Optical Coherence Tomography (OCT) Images,” *2018 2nd International Conference on Informatics and Computational Sciences, ICICoS 2018*, pp. 141–146, Oct. 2018, doi: 10.1109/ICICOS.2018.8621687.

[56] G. Lemaître *et al.*, “Classification of SD-OCT Volumes Using Local Binary Patterns: Experimental Validation for DME Detection,” *J Ophthalmol*, vol. 2016, 2016, doi: 10.1155/2016/3298606.

[57] N. Tasnim, M. Hasan, and I. Islam, “Comparisonal study of Deep Learning approaches on Retinal OCT Image,” pp. 23–24, 2019, Accessed: Jan. 06, 2024. [Online]. Available: https://www.kaggle.com/paultimothymooney/kermany2018

[58] F. Li *et al.*, “Deep learning-based automated detection of retinal diseases using optical coherence tomography images,” *Biomedical Optics Express, Vol. 10, Issue 12, pp. 6204-6226*, vol. 10, no. 12, pp. 6204–6226, Dec. 2019, doi: 10.1364/BOE.10.006204.