Ocular Toxoplasmosis Classification from Low-Resource Dataset Leveraging DiNet-Fusion

by

Tasnim Ferdous Nishat
20201100
Shipon Chakraborty
20201091
Methela Fariana Nijhum
20201097

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
Brac University
December 2025

Declaration

It is hereby declared that

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

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

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

4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Tasnim Ferdous Nishat
20201100

Shipon Chakraborty
20201091

Methela Fariana Nijhum
20201097

Approval

The thesis titled "Ocular Toxoplasmosis Classification from Low-Resource Dataset Leveraging DiNet-Fusion" submitted by:

1. Tasnim Ferdous Nishat (20201100)
2. Shipon Chakraborty (20201091)
3. Methela Fariana Nijhum (20201097)

of Fall, 2025 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science in December 6, 2025.

Examining Committee:

Supervisor:
(Member)

Dr. Md. Golam Rabiul Alam
Professor
Department of Computer Science and Engineering
School of Data and Sciences, BRAC University

Co-Supervisor:
(Member)

Mr. Md. Tanzim Reza
Senior Lecturer
Department of Computer Science and Engineering
School of Data and Sciences, BRAC University

Thesis Co-ordinator:
(Member)

Dr. Md. Golam Rabiul Alam
Professor
Department of Computer Science and Engineering
School of Data and Sciences, BRAC University

Head of Department:
(Chair)

Dr. Sadia Hamid Kazi
Associate Professor & Chairperson
Department of Computer Science and Engineering
School of Data and Sciences, BRAC University

Abstract

Ocular toxoplasmosis (OT) is a vision-threatening retinal illness, and its automated detection is challenging due to the scarcity, imbalance, and heterogeneity of fundus imaging datasets. Beneath these constraints, traditional deep learning approaches are challenged with providing accurate predictions, often resulting in inconsistent, non-clinically relevant multi-class classification between disease classes. This paper introduces DiNet-Fusion, a data-efficient fusion model that integrates EfficientNet-B0 and self-supervised DINO ViT-Small to generate comprehensive local and global representations for multi-class OT classification. Following the cleansing and stratification of a low-resource clinical dataset, features from both backbones were integrated and refined by weighted training and label smoothing to address class imbalance. The findings of the model shows DiNet-Fusion as a robust and resource-efficient technology effective at facilitating OT diagnosis in low-resource dataset clinical environments. Finally, the model demonstrated robust performance, exceeding 92% accuracy on the test set, with threshold optimization enhancing the recall of the clinically significant active class from 78.57% to 92.86%. Additionally, high AUC values (0.97–0.99), consistent learning curves, and assured probability distributions further prove its dependability.

Keywords: Ocular toxoplasmosis, Fundus image, Deep learning, EfficientNet-B0, DINO, Vision transformer, Feature fusion, Medical image classification, Classification, Neural Network.

Table of Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

AI Artificial Intelligence

AUC Area Under the Curve

CNN Convolutional Neural Network

DL Deep Learning

Grad-CAM Gradient-weighted Class Activation Mapping

LIME Local Interpretable Model-Agnostic Explanations

LLM Large Language Model

MaxViT Multi-Axis Vision Transformer

ML Machine Learning

MLP Multi-Layer Perceptron

OCT Optical Coherence Tomography

OT Ocular Toxoplasmosis

PCR Polymerase Chain Reaction

ROC Receiver Operating Characteristic

SVM Support Vector Machine

Chapter 1

Introduction

1.1 Introduction

Ocular Toxoplasmosis (OT) is an infectious disease of the eye caused by the parasite *Toxoplasma gondii*, which is a type of protozoan that can only live inside cells and can infect practically all types of mammalian cells. This parasite can affect other organs of the human body along with eyes and as well as animals. OT is still one of the most common causes of posterior uveitis around the world. It often causes inflammation that comes back, chorioretinal scarring, severe vision loss, or even blindness. The parasite can remain in retinal tissue in a dormant condition and may reactivate years later when the host's immune system is weakened. Because of this potential for reactivation, it is important to get a quick and correct diagnosis to avoid permanent harm to the eyes.

Diagnosing OT is clinically difficult due to the symptom overlap with several eye disorders. Ophthalmoscopy or fundus image examinations are commonly employed to detect distinctive retinal abnormalities; nevertheless, their ability diminishes in atypical or subtle instances. Although molecular tests like polymerase chain reaction (PCR) can aid in the diagnosis, the sensitivity is variable and these tests are technically demanding and require ocular fluids for specimen collection. Moreover, experts are not always accessible in low-resource environments, leading to discrepancies in the quality of diagnosis and patient outcomes.

Deep learning (DL) based analysis of fundus pictures has shown a lot of promise for automated illness identification in the last several years. But typical convolutional neural networks (CNNs) need big, labeled datasets to operate well in general. This is not possible for rare diseases like OT, where there are not enough data points and the classes are not balanced. Because of this, traditional deep learning (DL) and machine learning (ML) models generally make biased predictions, are not very robust, and have high rates of false positives and false negatives when trained on small OT datasets.

To address these issues, this research introduces DiNet-Fusion, a fusion-driven architecture that combines the hierarchical feature extraction ability of EfficientNet-B0 with the self-supervised representation learning ability of DINO ViT-Small. EfficientNet provides strong multi-scale convolutional features that are learnt from training on the large scale ImageNet data, while DINO (self-distillation with no labels) gives rich global transformer-encoded representations that generalise well from small amounts of data. We propose to learn a joint representation by integrating these complementary representations at the feature level, with the intention of improving classification performance for low-resource OT datasets.

In this work, we aim to classify fundus images into three clinically important categories - active, inactive and healthy using a data-efficient model that has been designed to operate effectively in a real world clinical environment where both annotated datasets as well as specialist expertise are scarce. Beyond OT, this methodology shows that transformer CNN fusion models can be extended to work on other rare or underrepresented diseases as well, making possible advances in automated medical image analysis across the board.

1.2 Problem Statement

Although the symptoms of Ocular Toxoplasmosis (OT) can occasionally be confused with other eye illnesses, there are certain indicators that help us identify this rare condition. The parasite *Toxoplasma gondii* is the cause of OT. If not properly identified and treated, it can result in blindness or severe visual impairment. Determining whether a patient has OT can be extremely difficult using the traditional clinical diagnostic method only. Furthermore, unavailability of specialized ophthalmic expertise in many regions makes it more difficult to identify OT using even the traditional method. Hence, an alternative and automated diagnostic approach is required.

To find a better approach, AI or ML can be utilized. Deep Learning (DL) is widely used to evaluate fundus images in order to identify various eye disorders. DL models can be utilized with fundus images, although sometimes the results are not accurate on small datasets. Sometimes, they respond negatively to a true positive opinion and vice versa. Additionally, DL is unable to function consistently due to the limited quantity of training examples. The study we looked at demonstrated that they required a high quality dataset and that the success of DL models depended on the availability of size of datasets. For instance, models such as SVMs, decision trees, or even standard CNN-based classifiers tend to underperform due to the limited number of training samples. Additionally, OT can present with retinal features that overlap with other ocular conditions, leading to misclassification or inconsistent predictions.

In research [20], this imbalance can lead to biased model performance, favoring the majority classes, as there are 36 images in the active lesions category, whereas 188 are inactive and 133 are healthy. In order to create a larger dataset, researchers increased the number of images to 700; this is useful for training but does not entirely replace the need for more authentic and varied data. Transfer training with previously trained model data is a major component of their model. The VGG16, VGG19, and ResNet50 models for transfer learning were the primary focus of their study; however, other models that may operate more quickly and effectively were not examined. The dataset was collected from a hospital, so the data were only from a specific region, but there might be many other places from where more data can be collected for better performance with their model (Choudhury et al., 2024).

Because of these problems, there is still a big gap in creating a diagnostic model for OT that is accurate, scalable, and uses resources wisely. A strategy is needed that can get beyond the lack of labeled images, lessen the need for a lot of clinical skill, and reliably sort fundus images into clinically useful categories. Our paper addresses these challenges by presenting a sophisticated fusion-driven transformer-based model designed for low-resource contexts.

1.3 Research Objectives

When using Deep Learning or Machine Learning algorithms for toxoplasmosis classification, a low-resource dataset is insufficient. As in several cases, the sickness could not be recognized by traditional DL or ML. Traditional machine learning models, such as decision trees and support vector machines (SVMs), usually require a large amount of data in order to achieve high classification accuracy. Thus, using SVMs Models can result in similar problems. Consequently, datasets with limited resources are likewise unsuitable for decision tree models. Furthermore, classical convolutional neural networks (CNNs) often struggle to generalize effectively, leading to misclassification and generating more false positives and false negatives in low-resource scenarios. These limitations indicate the necessity of a more robust and capable method of extracting features from a small set of medical images effectively.

The objective of this study is to design a reliable and data-efficient fusion model for ocular toxoplasmosis classification when the fundus image dataset is limited. The specific objectives are:

- To address the limitations of conventional ML and CNN-based DL methods, regarding requirements for extensive annotated training datasets to exhibit reliable performance in medical image analysis.

- To employ self-supervised DINO-ViT features to obtain better global representations that will lead to better generalization in low-resource settings from fundus images.

- To combine DINO embeddings with EfficientNet-B0 based features via feature-level fusion strategy (DiNet-Fusion) for improving multi-class OT classification performance.

- Minimize the misclassification (especially, reduce false positives and false negatives) by better representing the features and adopting class-aware training tactics.

- To establish an extendable diagnostic model that can detect the active, inactive, and healthy retinal categories with limited clinical data settings.

By achieving these goals, the study shows the possibility of combining CNN and transformer representations for classifying medical images with limited resources.

The study aims to evaluate the efficiency of the proposed DiNet-Fusion architecture in classifying ocular toxoplasmosis within the limitations of a low-resource dataset, and to determine whether its fusion-based representation learning offers demonstrable advantages over traditional deep learning methods. A primary objective is to assess the degree to which self-supervised DINO features minimize misclassifications, especially false negatives within the active class, which is clinically significant while enhancing generalization across diverse fundus images. The importance of this study is that it could lead to a diagnostic solution that is both accurate and easy to get, especially in places where there are limited specialists in eye care. DiNet-Fusion shows that it can work well with very little labeled data. This makes it a potential framework not only for OT classification but also for other uses involving uncommon diseases with few medical imaging resources. The work ultimately intends to provide a scalable, clinically suitable, and resource-efficient system that enhances diagnostic reliability and diminishes reliance on expert graders in low-resource healthcare facilities.

1.4 Report Organization

The rest of the report has been organized as follows: To address the aim of our research, Chapter 2 presented several literature reviews that were conducted on the similar domain. A detailed workflow including information associated with the dataset we have used to conduct the research and; details of our proposed model and working procedure is explained thoroughly in Chapter 3 along with diagrams and tables. This is followed by Chapter 4 that provides experimental results, evaluation matrices and a comparative analysis of our work. This chapter represents how well our proposed model has worked along with evaluating factors' graphs. Additionally, it contains a discussion on the overall research. Lastly, Chapter 5 is about summarization of our research followed by limitations and future work.

Chapter 2

Literature Review

2.1 Preliminaries

There are several researches already conducted on this similar domain. In this chapter, we have discussed those research thoroughly.

2.2 Review of Existing Research

In [4], Islam et al. (2019) present a CNN-based model designed to automatically detect eight ocular diseases, such as glaucoma, cataract, myopia, age-related macular degeneration (AMD), diabetes, hypertension, and other abnormalities. The model is robust across various imaging systems because it was trained on the ODIR-5K dataset, a real-world, multi-camera dataset containing fundus images of 5,000 patients. The study aids in early and precise diagnosis, reducing the need for costly and time-consuming manual methods. To address the limitations of conventional approaches for identifying ocular diseases, the paper proposes an automated system based on deep learning to classify multiple conditions. The preprocessing steps include data augmentation (zoom, rotation, flipping, brightness adjustments) and Contrast-Limited Adaptive Histogram Equalization (CLAHE) to improve image quality. The CNN architecture employs convolutional layers with activation functions, max pooling, dropout layers, and batch normalization, and is optimized using a cross-entropy loss function. The model achieved an F-score of 85%, a Kappa score of 31%, and an AUC value of 80.5%, demonstrating strong diagnostic performance. However, the study is affected by dataset imbalance, multi-label classification challenges, and lower predictive performance in certain disease categories.

In [21], Dragun & Roszczyk (2024) described the use of Convolutional Neural Networks (CNNs) for multiclass classification of 40 ocular diseases. 3200 fundus photos of 45 disease categories with a healthy class were taken from the RFMiD dataset. They did a lot of preprocessing and augmentations on the dataset, such as rotations, shears, zooms, and translations, in order to achieve class uniformity because it was so unbalanced. This study tackles the issue of a shortage of specialists in many developing nations, which makes manual eye disease diagnosis difficult and time-consuming. The suggested solution combines transfer learning with cutting-edge convolutional neural network (CNN) architectures, like ResNet and RegNet, which are pre-trained and optimized for the multi-class task. Using various augmentation strategies, fine-tuning techniques, and optimizers (SGD, Adam, Adagrad, and RMSProp), the methodology comprised four training phases. Specifically, RegNet-Y-3.2GF emerged as the top-performing model in this study. The model achieved an accuracy of 65.5%. For some of the diseases, the performance was higher than 95%, but due to a lower number of training samples for other diseases, the accuracy became less than 75%. The study has several limitations such asthe dataset distribution, computation cost and poor awareness of rare diseases. To achieve better accuracy in the future, we will use model ensembling and augmentation techniques, along with training on larger and more balanced datasets.

In [1], Nikam & Patil (2017) presented an image processing-based method of glaucoma detection using the Cup-to-Disc Ratio (CDR) analysis. The study's dataset is DRISHTI-GS1, which consists of publicly accessible fundus images that have been preprocessed to improve contrast and converted to grayscale. The issue under discussion is the early detection of glaucoma, a condition that damages the optic nerve and causes irreversible blindness due to elevated intraocular pressure (IOP). In order to determine the CDR for a normal or glaucomatous eye, an automatic method utilizing MATLAB GUI is suggested that divides the optic disc and optic cup and determines their boundaries using threshold-based segmentation and edge detection. Image preprocessing, optic disc segmentation using thresholding, opting smooth using an ellipse-fitting algorithm, cup extraction using a similar method, and final Cup-to-Disc Ratios (CDR) using the following formula are the steps in this methodology: Cup Diameter over Disc Diameter is equal to CDR. Glaucoma is indicated by a Cup-to-Disc Ratio (CDR) greater than 0.3. With CDR values that closely matched the observations made by the ophthalmologists in the dataset, this model demonstrated accuracy. However, they also point out drawbacks like sensitivity to image noise and a lack of deep learning integration, which could result in segmentation errors. They suggest resolving these issues in future research by utilizing morphological techniques, fuzzy logic, and novel feature extraction methods.

In [25], Mulyani et al. (2024) used a Convolutional Neural Network (CNN) that has been optimized using the Adam optimizer to enhance the classification of diabetic retinopathy fundus images. This study used Messidor, which has 1200 images that are categorized into four styles: Proliferative Diabetic Retinopathy, Severe Nonproliferative Retinopathy, Moderate, and Mild. The model employs SMOTE to handle the unbalanced data and Gabor filters to extract features. Performance increases by 78% to 81% as a result of training that is more accurate and efficient thanks to the Adam optimizer. Since AUC values range from 0.93 to 1.00 for each class, evaluation metrics like the Confusion Matrix and AUC-ROC curve verify that the model's classification ability has improved. Future research should look into the use of larger datasets and more advanced optimization techniques to further increase detection accuracy, even though the study shows improved performance. Its limitations include a small dataset and the lack of hybrid optimization methods.

In [6], their analysis focuses on the automatic identification of nuclear cataracts in retinal fundus images using pertinent image processing and machine learning techniques, particularly Support Vector Machines (SVM) with various kernels. The dataset, which was gathered from GitHub and Kaggle repositories, includes 800 fundus images, 500 of which are of people with cataracts and 300 of which are normal. To prevent cataracts from impairing vision, this project tackles the odd problem of early detection. Segmentation is used to measure retinal blood vessel features in the proposed method, and then SVM (Support Vector Machine) is used for classification using linear, polynomial, and kernel (RBF) techniques. The process starts with image data preparation (filtering, resizing, and contrast enhancement), then moves on to feature extraction (blood vessel segmentation), and finally, SVM kernel classification. With an accuracy of 95.2% and sensitivity and specificity values of 99.8% and 90.5%, respectively, the RBF-based SVM outperforms all other approaches, according to the results (Behera et al., 2020). This study's shortcomings include its exclusive dependence on blood vessel properties; to improve accuracy and generalization across diverse datasets, future iterations of this methodology might integrate deep learning methods like CNNs.

In [26], Sharma et al. (2024) focus on the classification of eye diseases using color fundus images by proposing an automated diagnosis system by leveraging various Machine Learning (ML), Deep Learning (DL) and transfer learning methods. This paper encompasses the following models: Deep Convolutional Neural Networks (DCNN), Efficient-Net B4 and B7, Entropy Segmentation Survival Analysis Optimization (EsSO), Convolutional Neural Network (CNN), Gradient Boosting, Support Vector Machines (SVM), Residual Deep Neural Networks (RDNN), VGG-19 and lastly, Transfer Learning with Multi-Layer Perceptron (TL-MLP). The problem statement of this paper addresses the critical issue of accurately detecting eye diseases from fundus images. Their methodology involves training these models on the dataset and utilizing augmentation and tuning the necessary hyperparameters to achieve the best results. Following this, the accuracy, precision, recall, and F1 scores were used to evaluate the model performance. The results showed that models like Efficient-Net B4 and B7 acquired more than 96% accuracy. Additionally, models like DCNN and EsSO got up to 96% accuracy. Therefore, these findings suggested that the optimized DL models can outperform previously existing traditional models. However, the research also mentions some limitations which highlight the negative impact classification performance overall as the efficacy of models based on ML-DL-TL is limited due to overlapping labels in the dataset. Moreover, the authors also mention the need for additional research to understand ensemble models on diverse datasets to better detect a wider variety of eye diseases.

In [2], Patel and Patel (2018) focus on the analysis of the Cup to Disc Ratio (CDR) in retinal fundus images for the automatic detection of glaucoma. The researchers utilized the Support Vector Machine (SVM) model for the classification and distinguishing between normal and glaucomatous images which was done based on the extraction of desired features from fundus images. Following this, component analysis is carried out in order to extract the optic cup component from the images for then calculating the CDR values. The dataset utilized consisted of categorized retinal fundus images. The categories marked whether the images were healthy or glaucoma-affected, and the study also includes multiple samples for better results. The main objective for the researchers was to accurately measure the CDR value of fundus images for accurately detecting glaucoma at an earlier stage. Therefore, they suggested an automated algorithm that can calculate the CDR value of a fundus image based on different color channels (red and green) to get an accurate and improved detection of glaucoma. Initially, the image data requires the Region of Interest (ROI) to be defined and pre-processing to make the data for the main steps. The methodology includes segmenting the image data into an optic disc and cup based on both red and green color channels. Consequently, the CDR value is calculated using the area of the disc and cup obtained in the previous step. Lastly, the extracted features are then classified using the SVM model to detect the presence or absence of glaucoma in the input fundus image data. The research results suggest that the method proposed can accurately differentiate between healthy and glaucoma-affected eyes based upon the CDR values respectively. The study also demonstrates its effectiveness by including results of different combinations of the color channels. However, some limitations of this study by Patel et al. show that there was a requirement of human intervention to define the region of interest (ROI) in the dataset which can hinder the automation process in general. Moreover, the study has not shown any significant comparison to other existing models in terms of accuracy and effectiveness.

In [7], Thanh et al. (2020) suggested the use of a real-time system that uses Artificial Intelligence for the effective detection of glaucoma from an optic nerve image dataset. The researchers utilized a dataset of 1677 labeled for both healthy and glaucomatous optic disc images. Furthermore, they also used a validation dataset of 500 images to test the performance of the model. The problem statement highlights the importance of early detection of glaucoma for which the authors suggested the use of a real-time glaucoma screening system called the EyeDr. built to leverage AI classifying these optic nerve images. Their research can be divided into two essential steps, which are object detection and glaucoma classification. They used the YOLOv3 system to detect the optic disc and cup from the fundus images. The model training process was carried out separately for both regions for over 10,000 iterations. The results from this model showed a significant performance with an accuracy of 92% and precision of 91.2% with a specificity of 91.34% and a sensitivity of 92.68%. Although the research indicates an overall significant result, the model is highly reliant on a limited dataset; thus, a comprehensive evaluation in a diverse population should be implemented to better understand the scopes of this model.

In [15], according to Jerry et al. (2023), diabetes has an impact on the eyes, particularly the retina, which initially exhibits no symptoms but may eventually lead to blindness. The risk of developing this eye complication increases with the duration of diabetes and the severity of the blood sugar level. Diseases like diabetes, stroke, hypertension, and coronary heart disease may be brought on by the retina's vulnerability due to the complex structure of the eye. This disease's main consequence is that it can cause serious issues, including permanent blindness, if treatment is delayed. The issue is that there are no symptoms at the early stages of diabetic retinopathy. The eye blood stops and remains after a few months. The study employed a deep learning model to identify this. To separate them for testing and training, they pre-processed them. Some of the pictures were saved for the test, while others were saved for the training. They then carried out the work using VGG16. They begin with scanning and improving images and end with photograph classification. Based on the patient's photo showing the affected eye area, they were able to diagnose diabetic retinopathy. Two thousand pictures were sent to them. The 3*3 filters used in VGG16 were applied to 224*224 RGB images with a stride of 1. Following the convolution, stride 2 was used for 22 frames with max-pooling. 94% accuracy, 96% sensitivity, 92% specificity, and 92% precision were achieved.

In [11], Rokhana et al. (2022) highlight the use of weighted MobileNet V2 machine learning models to get better classification results from fundus images. The research dataset was inclusive of 601 fundus images which were categorized into four specific classes, which are: Normal, Cataract, Retina Disease and Glaucoma. Additionally, the images with the Normal class label were three times in quantity, thus creating an imbalanced dataset. The problem statement of this study brings to light the challenges of accurately classifying the various ocular diseases when datasets are imbalanced. Thus, the authors suggest the use of a weighted cost function in the training phase of the models to enhance the system to better work with the minority classes and consequently perform better in the overall detection process. Their methodology includes splitting the dataset into a 7:3 ratio of 420 training images and 181 testing images. The training model was pre-trained on an ImageNet dataset for understanding the essential details which was later fine-tuned according to the specific fundus image dataset. They utilized the weighted cost function by adding more weight to the minority classes to reduce the effects of the dataset imbalance. The models utilized augmentation and trained it using the Adam optimizer. Consequently, the best results from the paper showed an accuracy rate of 66% with precision of 61%, and the recall and F-1 score were 58% and 57% respectively. Thus, these results suggest that the complexity of classification causes lower performance in comparison to other studies which have accuracy above 90%, even though the model seems to perform reasonably well. However, some shortcomings of this paper by Rokhana et al. suggest that the accuracy of only 66% is still low compared to recent deep learning models. Additionally, the complexity of classifying multiple ocular diseases from the image dataset was difficult using their suggested model which may have led to the inadequate performance of the system. And there are other studies which indicate that simplified classification and the use of a larger dataset can be utilized to overcome these shortcomings.

Till now, we have seen retinal disease-related research work where retinal fundus images were used to classify the diseases. Now, we will see many other papers which are related to our research motivation. There are many studies given below which demonstrate the detection of OT using various DL and ML models.

In [3], Chakravarthy et al. (2019) detected Ocular Toxoplasmosis from fundus images where they used an automated approach of convolutional neural networks (CNNs). The fundus images can be classified according to the healthy and unhealthy images based on Toxoplasma gondii which destroys some healthy parts of the retina. Normally the diagnosis for toxoplasmosis is a complex procedure as the dataset is very small. The OT lesions have different types, which makes it very tough to determine the types. Additionally, the availability and access to good clinical expertise as well as components is not enough for which misdiagnosis and mismanagement are common issues. In the study, they used a multi-test approach to differentiate between the healthy and unhealthy images. The dataset they used was collected from three countries' patients. There were 246 eye images of 215 patients with OT from three different academic centers in Argentina, Turkey and the United States. These images were included with color images of different modalities, and the dimensional range was 500 to 1000 pixels. Moreover, for evaluation, 66 healthy images were used. During pre-processing, the pixels of the images were resized to 512x512 for maintaining the same performance, and patches were extracted from images that contained lesions. If the patches contain a single OT entity at that time the image was identified as an unhealthy image. Image patches sizes were varied from 2 to 362 pixels in height and width. Furthermore, they used Deep CNNs for the classification of toxoplasmosis. From deep CNNs, the VGG16 model was fine-tuned for lesion detection. In the second step, they used a hybrid dual-input CNN for enhanced image-level classification. The evaluation was done by using AUC (area under receiver operating characteristic curve), sensitivity, and specificity metrics for the toxoplasmosis classification, where sensitivity defines how the model classifies the patches as a true class or not and true means unhealthy image, whereas specificity defines how the model classifies the patches as a false class or not where false means healthy image. In experiments and results, it worked better with patch-level classification rather than using the sliding window protocol for image-level classification and Hybrid Dual-Input CNN. Finally, the heat maps from the image-level classifier worked as additional input, which developed the hybrid dual-input CNN and got better accuracy. One of the limitations was not having enough data which caused problems in getting a good result using the deep learning model. Furthermore, image quality, lesions size, and lesions' location (e.g., legions in darker regions) can affect the classification of the images as well as the OT disease detection.

In [20], Choudhury et al. (2024) analyzed the fundus images of the eye, and developed an innovative Convolutional Neural Network (CNN) model to detect and classify Ocular Toxoplasmosis (OT). Their study addresses a previously unexplored gap in the application of deep learning for early OT diagnosis. Initially, the CNN model classified OT images into four categories: healthy, active, inactive, and mixed (active-inactive). After refinement, the classification was simplified into a binary system, such as healthy and unhealthy classes, where all non-healthy classes were consolidated into a single category. The research team benchmarked the performance of their custom-trained CNN model against four popular pre-trained architectures (VGG-16, VGG-19, MobileNet, and ResNet50) using the same dataset. The proposed model achieved 95% accuracy against these pre-trained models, demonstrating its potential to diagnose ocular toxoplasmosis and possibly outperform existing models for retinal disorder diagnosis. However, a key limitation of the study by Choudhury et al. lies in its binary classification of OT as either "normal" or "abnormal". While this simplification streamlines diagnosis, it fails to account for clinically relevant distinctions between active, inactive, and mixed disease states. Furthermore, the model was trained on a specific dataset, raising questions about its generalizability to other datasets or real-world clinical environments. Future research could focus on incorporating finer-grained classifications, expanding the diversity and size of training datasets, and validating the model in real-world clinical settings to enhance its practicality for end-users.

In [5], Abeyrathna et al. (2020) applied Directed Fine Tuning to Toxoplasmosis Fundus Images. In medical imaging, the segmentation of images is often used for early detection and management of increases in many types of diseases. They chose to localize rational lesions and scars which result from Ocular Toxoplasmosis [OT] as an instance segmentation problem of fundus images during their research. The dimensions and shape of OT entities vary significantly, which makes manual detection difficult to determine whether OT consists of an actual type or not. In this work, it was argued that learning a well-tuned deep network can be beneficial for this purpose. This paper is threefold: an instance segmentation network was developed based on Mask R-CNN, and finally, a novel unsupervised learning-based approach for directed analyses was proposed to advance the state-of-the-art. They utilized 246 images from three academic institutions in the US, Turkey, and Argentina. All images were annotated by trained medical professionals. Instance segmentation Mask R-CNN can also be used for instance segmentation. First, they fine-tuned a simple Mask R-CNN segmentation on a training set composed of positive OT images. After that, they optimize for upcoming clustering. The method requires four inputs: i) training images with their GT annotations, ii) predicted mask images P, iii) basic model BM, and iv) effort parameters [0, 1]. The fine-tuned model (FM) is the outcome of this process. The method can be outlined by the following main steps: they resize and convert the images to the required dimension, adjust contrast, etc. Moreover, for preprocessing, they also remove unwanted features from images. The archive data formats were SVG/JPG/PNG. Resize image: width = 512 and height = 512. Finally, for evaluation, they employed Intersection over Union (IoU), generally referred to as Jaccard Index for predicted masks; the evaluation Matrix was frequently used in the segmentation task. In the last, they trained the fundamental model after evaluating it and subsequently set up fine-tuning.

In [18], Yilmaz (2023) has explored several approaches, such as examining the fundus, fluorescein angiography has been used to diagnose toxoplasmosis, an infection caused by Toxoplasma gondii, that is associated with retinal inflammation. Nine different deep learning models were evaluated using PAM (Polygonal Area Metric) on a dataset of 412 OCT images (131 healthy and 281 diseased). Among these, the EfficientNetB0 model was the best performing, and a modified version of this model named modefficinet v1 outperformed it. In the classification of toxoplasmosis, Yilmaz claims that ModEfficinet v1 is an impressive model choice, as it shows promise for enhancing diagnostic accuracy and patient care. Additionally, future studies might aim to validate this model as a screening tool before implementing it in clinics or use other retinal diseases during the modeling process. However, while Yilmaz's work shows considerable progress in classifying toxoplasmosis, there are limitations. The dataset used contained only 412 OCT images, which may be too small or homogeneous to generalize the model's performance across different populations or variants of the disease. Furthermore, the study is primarily concerned with classification accuracy and does not delve into other important factors such as interpretability or generalization to real-world variability in imaging conditions. Finally, for future work, larger and more heterogeneous training datasets could be used to address this limitation. Additionally, the model's robustness and clinical feasibility should be tested across different populations.

In [13], Gupta et al. (2023) stated that cataracts are a very common eye disease, especially affecting elderly people. Cataracts can be detected by studying ocular fundus images. If detected at an early age, they are treatable. Machine Learning can detect this disease from a dataset of fundus images. SVM is also used for binary classification tasks, distinguishing between true cataract patients and false cataract patients. In this paper, they used many machine learning methods to detect cataracts, including Support Vector Machine (SVM), random forests, decision trees, logistic regression, naive bayes, k-nearest neighbors, XGBoost, light gradient boosting, and voting classifiers. Finally, the outcome was achieved by using light gradient boosting.

In [16], Noguera et al. (2023) used deep learning to classify fundus images of the disease. This study was the first to perform multi-class classification to differentiate between active and inactive Ocular Toxoplasmosis. The study described the classical ocular manifestation of toxoplasmosis, which is a nidus and fluffy white, and adjacent to a variable pigmented chorioretinal scar, indicative of focal necrotizing retinitis or retinochoroiditis. Although many diagnosis procedures are effective, many issues can be created such as giving false-negative or false-positive results, and sometimes this kind of error can be made by experienced doctors. In such cases, Machine Learning (ML) methods could be beneficial as used here. Additionally, Deep Learning (DL) is a subfield of ML involving Artificial Neural Networks (ANNs), and was used alongside Conventional Neural Networks (CNNs). This study determines the OT through binary classification, not only distinguishing healthy or unhealthy images but also differentiating between active and inactive lesions or inactive scars in unhealthy images. Furthermore, multi-class classification can be used in the future to enable automated diagnosis for specific cases. Their dataset included 412 fundus images captured using two different cameras. And one ophthalmologist classified the images based on OT lesions as either active or inactive, where active lesions have various sizes and colors (e.g., white, yellow). Moreover, in some cases, active lesions can be challenging to identify due to vitreous opacities. During data pre-processing, images were resized to 512x512 pixels. Pooling and convolution layers reduced the number of parameters in the model. A Residual Neural Network (ResNet18) was used to classify images. The model was trained by ImageNet and then fine-tuned with their data. CNN was applied with a pre-trained model (VGG16) with additional convolutional layers. A stopping criterion was used to stop the execution when the loss plateaued, and binary cross-entropy loss was used for optimization. After calculating everything, they calculate the sensitivity and specificity, where the model achieved 91% sensitivity and 98% specificity. The research team claims that this is the first work to address OT disease using a multi-class approach. However, the limited dataset can be a limitation of the work since the result may only work well sometimes. While the dataset contained more than 400 images, its limited size and class imbalance (132 healthy, 91 active, 190 inactive) may affect generalizability. Labeling bias in active/inactive criteria and the lack of comparison to other diagnostic methods (e.g., PCR, serological tests, other lab tests) were noted limitations.

In [24], Milad et al. (2024) developed an Automated Machine Learning (AutoML) model for classifying fundus images without having any coding experience but they worked with deep learning (DL) models. The team trained the model using 304 labeled fundus images, focusing on ophthalmology. As a result, their AutoML model had an area under the precision-recall curve (AuPRC) with sensitivity of 100%, specificity of 83% and accuracy of 93.5%. Their model was labeled with 15 normal fundus images and 15 confirmed OT fundus images, where they got satisfying accuracy. AutoML offers many advantages over other traditional models to classify fundus images, which the authors argue could improve OT detection. Despite lacking coding skills, the team successfully applied a machine learning model to determine Ocular Toxoplasmosis from fundus images. However, in the case of balancing precision and recall, they may face some challenges.

In [23], M and Narayanan (2024) extensively explored DL-based ocular disease detection with fundus and optical coherence tomography images. First, the authors discuss the early detection of ocular diseases such as Diabetic Retinopathy (DR) and Glaucoma, which can result in permanent vision loss without timely treatment. They demonstrate the role of a variety of deep learning frameworks, including convolutional neural networks [CNNs], recurrent neural networks [RNNs], generative adversarial networks, U-Net, Y-Net, and transformer-based models, in tasks such as fundus image classification and retinal structure segmentation. Their survey spans numerous applications, ranging from retinal vessel, optic disc, and cup segmentation to DR classification or grading. Finally, based on the findings of this review, they address remaining challenges in the field, with emphasis on handling large-scale unlabeled data and coping with computational limitations. They suggest such options as running on GPU workstations, moving to federated learning for the privacy of data, and using Generative AI technology features that shall bring higher performance. They conclude by discussing the implications of catastrophic forgetting in epilepsy, opportunities for future work such as utilizing weight decay and generative replay to prevent this phenomenon, and how transformer-based models may aid seizure detection within epileptology. In contrast, their review of deep learning (DL) in ocular disease detection is well-considered but one that can be further expanded on. Their review provides little information about the diversity of datasets used in studies, which is pivotal to establishing general equilibrium and performance over different populations or demographics. Furthermore, the interpretability of these models is often ignored despite it being a requirement for clinical acceptance. It might also explore the specific challenges associated with learning DL models across different implementations, imaging modalities and institutions. Moreover, there is a scarcity of detail regarding real-time application to clinical workflow and hardware specifications. Some ethical issues, such as respect for user privacy and neutrality of the model during training were unexplored.

In [14], Hassan et al. (2023) created a DL model to screen ocular toxoplasmosis using fundus images. Using Google's automated machine learning (AutoML) platform, the researchers constructed the model without manual coding. They collected fundus images of patients who had Ocular Toxoplasmosis from centers in Argentina, Turkey and the USA, and normal subjects were obtained using public databases. A total of 712 fundus images (552 Ocular Toxoplasmosis, 160 normal) were split into training, validation and testing sets. They reported excellent performance, with an area under the precision-recall curve (AUPRC) of 0.99 and sensitivity, specificity, positive predictive value and overall accuracy equal to 96.5%, 100%, 100% and 97%.

This paper concluded that an AutoML-driven model holds promise in aiding clinicians with Ocular Toxoplasmosis diagnosis and the possibility of identifying new disease biomarkers, emphasizing the potential applications of automated deep learning for medical diagnostics. However, although this study demonstrated excellent performance for detecting Ocular Toxoplasmosis using AutoML-based deep learning model systems, there are a few possible limitations and directions for future research. The dataset is varied; however, it might not be representative globally as the fundus images were captured from limited countries. Generalizability could be improved by expanding the dataset with images across diverse populations and regions. One limitation to the external validation of the model, which was done retrospectively on patients without concomitant ocular diseases, may limit its application in real-world situations where most confined older adults present with multiple conditions. Subsequent work may focus on continuously training the model on Ocular Toxoplasmosis cases alongside a mixed pool of other ocular diseases, enhancing the diagnostic versatility for complicated cases.

In [9], Parra et al. (2021a) used deep learning strategies with the aim of improving ocular toxoplasmosis (OT) diagnosis on fundus images by applying a pre-trained Residual Neural Network (ResNet18). Numerous studies underscore the difficulty of OT diagnosis, especially in areas lacking expert ophthalmologists, and demonstrate the possible usage of predictive computational models with high sensitivity (94%) and specificity (93%). By developing a new dataset of 160 fundus images obtained from the Hospital de Clínicas in Asunción, Paraguay collected specifically for this study that allowed researchers to fine-tune the ResNet18 model and motivated them to release an open-labelled image set to benefit more researchers. The results indicate that deep learning algorithms, especially when trained on broad datasets, hold promise as diagnostic tools to support ophthalmologists in recognizing atypical cases and potentially optimizing the clinical outcomes of patients in regions with mixed disease burden. Although the study by Parra et al. shows potential in automatic OT diagnosis, the model still lacks some abilities. The model was trained on a relatively small dataset (N = 160 fundus images), which may affect its generalizability across a broader population. Alternatively, depending on the nature of the cases, amplifying the dataset to accommodate more generic use cases thereby can enhance robustness. Although class imbalance was addressed during preprocessing, lower performance under imbalanced conditions suggests a need for advanced techniques like synthetic data generation or complex sampling. The current model focuses on binary classification (healthy vs. OT), but additional labels (e.g., active or inactive lesions) could improve diagnostic ability of ophthalmologists. Finally, since results were validated only on a controlled private dataset, external testing across diverse clinical settings with real-world images is critical to assessing noise tolerance.

In [22], Kulkarni et al. (2024) make good progress through few-shot learning (FSL), and as for now, it remains amongst the only works on applying meta-learning to diagnose toxoplasmosis chorioretinitis, tackling smaller labeled data issues in medical imaging. Given the known challenges regarding over-fitting and difficulty in defining complex features using traditional machine learning approaches, they chose to use transfer learning along with a few well-recognized Convolutional Neural Network (CNN) architectures namely ResNet50, VGG-16 and InceptionV3. It has been shown that Reptile meta-learning algorithms can help improve classification accuracy better than other types of Transfer Learning methods when it comes to such kinds of problems. Through their robust assessment, they found the accuracy of Inception V3 was best (84%), next ResNet50 (80%) and then for VGG-16 at only (66%), indicating its power in differentiating toxoplasmosis retinitis. Their work demonstrates the ability of Reptile to adapt various models trained with deep convolutional neural networks (CNNs) on ocular disease datasets and highlights opportunities for integrating advanced CNN architectures into meta-learning methods to improve medical imaging diagnosis by validating progress regarding developments in this field. In contrast, although this study provides important information regarding few-shot learning and meta-learning approaches for diagnosing toxoplasmosis chorioretinitis, several limitations need to be considered. The success of their solution would likely hinge on the heterogeneity and representativeness of the dataset, serving to restrict how well a model that is trained upon it can generalize to real-world scenarios devoid or limited set of demographics among patients or disease stages. Moreover, a metric such as accuracy is not the ideal performance measure for imbalanced data since it can miss important points of model evaluation; using metrics like precision and recall will add value to the overall evaluation (F1-score). The complexity of the CNN architectures used may also affect model interpretability, making it difficult for clinicians to understand predictions; therefore, improving interpretability with methods like attention visualization might be useful. Applying their method to other ocular diseases would also broaden the usefulness and generalizability of their approach. Finally, the scalability and computational performance of the models are also vital for practical purposes in addition to clinical validation with real-world patient data.

In [8], Parra et al. (2021b) developed a methodology to assess deep learning (DL) models for automatic diagnosis of ocular toxoplasmosis (OT) based on fundus images. One of them was a trust-based evaluation framework which aims to tackle the challenge of building trust between the medical community and AI models. This axiomatic framework assesses the concordance between the DL models and human decision-making by calculating variant distance over attribution matrices derived from interpretable methods like Integrated Gradients and comparing them to accurate segmentation masks of OT lesions, annotated by experts. They tested three DL architectures: Vanilla CNN, VGG16, and ResNet18. Their more complex models consistently performed best in terms of accuracy but scored worst in perceived trustworthiness. They propose a key trust measure for evaluating DL models in medical imaging and support openness since transparency is a paramount reliability metric that should be considered when applying AI to healthcare applications. The team also provided an open-source dataset of annotated fundus images for further research in this field. In contrast, the work by Parra et al. provides a trust-based approach to reproduce the results of deep learning models in OT diagnosis from fundus images, which has several drawbacks. In future work, one could expand the work on the reliance on Integrated Gradients as an attribution method by including different interpretability methods to improve the understanding of the model's trustworthiness. Additionally, regarding the trade-off between model complexity and trust highlighted in the study, more specific insights into why some architectures like VGG16 or ResNet18 score lower in trust would strengthen the analysis.

This work has multiple limitations including the relatively small dataset (160 fundus images) and the lack of direct validation with ophthalmologist perception to confirm alignment between the trust score and expert judgments. Furthermore, the approach is limited to only OT diagnosis, and it remains to assess its potential across other medical sectors, while overcoming these limitations would enhance the strength and clinical relevance of their approach.

In [17], Tiwari et al. (2023) performed Toxoplasmosis Chorioretinitis Detection by using a Hybrid ResNet-YOLO Classifier. The authors proposed a new method that works with Hybrid ResNet and the YOLO (You Can Look Once) classifier, achieving a high recognition rate with their hybrid deep-learning model. Their model achieved 99% accuracy. The study focused on medical picture analysis using YOLO models. Toxoplasmosis Chorioretinitis is a very severe condition that can cause vision loss of a person; and the disease affects approximately one-third of the population in the world. Ocular lesions occur when the reactivation of immunosuppression takes place, such as HIV infection that causes blurred vision, floaters and some severe visual impairment. This is treatable only if the treatment starts early. Additionally, early treatment is critical, and tools like deep learning and computer-aided diagnostic systems provide better diagnosis as well as treatment facilities to get rid of this disease. Existing diagnosis methods for Toxoplasmosis including checking medical history, ophthalmoscopy (using an ophthalmoscope to examine the retina and choroid), fluorescein angiography (imaging eye blood vessels), serological tests (detecting antibodies in blood), molecular testing (Toxoplasma DNA detection by polymerase chain reaction), biopsy and differential diagnosis. Deep learning applications in medical imaging for Toxoplasmosis include image classification and segmentation, disease detection and diagnosis, personalized medicine, object detection and localization, feature extraction, and multi-modal fusion. In the data pre-processing, the authors standardized and enhanced the dataset through scaling, normalization, and image augmentation. Initially, 355 fundus images were used and then the dataset was expanded to 700 images via zooming, rotating, flipping, shifting and so on. They used the FS-SSO technique during fitness calculation. Their training and validation accuracy were closely aligned. However, they had some limitations such as the data scarcity, raising concerns about overfitting. While deep learning methods were used, the authors noticed that accuracy may not consistently be maximized due to low-resource datasets. Finally, for complex disease presentations, alternative machine learning models could complement their approach.

In [12], Ferdous et al. (2023) investigated the determination of proteomic diagnosis for ocular toxoplasmosis (OT), a condition caused by Toxoplasma gondii. Their study displays the techniques to distinguish between active and inactive toxoplasmic lesions by applying DL models to ocular fundus images. The steps taken in the study included designing a workflow, compiling a dataset of 412 fundus images, and pre-processing them with normalization and augmentation for model training. They used pre-trained models such as VGG16 and developed two models: one was Artificial Neural Networks (ANNs) and the other was Convolutional Networks (CNNs). Results showed that the ANN model achieved 98% accuracy, while the CNN model achieved 95% accuracy. This study underscores the robustness of deep learning methods for medical diagnosis, particularly in conditions like OT where current guidelines rely on complex clinical examinations and patient histories. The findings suggest a significant potential for AI-supported diagnosis in ophthalmology, promoting the advancement for further improvements in data pre-processing and model optimization. However, while there are numerous opportunities for improvement, several limitations also have been noted. For instance, although the dataset size was expanded to 412 images, it remains relatively small, which may limit deep learning model performance. Additionally, the exclusion of images depicting coexisting active and inactive lesions (to avoid ambiguity) might reduce model efficacy in complex cases. While the current study focused on ANN and CNN architectures, exploring other models or hybrid approaches could provide new directions to improve accuracy. In addition, though preprocessing techniques like CLAHE or segmentation were discussed, none were applied which might have enhanced image resolution and model performance. Moreover, the interpretability of the models was not thoroughly examined; integrating methods to explain predictions using comprehensible features could bolster trust in AI diagnostics. Finally, the lack of clinical trials could be introduced to validate their hypothesis with real-world patient data since it is difficult for many to find practical applicability; and addressing these constraints could potentially improve the research's impact on large-scale clinical diagnosis of ocular toxoplasmosis.

In [19], Alam et al. (2024) studied the performance of transfer learning in medical image analysis with an emphasis on Optical Coherence Tomography (OT) diagnostics. The study aimed to evaluate the ability of various pre-trained convolutional neural networks (CNNs) such as VGG16, InceptionV3, ResNet50, DenseNet121 and MobileNetv2 for classification as well as segmentation tasks using retinal fundus images. Overall results summarized in a table suggest that MobileNetV2 achieved an outstanding classification with an accuracy of 0.976, outperforming the SoTA VGG16 model with an accuracy of 0.973 in their dataset. Furthermore, the authors showed superior segmentation results with MobileNetV2/U-Net which exhibited the least mis-segmentation of tissues compared to other models. The authors systematically tested various pre-processing methods and deep learning architectures, emphasizing the importance of model complexity, inference time and computational efficiency. They acknowledged limitations in their work, particularly the dataset size being small, and suggested the need to refine model performance with larger or more balanced datasets through self-supervised learning schemes. In conclusion, their work highlights the promise of modern deep learning methods to improve automated diagnosis and semantic segmentations in OT problems in medical image analysis. In addition, results validate promising achievements in the classification and segmentation of ocular toxoplasmosis (OT) using deep learning models. However, the study is not without limitations; where the major limitation was the relatively small size of the dataset that could reduce the robustness of the models for rare cases and for task-specific complex architectures. As a result, the authors opted to use a pre-trained model rather than custom networks. In addition, while the MobileNetV2/U-Net model achieved the best performance, minor mis-segmentation persisted in fine details near tissue edges. Moreover, the inference time for models such as DenseNet121 was too large for practical implementation on edge devices, requiring real-time processing. To address these limitations, the authors propose scaling up the datasets to improve accuracy without sacrificing resolution, tackling complex issues about expanding data as well as exploring architectures like transformers for enhanced image analysis.

2.3 Summary of Key Findings

Table 2.1: Comparison Result Analysis of Literature Review Papers with Proposed Model Evaluation

Paper Title	**Models Used**	**Evaluation Results**	**Binary or Multi-Class**
Alam et al. [25]	VGG16, MobileNetV2, InceptionV3, ResNet50, DenseNet121	Accuracy: 98.9% (MobileNetV2)	Binary (classification)
Chakravarthy et al. [11]	CNN	Accuracy: 92.5%	Binary (classification)
Noguera et al. [16]	ResNet-50	Accuracy: 86.7%, Recall: 91.2%	Multi-class (Healthy / Active / Inactive)
Our Proposed Model	DiNet-Fusion (EfficientNetB0, DINO ViT-Small)	Accuracy: 92% Active (F1: 89.66%, Recall: 92.86%, P: 86.67%) Inactive (F1: 90.91%, Recall: 89.29%, P: 92.59%) Healthy (F1: 95%, Recall: 95%, P: 95%)	Multi-class (Healthy / Active / Inactive)

Chapter 3

Methodology

The aim of this study is to create a strong and data-efficient classification model for OT utilizing a dataset with few data samples. Details of the methodology presented in Figure 3.1. Firstly, the data was collected from the Zenodo website [10], then preprocessed by resizing to 224 pixels. In the next step, the dataset was stratified, split into train-test-validation, and the class imbalance was handled. Subsequently, the extraction of the characteristics was performed using EfficientNet-B0, which is a baseline model that extracted a 1280-dim CNN feature and fused with DINO ViT-Small. After the feature-level fusion with EfficientNet and DINO (DiNet-Fusion), the model starts training. After model training, we evaluated it using the accuracy and loss curve, classification report (precision, recall, f1-score), per-class accuracy, ROC curve and AUC values, and probability distribution analysis. Lastly, threshold tuning and model testing was conducted. Finally, the result was analyzed. In short, the entire workflow in our study includes dataset download, preprocessing, splitting and data augmentation, architecture design, proposed DiNet-Fusion model training and testing.

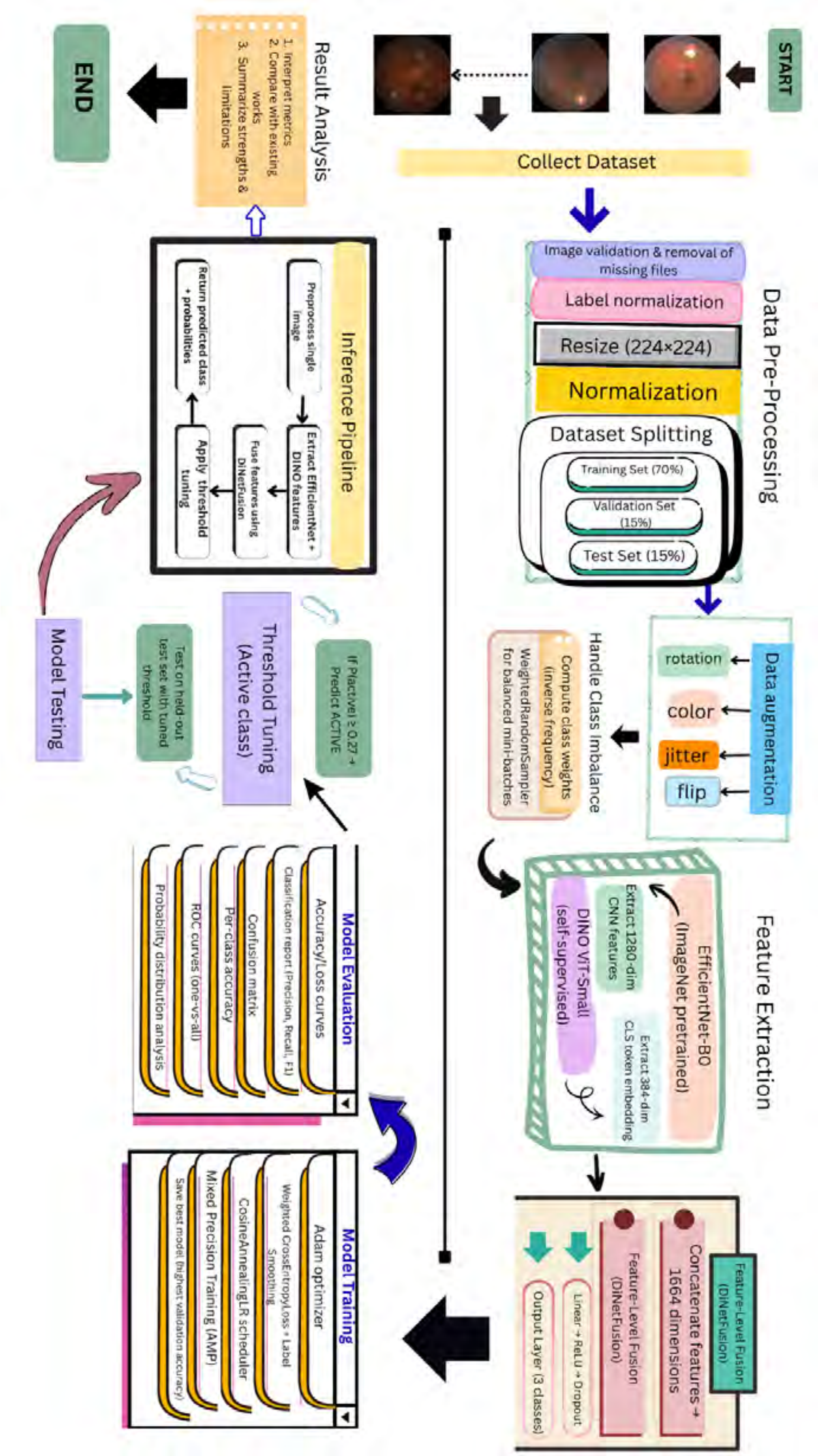

Figure 3.1: The Top Level Overview of the Proposed System

3.1 Dataset

The Ocular Toxoplasmosis Fundus Image Dataset (OTFID) - Version 3 has been used in this work where the dataset is publicly available on the Zenedo website [10].

The dataset contains a set of eye images that were collected from the Hospital de Clínicas and Hospital General Pedriático Acosta Ñu medical centres from Asunción, Paraguay. Samples of the dataset were collected between the years 2018 and 2021 from the two hospitals. The dataset includes two major folders where all the collected images are in one folder, and another folder contains the masks for the lesions of the images with OT. Moreover, the dataset includes a CSV file that contains the image labels. Furthermore, the masks were only for unhealthy images to make it easier to differentiate them from healthy images [10]. Additionally, 291 eye images (of size 2124 x 2156 pixels) and 121 images (of size 1536 x 1152 pixels) were collected from the Hospital de Clínicas and Hospital General Pedriático Acosta Ñu medical centers, respectively [19]. Therefore, there are a total of 412 fundus images in the dataset which has three different classes; such as Inactive, Active and Healthy. The masks of the images include the suffix "-a" for all the active lesion masks in order to differentiate the uncommon active lesion from the inactive lesion. Table 3.1 displays classes and the number of image samples each class contains. From Table 3.1 we can see how imbalanced the data samples are. Due to the restricted amount of available fundus images, the methodology prioritizes meticulous dataset preparation, proportional sampling, advanced feature extraction via pre-trained models, and a fusion-driven classification approach.

Healthy Inactive Active

Figure 3.2: Healthy, Inactive and Active Images

Figure 3.3: Mask of Active (Affected) Images

Table 3.1: Total Number of Images in Dataset with Classes

Label	Number of Images
Inactive	188
Healthy	132
Active	92

3.1.1 Data Prepossessing

To prepare the dataset for model training, certain types of preprocessing steps were taken: We have taken the dataset from the publicly open-source Zenodo where the dataset contains 412 images which were taken from two hospitals in Paraguay. The images have three classes, including Active, Inactive, and Healthy. We track the image class information from the CSV files. The CSV file contains the names of the images and the label (of which class they belong to). During preprocessing, we kept only the image name and label columns. In the label, "active" means Ocular Toxoplasmosis disease is present; "inactive" means there is another ocular disease but not OT. On the other hand, "healthy" means there are no eye illnesses present. To map the images according to their labels, we used integer labels 0, 1, and 2, where they are active, healthy, and inactive, respectively. The images were split into 70%, 15%, and 15% for Train, Test and Validation, respectively. The image size was resized to 224 pixels because EfficientNet-B0 requires an image size that is a multiple of 7. During augmentation the images was flipped and rotated. Moreover, ColorJitter is also used for brightness and contrast so that the model can identify the images carefully. A weighted random sampler was used, and then the classes of the fundus image dataset were used to load the images and transform them. The train, val and test dataloader was also created. Thereafter, the loaded pretrained EfficientNet-B0 backbone was attached to the pretrained DINO ViT-small model, and it was subsequently loaded. For loss calculation, we used cross-entropy, class weighting, and label smoothing. The Adam optimizer, Cosine AnnealingLR for the scheduler, GradScaler, and autocast for mixed precision tools were used.

After the training was completed, the best-performing model was loaded for the evaluation, and all the predictions, true labels, and class probabilities were collected. Based on these outputs, the classification report, confusion matrix, and accuracy of each class were generated. Thereafter, post-hoc threshold tuning was applied. If the predicted probability of active class was ≥ 0.27, in that case, the prediction was forced to active, and then a new classification report and confusion matrix were generated. Additionally, for visual representation, several graphs and diagrams were generated, including the training vs validation loss curve, the accuracy curve, the threshold tuning confusion matrix heatmap, the ROC curve for each class and the micro-average, the probability distribution for the three classes, a distribution plot, and a bar chart comparing correct and incorrect predictions. Finally, the threshold-aware model was saved, and the predict function was created, which can provide a threshold-based 3-class probability return if there is a new input image.

3.2 Techniques used in Training, Validation and Testing

A full pre-training pipeline was implemented ahead of model training which includes data cleaning, excluding missing images, normalizing columns and label encoding. The dataset was subsequently divided utilizing a stratified 70/15/15 approach, after which all transformations were defined as follows: the training set underwent augmentation (resize, flip, rotation, color jitter), whereas the validation and test sets employed solely deterministic resize and tensor conversion. FundusDataset and DataLoader objects were generated, and class imbalance was addressed with the application of a WeightedRandomSampler just to the training loader. The model was initialized by loading pretrained EfficientNet-B0 and DINO ViT-Small backbones, and a fusion head was defined with optimization managed using Adam, CosineAnnealingLR, class-weighted CrossEntropy with label smoothing, and mixed-precision techniques.

Throughout training, each epoch executed gradient updates utilizing weighted loss in mixed precision, while validation proceeded without gradients to assess loss and accuracy. The scheduler progressively decreased the learning rate, and the optimal model was preserved according to validation accuracy. Following the end of the training, the best model was evaluated on the test set to generate predictions, classification reports, confusion matrices, and accuracy metrics of each class. A post-training threshold adjustment was used, and thus active class probabilities of ≥ 0.27 were designated as "active", thereby enhancing clinical sensitivity. Finally, training vs validation curves, ROC plots, and probability analyses were produced, and a threshold-sensitive inference function with the finalized optimized weights was established for practical prediction.

3.3 Proposed Model Description

The DiNet-Fusion model comprises two simultaneous feature-extraction branches: EfficientNet-B0 and DINO ViT-Small resulting in a fully connected fusion head. The initial branch uses EfficientNet-B0 as a lightweight yet highly expressive convolutional backbone. The architecture starts with a 3×3 convolutional stem and then adds a series of Mobile Inverted Bottleneck (MBConv) blocks that include depthwise separable convolutions, expansion layers, and squeeze-and-excitation (SE) modules. As the channel depth increases, the spatial resolution decreases in each of the seven main stages. This gives rise to increasingly abstract semantic representations. A global-average pooling is then applied to the final feature map to give a 1280-dimensional visual descriptor. To use this network only as a feature extractor, we replace its default classification layer with an identity operation, allowing the model to output a compact global feature vector $f_{\text{eff}} \in R^{1280}$.

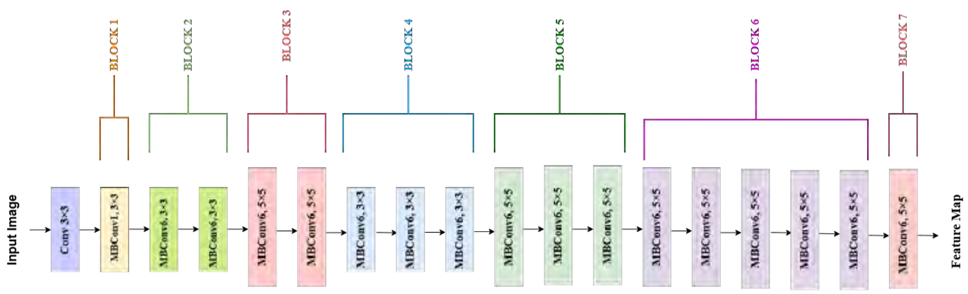

Figure 3.4: EfficientNet-B0 Architecture Diagram

The second branch uses DINO ViT-Small, a self-supervised Vision Transformer designed to capture global context and long-range dependencies. The ViT-Small/16 architecture segments each 224×224 image into 196 non-overlapping 16×16 patches, which are then flattened and projected into a 384-dimensional embedding space. A learnable CLS token is prepended to the patch sequence, and positional embeddings are added to preserve spatial information. The sequence is processed through twelve Transformer encoder blocks, each comprising multi-head self-attention, layer normalization, residual connections, and a GELU-activated feed-forward network. During inference, the forward_feature() function outputs the processed token sequence, from which the CLS token representation is extracted as the final image descriptor. The output of DINO is a feature vector, $f_{\text{dino}} \in R^{384}$, which encodes rich global contextual information learned via self-distillation.

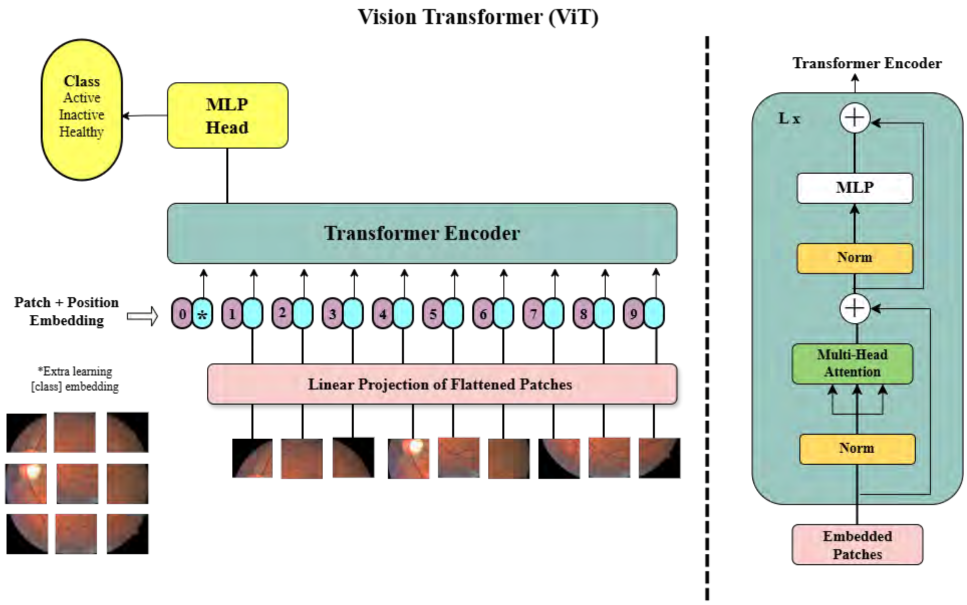

Figure 3.5: DINO Architecture Diagram

After representing the features for both branches, the vectors are merged into a unified vector of 1664-dimensional space. The concatenated vector is then passed into a multi-layer head with linear layers, ReLU activation, and dropout regularization. The fused feature goes through a fully connected layer that reduces its dimension from 1664 to 512, with ReLU activation and a dropout of 0.3. Following this section is a dense layer with a dimension reduction of 512 to 128, again using ReLU and a dropout of 0.2. The final layer is a linear transformation from 128 to 3. It outputs class values for active, healthy and inactive. This dual-stream architecture facilitates the incorporation of the high-resolution local lesion detail encoded by EfficientNet with the global context captured by DINO, providing a more comprehensive representation for ocular toxoplasmosis classification.

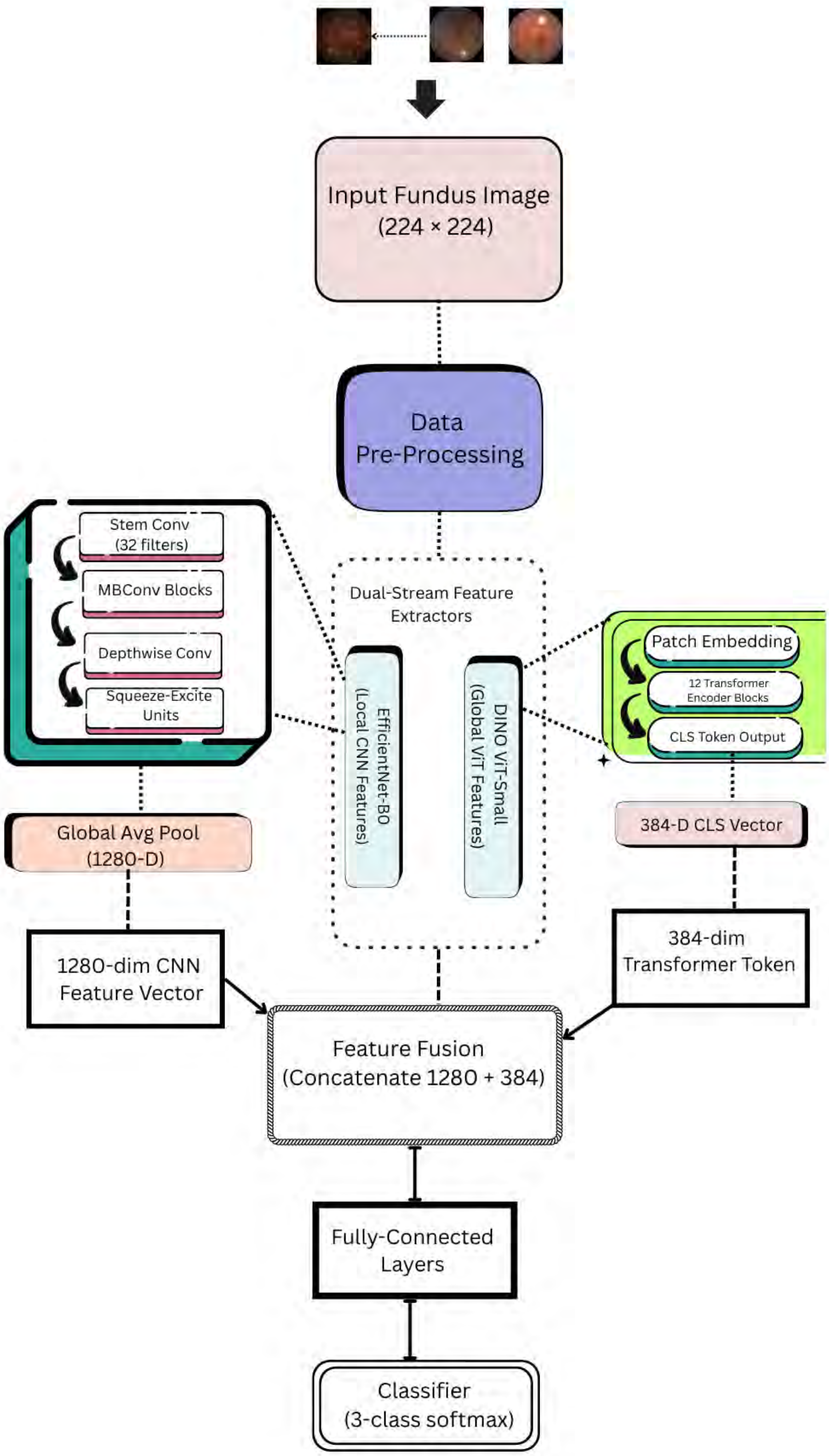

Figure 3.6: Top Level Overview of The Proposed Model Diagram (DiNet-Fusion Architecture)

3.3.1 Working Mechanism of the Proposed Model

The proposed model is an integration of two deep-learning architectures that modulate information sequentially: EfficientNet-B0 and DINO ViT-Small which uses fundus images showing distinct variants. As a convolutional network, EfficientNet-B0 is efficient at capturing local textural patterns or lesion-specific characteristics that may be critical to distinguishing subtle clinical markers. In sheer contrast, DINO ViT-Small is a vision transformer that captures long-range interdependence, global relationships and large spatial structures. Instead of freezing both backbones, they are jointly fine-tuned which permits the fused model to determine when local CNN features are beneficial and when transformer-derived global context is preferable. They concatenate their outputs and run it through a multi-layer perception fusion head that learns the most appropriate fusion strategy for classifying ocular toxoplasmosis.

To reduce class imbalance and enhance generalization, the model employs a class-weighted cross-entropy loss augmented with label smoothing. For a specified sample, the smoothed target distribution designates $q_y = 1 - \epsilon$ for the true class and $q_k = \epsilon/(K-1)$ for all other classes, where $\epsilon = 0.1$. The weighted loss is further calculated as,

$$L = -w_y \sum_{k=1}^{K} q_k \log p_k, \tag{3.1}$$

where $w_y = 1/n_y$ addresses class differences. This minimizes model overconfidence and enables optimized calibration. A post-hoc thresholding procedure is implemented to improve clinical sensitivity for the "active" class: if the predicted probability $p_{\text{active}} \geq 0.27$, the model overcomes the maximal prediction and pushes the output to be "active," hence reducing false negatives for clinically significant instances.

$$w_y = \frac{1}{n_y} \tag{3.2}$$

The following describes the full set of end-to-end workflow functions. Firstly, an input fundus image is resized to the size of 224×224 pixels; for training images, augmentation is applied while the validation, test, and inference images are only processed deterministically. Both backbones process the image, with EfficientNet-B0 outputting a global descriptor of depth 1280-dimensions and DINO ViT-Small outputting a CLS token with 384-dimensions. The two vectors combine into a 1664-dimensional fused feature, thereafter processed by the MLP classifier (1664 $\rightarrow$ 512 $\rightarrow$ 128 $\rightarrow$ 3) employing ReLU activations and dropout. During training, softmax probabilities are utilized to calculate the weighted, label-smoothed cross-entropy loss, and gradients update all parameters accordingly. During validation and testing, the same forward pass is conducted without gradient calculation, and predictions are derived using argmax. During evaluation, threshold tuning is implemented to modify active-class predictions, thereby producing improved confusion matrices and classification reports. The inference function analyzes new fundus images across the complete pipeline and implements the thresholding requirement to yield clinically dependable predictions and associated class probabilities.

Weighted Class Entropy + Label Smoothing.
Label smoothing $\varepsilon = 0.1 \rightarrow$ target distribution
for true class y: $\quad q_y = 1 - \varepsilon$
for other classes: $\quad q_k = \dfrac{\varepsilon}{(K-1)}$
One sample loss,

$$\mathcal{L} = -w_y \sum_{k=1}^{K} q_k \log p_k \tag{3.3}$$

Threshold tuning:
Normal Prediction: $\arg\max(p)$
Rule:

$$\hat{y} = \begin{cases} \text{active}, & \text{if } p_{\text{active}} \geq 0.27, \\ \arg\max_k p_k, & \text{otherwise.} \end{cases}$$

Chapter 4

Result Analysis

This chapter represents the experimental findings of our proposed model, DiNet-Fusion, a fusion-based architecture that integrates EfficientNet-B0 and DINO ViT-Small features to effectively classify OT into active, inactive, and healthy classes. This assessment aims to determine whether the combination of CNN and transformer embedding better discriminates across visually similar retinal conditions and how well our proposed architecture performs in datasets with limited resources. We present an evaluation consisting of comprehensive classification performance metrics, confusion matrices, training behaviour, threshold tuning to maximize the sensitivity of clinically meaningful classes, and several diagnostic plots to assess the reliability and generalization of the model.

4.1 Training and Validation Performance

To assess how the model was being trained, the accuracy and loss curves were plotted for each epoch during training. As we can see from figure 4.1, the training goes down almost significantly during the epochs and ends up pretty much close to 0.36 by almost the end of the training.

Although higher than the training loss, the validation loss indicates balanced generalization around 0.52 without any sign of overfitting. In small medical datasets, the validation variance is greater because of subtle image differences, so we expect minor variations in succeeding epochs.

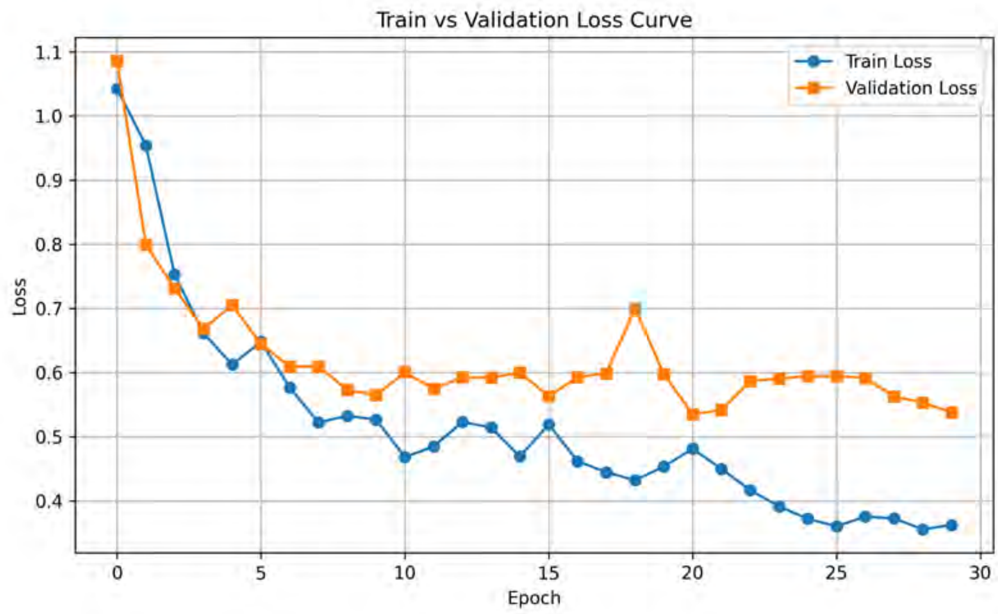

Figure 4.1: Train vs Validation Loss Curve

Similarly, Figure 4.2 shows a gradual increase in the accuracy of the model during the training period. About ten epochs later, the training accuracy passes 90% and stabilizes around 96% by epoch 30. Even though the validation accuracy is lower, it is between 82–90% throughout the training phase. This proximity of two curves signifies the learning efficiency and it also depicts the good generalization of the proposed fusion model when trained on small amount of data.

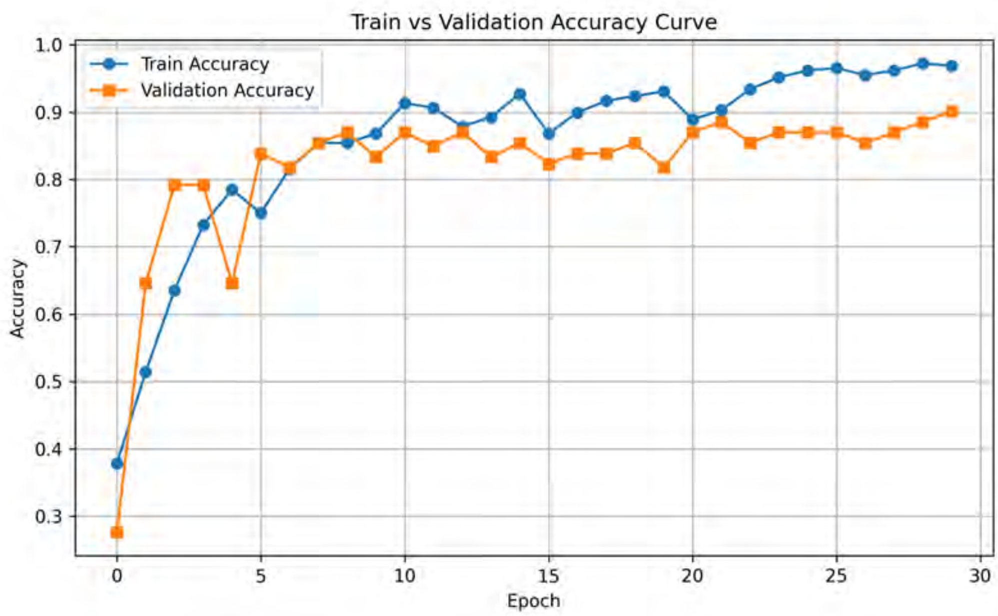

Figure 4.2: Train vs Validation Accuracy Curve

These two curves together demonstrate evidence that DiNet-Fusion gradually converges and prevents overfitting from weighted sampling and label smoothing while maintaining a consistent validation performance across datasets with a class imbalance.

4.2 Classification Results

Following training, an independent test set with 62 samples was used to assess the best-performing model, which was selected based on validation accuracy. Table 4.1, the initial classification report, shows an overall accuracy of 90.32%, indicating strong multi-class performance.

Table 4.1: Classification Report before Applying Threshold

	precision	**recall**	**f1-score**	**support**
active	0.9167	0.7857	0.8462	14
healthy	0.9500	0.9500	0.9500	20
inactive	0.8667	0.9286	0.8966	28
accuracy			0.9032	62
macro avg	0.9111	0.8881	0.8976	62
weighted avg	0.9048	0.9032	0.9024	62

This finding indicates that the model can differentiate accurately and reliably between the healthy and inactive cases (95% for the healthy class, 92.86% for the inactive class). This is expected in imbalanced datasets driven by class imbalance where the active class is the minority subset and hence has a lower recall of 78.57%. However, for the active class, it still gives a precision of 91.67%, meaning that most samples predicted as active are correct.
These outcomes suggest that the combination of CNN and transformer features affords more equitable predictions between each of the classes making up the multi-class OT classification task than traditional CNN model alone.

4.3 Threshold-Tuned Performance for Active Class

Because the active class signifies the most clinically urgent category, we implemented post-hoc probability threshold tuning to enhance sensitivity. A threshold of $T = 0.27$ was found to balance sensitivity and specificity optimally. Upon implementing this rule, predictions are mandated to be classified as "active" when the model yields a probability of ≥ 0.27 for the active class. The revised classification report (Table 4.2) demonstrates a notable improvement: The active class recall rises from 78.57% to 92.86% and the active class F1-score rises from 84.62% to 89.66%. This modification aligns the model with actual clinical requirements: minimizing false negatives (overlooked active infections), which indicate a greater risk than false positives.

Table 4.2: Classification Report after Applying Threshold

	precision	**recall**	**f1-score**	**support**
active	0.8667	0.9286	0.8966	14
healthy	0.9500	0.9500	0.9500	20
inactive	0.9259	0.8929	0.9091	28
accuracy			0.9194	62
macro avg	0.9142	0.9238	0.9185	62
weighted avg	0.9203	0.9194	0.9195	62

Figure 4.4 shows the threshold-adjusted confusion matrix similarly indicates this enhancement as active misclassifications have significantly decreased. Although two previously "inactive" samples have been reclassified as "active", this compromise is acceptable in a screening context that prioritizes the identification of more active cases. This case study illustrates that probability calibration may significantly enhance clinical applicability without requiring model retraining.

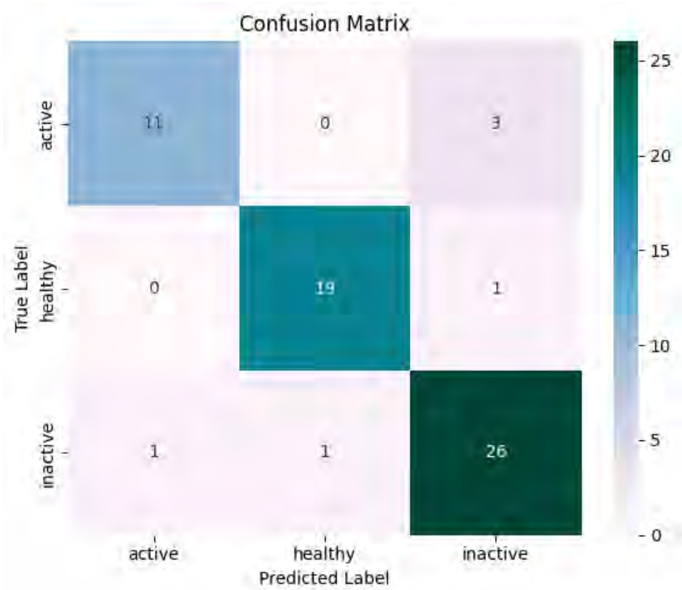

Figure 4.3: Confusion Matrix of the Proposed Model before Threshold Tuning

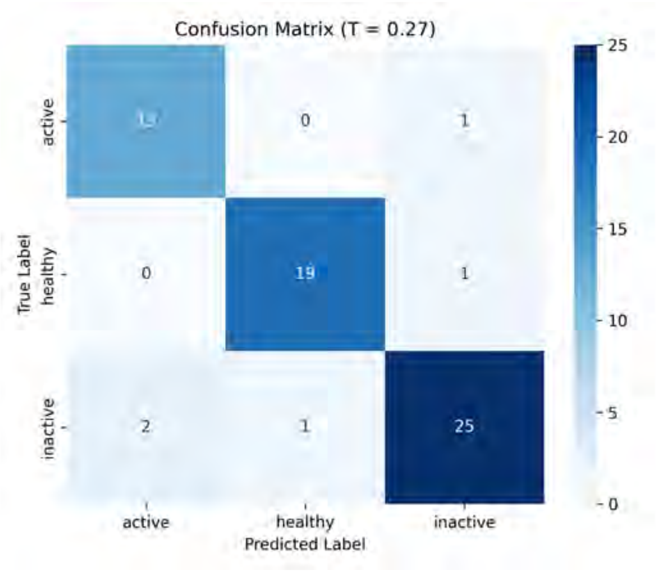

Figure 4.4: Confusion Matrix of the Proposed Model after Threshold Tuning

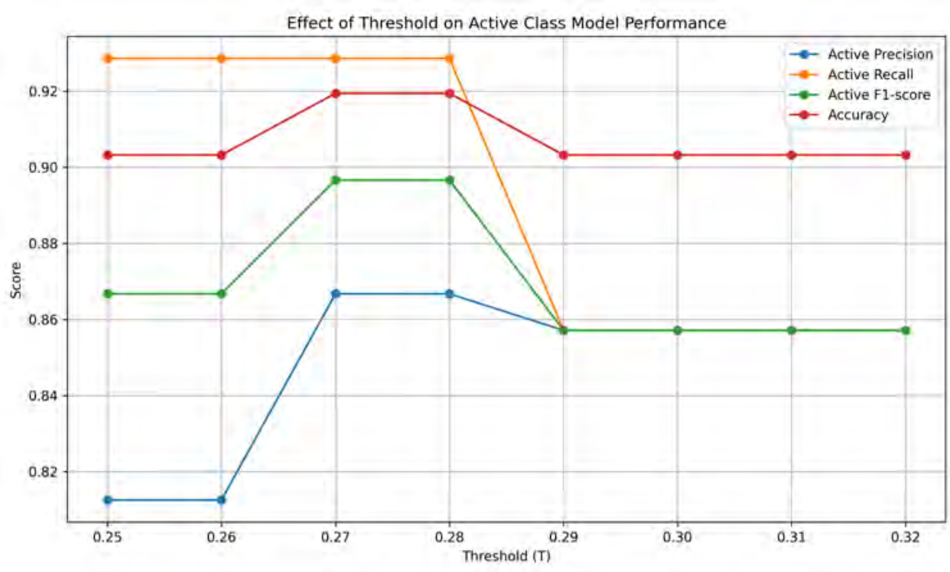

Figure 4.5: Effect of Threshold Tuning on Active Class Model Performance Diagram

4.4 ROC Curves and AUC Analysis

Figure 4.6 presents the Receiver Operating Characteristic (ROC) curves for each class to clarify the discriminative ability of the DiNet-Fusion model. All three classes demonstrate elevated AUC values, such as Healthy: 0.999, Active: 0.978, Inactive: 0.976, and Micro-average: 0.980. The elevated AUC scores signify outstanding separability among classes. The robustness of the feature-level fusion approach is demonstrated by the healthy class's near-perfect AUC value. The strong AUC values for both active and inactive classes validate the model's ability to differentiate subtle lesion variations, even in limited data environments.

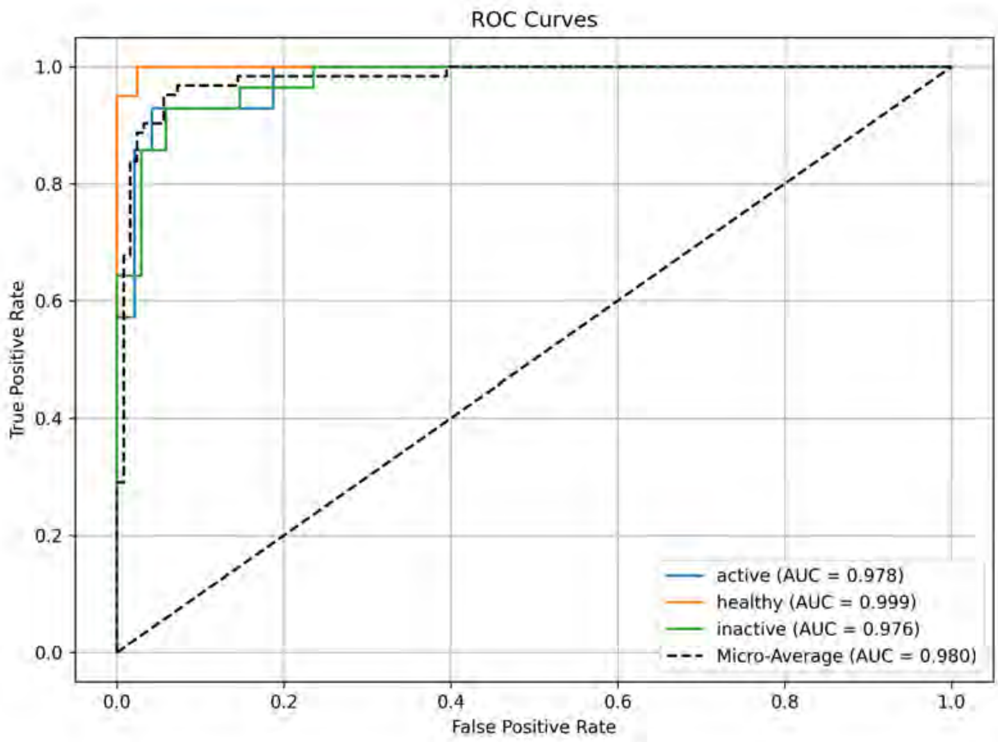

Figure 4.6: ROC Curve of the Proposed Model After Threshold Tuning

4.5 Softmax Probability Distribution

In Figure 4.7, a plot of the softmax probability distribution was created to assess how confidently the model makes predictions across all classes. The distribution displays two prominent peaks: one close to 0 (indicating incorrect or low-confidence predictions) and another near 1 (representing correct or high-confidence predictions). Furthermore, active samples exhibit a slightly broader distribution than healthy and inactive samples, indicating an increase in visual variability for active lesions. This distribution clarifies the advantages of threshold tuning: it helps in recovering borderline active cases in the mid-probability region.

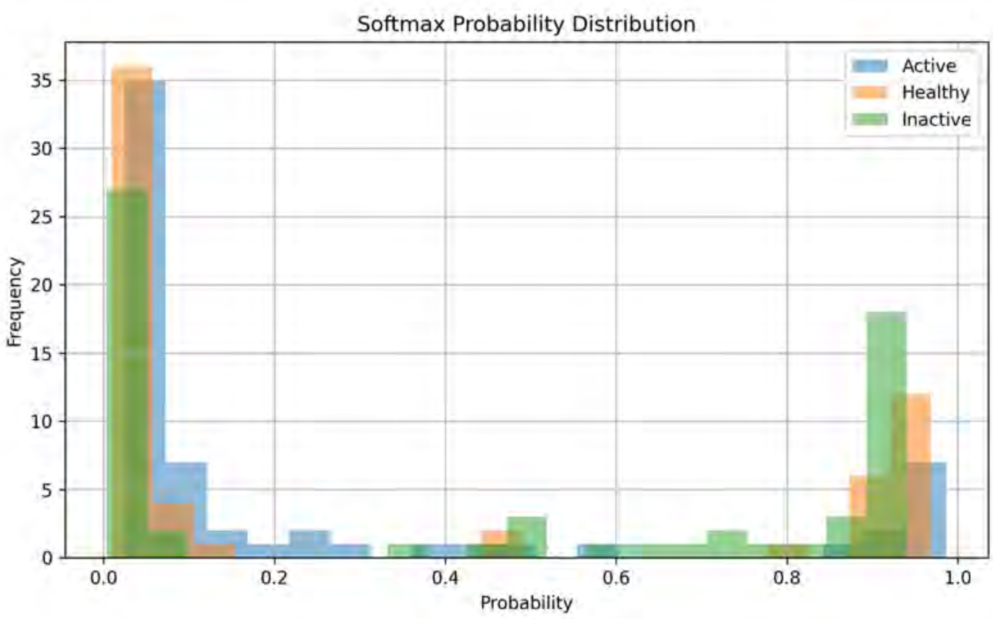

Figure 4.7: Softmax Probability Distribution Diagram

4.6 Error Analysis

Figure 4.8 shows that DiNet-Fusion correctly classified 57 out of the 62 samples after threshold tuning, with only a few mistakes. The majority of errors are due to active vs inactive class confusion, which correlates with clinicopathologic reports that scars can look similar to the early active lesion. More importantly, the number of missed active cases is greatly reduced after tuning the threshold, demonstrating the importance of a conservative threshold for activation.

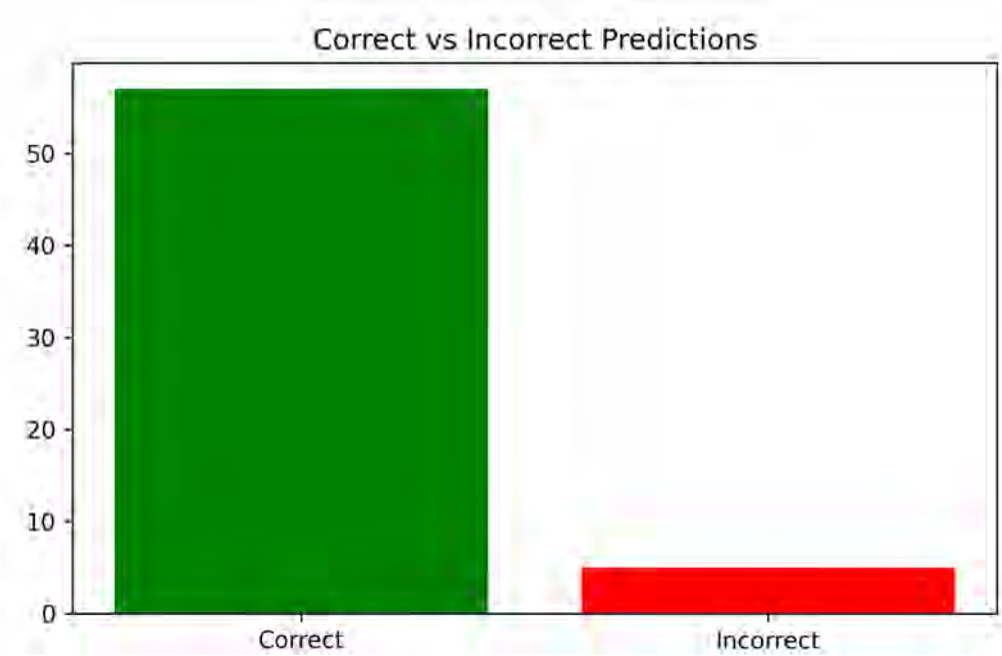

Figure 4.8: Correct (True Positive and True Negative) vs Incorrect (False Positive and False Negative) Prediction Diagram

4.7 Comparative Study

Table 4.3: Comparison of Our Other Approaches with Our Proposed Model

Model	Class	Precision (%)	Recall (%)	F1_Score (%)
MobileNet with LLM augmented fine-tuning	Active	80	89	84
	Inactive	93	65	79
	Healthy	78	100	88
	Overall Accuracy	83		
LLM+DINO+10 Shot Fusion	Active	51	75	60
	Inactive	84	58	68
	Healthy	79	89	84
	Overall Accuracy	72		
MaxVit+DINO	Active	73	100	84
	Inactive	95	67	79
	Healthy	82	95	88
	Overall Accuracy	83		
DiNet-Fusion (Proposed Model)	Active	86	92	89
	Inactive	92	89	90
	Healthy	95	95	95
	Overall Accuracy	92		

Table 4.4: Comparison of Existing Models with Our Proposed Model

Paper Title	**Models Used**	**Evaluation Results**	**Binary or Multi-Class**
Alam et al. [25]	VGG16, MobileNetV2, InceptionV3, ResNet50, DenseNet121	Accuracy: 98.9% (MobileNetV2)	Binary (classification)
Chakravarthy et al. [11]	CNN	Accuracy: 92.5%	Binary (classification)
Noguera et al. [16]	ResNet-50	Accuracy: 86.7%, Recall: 91.2%	Multi-class (Healthy / Active / Inactive)
Our Proposed Model	DiNet-Fusion (EfficientNetB0, DINO ViT-Small)	Accuracy: 92% Active (F1: 89.66%, Recall: 92.86%, P: 86.67%) Inactive (F1: 90.91%, Recall: 89.29%, P: 92.59%) Healthy (F1: 95%, Recall: 95%, P: 95%)	Multi-class (Healthy / Active / Inactive)

4.8 Discussion

This study aimed to propose an effective and data-efficient strategy for automatic classification of OT based on a limited dataset. Previous experience with small CNNs already revealed a strong decrease of performance in limiting the number of training samples, thus showing that standard deep learning models are unable to generalize well when very few data are available. To overcome this limitation, the proposed DiNet-Fusion model fuses EfficientNet-B0 and self-supervised DINO ViT-Small features at the feature-fusion level. This hybrid can provide both fine-grained local detail and global context information for fundus images, ultimately achieving a more balanced and robust classification in all three OT categories.

The experimental findings indicate that DiNet-Fusion excels in low-resource environments, surpassing conventional CNN-only benchmarks in multi-class scenarios. Weighted sampling and threshold tuning efficiently address class imbalance in the model, increasing recall for the clinically significant active class. Such findings suggest that the proposed fusion-based architectures can highly improve classification quality when large annotated datasets are not available. The results verify that self-supervised transformer representations with efficient convolutional features provide an effective and applicable approach for real-world medical imaging contexts.

Chapter 5

Limitations and Future Work

5.1 Conclusion

This study presents DiNet-Fusion, a fusion-based classification model designed for low-resource Ocular Toxoplasmosis datasets, the aim was to build an architecture which can perform better even after low datasets. By integrating features from EfficientNet-B0 and DINO ViT-Small, the model demonstrates great generalisation, high discriminative capacity, and stability, even with limited datasets because EfficientNet-B0 captures local lesion details and DINO ViT-Small understands the global context of the entire retinal structure, the model can learn powerful features even with limited data. The results further confirm that hybrid CNN–transformer architectures are able to fill the performance gap due to small dataset size and class imbalance, providing more stable predictions than conventional deep models, which shows that the variability seen when using only CNN or only Transformer is much less in the case of fusion models and more consistent results are obtained across all three classes. Additionally, the proposed method holds promise as a simple diagnostic support tool targeting the detection of Ocular Toxoplasmosis in minimal conditions, especially where there is a shortage of physicians or specialists, it can provide rapid decisions as a primary screening tool, and its integration into clinical practice is possible, such as in regions with limited access to professional ophthalmic assessment which further increases the likelihood of acceptance in real-world medical settings in the future.

5.2 Limitations

The DiNet-Fusion model uses a single imaging procedure, which is fundus images. The lesions does not have clear classification. This model depends on 224x224, which may be exclude small definitions from the lesions.

5.3 Future Work

The suggested model exhibits robust performance; yet, numerous avenues for future enhancement remain, like additional imaging methods such as OCT or fluorescein angiography, which might enhance diagnostic accuracy. Lesion localization or explainability techniques (e.g., Grad-CAM and LIME) could be further incorporated to help medical experts understand model predictions. More advanced fusion strategies, such as attention-based fusion or cross-modal transformers can be considered to achieve better performance. Deployment-driven optimizations of the model, such as pruning, quantization, or mobile responsiveness are expected to make this method practical for real-time OT assessment in low-resource healthcare settings. By further pursuing along these directions, DiNet-Fusion can become a basis for extending to general applications in automated retinal disease diagnosis and may be adapted to identify other rare or underrepresented eye diseases for which annotated data are very limited.

Bibliography

[1] S. M. Nikam and C. Y. Patil, "Glaucoma detection from fundus images using MATLAB GUI," in *2017 3rd International Conference on Advances in Computing, Communication & Automation (ICACCA) (Fall)*, Dehradun, India, 2017, pp. 1–4. DOI: 10.1109/ICACCAF.2017.8344700.

[2] S. C. Patel and M. I. Patel, "Analysis of CDR of fundus images for glaucoma detection," in *2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI)*, Tirunelveli, India, 2018, pp. 1071–1074. DOI: 10.1109/ICOEI.2018.8553707.

[3] A. D. Chakravarthy et al., "An approach towards automatic detection of toxoplasmosis using fundus images," in *2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)*, Athens, Greece, 2019, pp. 710–717. DOI: 10.1109/BIBE.2019.00134.

[4] M. T. Islam, S. A. Imran, A. Arefeen, M. Hasan, and C. Shahnaz, "Source and camera independent ophthalmic disease recognition from fundus image using neural network," in *2019 IEEE International Conference on Signal Processing, Information, Communication and Systems (SPICSCON)*, IEEE, 2019, pp. 59–63. DOI: 10.1109/SPICSCON48833.2019.9065162. [Online]. Available: https://doi.org/10.1109/SPICSCON48833.2019.9065162.

[5] D. Abeyrathna et al., "Directed fine tuning using feature clustering for instance segmentation of toxoplasmosis fundus images," in *2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)*, Cincinnati, OH, USA, 2020, pp. 767–772. DOI: 10.1109/BIBE50027.2020.00130.

[6] M. K. Behera, S. Chakravarty, A. Gourav, and S. Dash, "Detection of nuclear cataract in retinal fundus image using radial basis function based SVM," in *2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC)*, Waknaghat, India, 2020, pp. 278–281. DOI: 10.1109/PDGC50313.2020.9315834.

[7] T. H. P. Thanh, T. P. T. Thuy, T. N. Hieu, and M. S. Nguyen, "A real-time classification of glaucoma from retinal fundus images using AI technology," in *2020 International Conference on Advanced Computing and Applications (ACOMP)*, Quy Nhon, Vietnam, 2020, pp. 114–121. DOI: 10.1109/ACOMP50827.2020.00024.

[8] R. Parra et al., "A trust-based methodology to evaluate deep learning models for automatic diagnosis of ocular toxoplasmosis from fundus images," *Diagnostics*, vol. 11, no. 11, p. 1951, 2021. DOI: 10.3390/diagnostics11111951.

[9] R. Parra et al., "Automatic diagnosis of ocular toxoplasmosis from fundus images with residual neural networks," in *Studies in Health Technology and Informatics*, 2021. DOI: 10.3233/SHTI210143.

[10] R. Parra et al., *Ocular toxoplasmosis fundus images dataset (3.0.0)*, Zenodo, 2021. DOI: 10.5281/zenodo.13883485. [Online]. Available: https://doi.org/10.5281/zenodo.13883485.

[11] R. Rokhana, W. Herulambang, and R. Indraswari, "Classification of ocular diseases on fundus images using weighted MobileNetV2," in *2022 International Electronics Symposium (IES)*, Surabaya, Indonesia, 2022, pp. 570–575. DOI: 10.1109/IES55876.2022.9888652.

[12] S. Ferdous, M. Rifat, M. Ayan, and R. Rahman, "An approach to classify ocular toxoplasmosis images using deep learning models," in *ICCIT 2023*, 2023, pp. 1–6. DOI: 10.1109/ICCIT60459.2023.10441263.

[13] V. Gupta, A. Jaiswal, T. Choudhury, and N. Sachdeva, "Cataract detection on ocular fundus images using machine learning," in *Emerging Technologies in Data Mining and Information Security*, ser. Lecture Notes in Networks and Systems, vol. 490, Singapore: Springer, 2023. DOI: 10.1007/978-981-19-4052-1_20.

[14] M. Hassan et al., "Utilization of automated deep learning approach toward detection of ocular toxoplasmosis using fundus photographs," *Investigative Ophthalmology & Visual Science*, vol. 64, no. 8, p. 1093, 2023.

[15] J. M. Jerry, T. Thomas, R. Pugalenthi, and T. Thomas, "Automatic classification of retinal fundus images for diabetic retinopathy detection," in *2023 International Conference on Recent Advances in Electrical, Electronics, Ubiquitous Communication, and Computational Intelligence (RAEEUCCI)*, Chennai, India, 2023, pp. 1–5. DOI: 10.1109/RAEEUCCI57140.2023.10134335.

[16] J. L. V. Noguera et al., "Multiclass classification of ocular toxoplasmosis from fundus images with residual neural networks," in *2023 18th Iberian Conference on Information Systems and Technologies (CISTI)*, Aveiro, Portugal, 2023, pp. 1–5. DOI: 10.23919/CISTI58278.2023.10211306.

[17] R. Tiwari, H. Maheshwari, V. Gautam, and A. Agarwal, "Enhancing toxoplasmosis chorioretinitis detection: A hybrid resnet-yolo classifier with fitness sorted-shark smell optimization," in *SMART 2023*, 2023, pp. 77–82. DOI: 10.1109/SMART59791.2023.10428594.

[18] E. Yilmaz, "A modified efficientnet model for the classification of toxoplasmosis disease on OCT images," in *2023 Innovations in Intelligent Systems and Applications Conference (ASYU)*, Sivas, Turkiye, 2023, pp. 1–4. DOI: 10.1109/ASYU58738.2023.10296557.

[19] S. S. Alam, S. B. Shuvo, S. N. Ali, F. Ahmed, A. Chakma, and Y. M. Jang, "Benchmarking deep learning frameworks for automated diagnosis of ocular toxoplasmosis: A comprehensive approach to classification and segmentation," *IEEE Access*, vol. 12, pp. 22 759–22 777, 2024. DOI: 10.1109/ACCESS.2024.3359701.

[20] P. K. Choudhury et al., *Deep learning based automated diagnosis of ocular toxoplasmosis in fundus images using convolutional neural network*, BRAC University, 2024. [Online]. Available: http://hdl.handle.net/10361/22916.

[21] M. Dragun and R. Roszczyk, "Neural network-based approach to eye disease classification from fundus images," in *2024 25th International Conference on Computational Problems of Electrical Engineering (CPEE)*, Stronie Śląskie, Poland, 2024, pp. 1–4. DOI: 10.1109/CPEE64152.2024.10720421.

[22] P. Kulkarni, S. N. Habib, A. P. Bidargaddi, S. N. B., C. U. M., and M. S. M., "Optimizing few-shot learning via reptile meta-learning approach for toxoplasmosis chorioretinitis detection," in *2024 5th International Conference for Emerging Technology (INCET)*, Belgaum, India, 2024, pp. 1–7. DOI: 10.1109/INCET61516.2024.10593615.

[23] R. M. and S. Narayanan, "Deep learning of fundus images and optical coherence tomography images for ocular disease detection – a review," *Multimedia Tools and Applications*, vol. 83, pp. 88 745–88 789, 2024. DOI: 10.1007/s11042-024-18938-x.

[24] D. Milad, F. Antaki, A. Bernstein, S. Touma, and R. Duval, "Automated machine learning versus expert-designed models in ocular toxoplasmosis: Detection and lesion localization using fundus images," *Ocular Immunology and Inflammation*, vol. 32, no. 9, pp. 2061–2067, 2024. DOI: 10.1080/09273948.2024.2319281.

[25] A. Mulyani, Muljono, Purwanto, and M. A. Soeleman, "Fundus image classification for detecting diabetic retinopathy using convolutional neural network with Adam optimizer," in *2024 International Conference on ICT for Smart Society (ICISS)*, Bandung, Indonesia, 2024, pp. 1–6. DOI: 10.1109/ICISS62896.2024.10751509.

[26] H. Sharma, J. Wasim, and P. Sharma, "Analysis of eye disease classification by fundus images using different machine/deep/transfer learning techniques," in *2024 4th International Conference on Innovative Practices in Technology and Management (ICIPTM)*, Noida, India, 2024, pp. 1–6. DOI: 10.1109/ICIPTM59628.2024.10563429.