International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7031
***Article Received**: 01 April 2023 **Revised:** 18 May 2023 **Accepted:** 08 June 2023*

Enhancing Retinal Scan Classification: A Comparative Study of Transfer Learning and Ensemble Techniques

Dr. Ajitkumar Shitole[1], Aryan Kenchappagol[2], Rutuja Jangle[3], Yashowardhan Shinde[4], Akalbir Singh Chadha[5]
[1]Associate Professor – Department of Computer Engineering
Hope Foundation's International Institute of Information Technology
Pune, India 411057
ajitkumars@isquareit.edu.in
[2]Department of Computer Engineering
Hope Foundation's International Institute of Information Technology
Pune, India 411057
aryan.kenchappagol@gmail.com
[3]Department of Computer Engineering
Hope Foundation's International Institute of Information Technology
Pune, India 411057
rutujarjangle@gmail.com
[4]Department of Computer Engineering
Hope Foundation's International Institute of Information Technology
Pune, India 411057
yashowardhanshinde@gmail.com
[5]Department of Computer Engineering
Hope Foundation's International Institute of Information Technology
Pune, India 411057
chadhaakalbirsingh@gmail.com

Abstract—Ophthalmic diseases are a significant health concern globally, causing visual impairment and blindness in millions of people, particularly in dispersed populations. Among these diseases, retinal fundus diseases are a leading cause of irreversible vision loss, and early diagnosis and treatment can prevent this outcome. Retinal fundus scans have become an indispensable tool for doctors to diagnose multiple ocular diseases simultaneously. In this paper, the results of a variety of deep learning models (DenseNet-201, ResNet125V2, XceptionNet, EfficientNet-B7, MobileNetV2, and EfficientNetV2M) and ensemble learning approaches are presented, which can accurately detect 20 common fundus diseases by analyzing retinal fundus scan images. The proposed model is able to achieve a remarkable accuracy of 96.98% for risk classification and 76.92% for multi-disease detection, demonstrating its potential for use in clinical settings. By utilizing the proposed model, doctors can provide swift and accurate diagnoses to patients, improving their chances of receiving timely treatment and preserving their vision.

Keywords-Deep Learning, Ensemble Learning, EfficientNetV2M, Retinal scans, Transfer Learning

I. INTRODUCTION

Corneal disease is the second most common cause of reversible blindness worldwide, following cataracts. In India, approximately 6.8 million people and 3.2 million people in China suffer from corneal blindness in at least one eye. Timely detection and effective treatment can prevent vision loss caused by corneal disease. The diagnosis of corneal disease relies heavily on the expertise of clinicians who examine the cornea and conjunctiva using a specialized microscope called a slit-lamp. However, this process is time-consuming and can lead to variations in diagnosis among different observers. To address these challenges, automatic grading of medical photographs can be employed. This approach reduces the workload on physicians, enhances the efficiency and consistency of screening programs, and improves patient outcomes by enabling early diagnosis and treatment [1, 2]. The main objective in this context is to identify eye diseases at an early stage and provide appropriate treatment to maintain vision and enhance quality of life. Integrating artificial intelligence (AI) extensively into the field of ophthalmology can be advantageous for achieving these goals. AI technology can accelerate the diagnostic process and minimize the requirement for human resources. AI, a field of computer

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7031
***Article Received**: 01 April 2023 **Revised:** 18 May 2023 **Accepted:** 08 June 2023*

science, focuses on developing algorithms that imitate human intelligence [3]. It can be difficult for doctors to detect eye diseases early enough using fundus photographs. Hand diagnosing eye disorders is time-consuming, error-prone, and complex. An automated ocular disease detection system equipped with computer-aided tools is necessary for the detection of diverse eye conditions through the utilization of fundus images. The feasibility of a system that utilizes deep learning algorithms to enhance image classification capabilities has been enhanced [4]. Deep learning (DL) holds significant potential in the field of ophthalmology, particularly for screening diseases such as diabetic retinopathy (DR) and retinopathy of prematurity (ROP), which have established protocols. Additionally, disorders like glaucoma and age-related macular degeneration (AMD) may require regular screening and monitoring. However, the process of screening requires substantial manpower and financial resources for healthcare systems, regardless of whether they are in developed or developing countries. To address this challenge, combining DL with telemedicine can offer a sustainable solution for screening and monitoring patients in primary eye care settings in the long run [5]. This study summarizes new DL systems for ophthalmology applications, prospective clinical implementation issues, and potential future directions.

II. LITERATURE REVIEW

This section covers literature review of four main topics. Transfer Learning, Ensemble Learning, State-of-the-Art (SOTA) CNN models and Approaches to solve ophthalmic disease detection problem.

Tan, C et al. examine the advantages of deep learning across different areas, but emphasize the difficulty of gathering and annotating extensive datasets in specific fields. To overcome this challenge, transfer learning is introduced as a solution that allows for the use of previously trained models to improve performance on new datasets with limited data. The article reviews recent research on transfer learning using deep neural networks and its applications. The focus is on deep transfer learning approaches and their potential to address the problem of insufficient training data [6]. In a study by Wang et al., the challenge of training deep convolutional neural networks with limited datasets is addressed, and the concept of transfer learning is explored as a solution. However, blindly transferring all learned features from one dataset to another can lead to unnecessary computations and reduced performance on the target task. To overcome this, the authors propose Attentive Feature Distillation and Selection (AFDS) technique that modifies the regularization of transfer learning while also identifying which features should be transferred dynamically. When AFDS is applied to ResNet-101, it achieves state-of-the-art computation reduction while maintaining accuracy. In fact, a ResNet-101 model with AFDS, trained on ImageNet and then applied to Stanford Dogs 120 dataset, outperforms other transfer learning algorithms while operating within a reduced computation budget of 10 times fewer multiply-accumulate operations (MACs) [7]. Rezaei, et al. research talks about the security risks associated with transfer learning, which involves using a pre-trained model for a new task with a small dataset. The pre-trained model is often publicly available, which can be exploited by attackers to create instances of input that can trigger each target class with high confidence. The authors propose a brute force attack that is target-agnostic and does not require any target-specific information, such as re-trained models or SoftMax probability assigned to each class. The attack's effectiveness is demonstrated in face and speech recognition tasks, revealing a fundamental security flaw in the SoftMax layer when used in transfer learning scenarios [8].

The authors study how ensemble learning and deep learning are superior to standard algorithms in machine learning. Ensemble learning combines various models to create a more effective model, while deep learning uses complex architectures to enhance predictive accuracy. Tuning the ideal hyper-parameters in deep learning can be time-consuming, so there have been recent efforts to approach ensemble learning through deep learning. The article provides an extensive evaluation of various ensemble methodologies, especially deep learning, and categorizes research projects that use ensemble learning in various fields [9]. Su et al. present TPE-DEM, an adaptive deep ensemble learning method designed to address challenges related to interpreting complex machine learning models in diagnosis and the performance variations caused by differences in data. TPE-DEM combines the Deep Ensemble Model with the tree-structured Parzen Estimator to aggregate simpler models that are more comprehensible to physicians and require less training data. The proposed model determines the optimal number of layers and basic learners based on the distribution of data and characteristics of the diagnostic task, resulting in improved performance compared to other baseline models. This approach offers a novel solution for developing straightforward and interpretable machine learning models in computer-aided diagnosis tasks that involve diverse datasets and feature sets [10]. Ensemble learning algorithms are widely used in medical image classification pipelines to combine different models and improve prediction performance. The effectiveness of these algorithms in deep learning-based pipelines for categorizing medical images remains uncertain. This study introduces a medical image classification pipeline to explore the influence of ensemble learning techniques, such as

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7031
***Article Received**: 01 April 2023 **Revised:** 18 May 2023 **Accepted:** 08 June 2023*

augmentation, stacking, and bagging, on performance. The pipeline was tested on four medical imaging datasets and demonstrated that stacking achieved the highest performance boost, with up to a 13% rise in F1-score. Augmenting and bagging also exhibited considerable performance gains, and basic statistical pooling functions were found to be as good as, if not better than, more sophisticated ones. The study concluded that incorporating ensemble learning techniques is a powerful way to improve resilience and performance in any medical picture classification pipeline [11]. E. TaSci and A. Ugur present a new method for image classification that combines pre-trained convolutional neural networks with hand-crafted features. The proposed method uses four hand-crafted features and 4096 deep learning features from the CIFAR-10 dataset, and the classification accuracy rate is used to evaluate the system performance. The results of the experiment demonstrate that the combination of hand-crafted and deep learning features outperforms using only deep learning features [12].

The work done by Naseer I, Akram S, et al. evaluates and compares several convolutional neural network (CNN) architectures such as LeNet, AlexNet, VGG16, ResNet-50, and Inception-V1 using public LUNA16 datasets for identifying lung cancer. Different performance optimizers such as RMSProp, Adam, and SGD were used, and the accuracy, sensitivity, specificity, and other measures were assessed. AlexNet with the SGD optimizer achieved the highest validation accuracy for CT lung cancer, outperforming other CNN designs, with an accuracy of 97.42%, a sensitivity of 97.58%, a specificity of 97.25%, and an F1 score of 97.58% [13]. A. A. Ardakan et al. propose an AI-based method for quick and accurate COVID-19 diagnosis using CT scans. The authors used ten different convolutional neural networks to distinguish between COVID-19 and non-COVID-19 cases, and ResNet-101(sensitivity, 100%; specificity, 99.02%; accuracy, 99.51%) and Xception (sensitivity of 98.04%, specificity of 100%, and accuracy of 99.02%) achieved the best performance. The radiologist's performance was moderate compared to the AI models. ResNet-101 can be used as an adjuvant tool in radiology departments for characterizing and diagnosing COVID-19 infections with high sensitivity [14]. The authors Porag, Al Mohidur Rahman, et al. focus on the development of a system for automatically detecting bacterial and viral pneumonia in digital X-ray images, which is crucial for early diagnosis and treatment. The researchers used deep convolutional neural networks and compared different architectures, including VGG19, ResNet with 152v2, Resnext 101, Seresnet 152, MobileNetv2, and DenseNet with 201 layers, and found that DenseNet201 performed the best, achieving 95% testing accuracy with substantially fewer parameters and in an acceptable computing time. The authors also discuss recent achievements in trustworthy pneumonia diagnosis and provide an empirical comparison of deep learning architectures for plant disease classification based on photographs [15].

K.Prasad et.al. talks about how India has 15 million blind people, and 75% of them could have been treated. Diabetic retinopathy and glaucoma are the primary causes of blindness in India, and the doctor-patient ratio is 1:10,000. As these disorders are asymptomatic in their early stages, detection is difficult, and if left untreated, they can cause irreversible eyesight impairment. The proposed deep neural network model can aid in the early detection of diabetic retinopathy and glaucoma by notifying people to see an ophthalmologist for screening. The model is accurate up to 80%, less complex, and can help address the challenges of blindness in India [16]. Topaloglu, Ismail, et al. implemented a deep learning-based convolutional artificial neural network approach for image classification, with a sample application carried out for diabetic retinopathy. The care model is a method that includes the process of rescaling data prior to generating an average data pool. Additionally, it involves multiplying all data by the number of elements by the number of epoch time eight tensors. The proposed model is a combination of the VGG-19 image classification model and a mathematical model that has been developed. The model achieved a train accuracy of 87%, test accuracy of 88%, precision of 93%, and recall of 83%. The pre-trained model and image datasets were taken from Kaggle and Keras for the case study [17]. Diabetic patients are at a higher risk of developing eye illnesses such as diabetic retinopathy (DR), diabetic macular edema (DME), and glaucoma, which can be difficult to detect in the early stages. Ali Javed et.al. developed an automated system to assist in early detection and screening. The research proposes an automated disease localization and segmentation method using the Rapid Region-based Convolutional Neural Network (FRCNN) algorithm and fuzzy k-means (FKM) clustering. The FRCNN requires bounding-box annotations, which were produced using ground-truths since the datasets did not offer them. The approach was evaluated using several datasets, and a comparison with recent methods confirmed its efficacy in disease diagnosis and segmentation [18]. DenseNet is a deep neural network that has shown excellent performance on image classification tasks. A study by Liu et al. (2020) used a pre-trained DenseNet model to classify four different types of retinal diseases. The authors fine-tuned the last few layers of the network using a dataset of retinal scans and achieved an accuracy of 96.54% for the detection of age-related macular degeneration, diabetic retinopathy, glaucoma, and hypertensive retinopathy [19]. ResNet is another popular deep neural network architecture

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7031
***Article Received**: 01 April 2023 **Revised:** 18 May 2023 **Accepted:** 08 June 2023*

that has been used for image classification tasks. A study by Rajalakshmi et al. (2018) used a pre-trained ResNet model to classify diabetic retinopathy using retinal scans. The authors fine-tuned the last few layers of the network using a dataset of retinal scans and achieved an accuracy of 86.2% for the detection of diabetic retinopathy [20]. EfficientNet is a relatively new deep neural network architecture that has shown state-of-the-art performance on image classification tasks. A study by Zebin et al. (2020) used a pre-trained EfficientNet model to classify five different types of retinal diseases. The authors fine-tuned the last few layers of the network using a dataset of retinal scans and achieved an accuracy of 97.11% for the detection of age-related macular degeneration, diabetic retinopathy, glaucoma, hypertensive retinopathy, and myopia [21].

III. DATASET DESCRIPTION

The Drive dataset (RFMiD) has been considered for all experiments in this paper [22]. This Dataset was taken from a global challenge named "Retinal Image Analysis for multi-Disease Detection Challenge" that was hosted on Biomedical Imaging Platform. It is a public Dataset consisting of 3200 retinal fundus images collected from various sources which are categorized into 46 (27 main disease categories + other category) ocular diseases [22]. The images were captured using three retinal fundus cameras namely: TOPCON 3D OCT-2000 (each image size is 2144 × 1424 pixels), Kowa VX-10 (each image size is 4288 × 2848 pixels) and TOPCON TRC-NW300 (image resolution is 2048 × 1536) [22]. The retinal fundus images are labeled based on the presence of diseases shown by the image. The dataset is split into the following subsets: Training set (60%) consisting of 1920 images, Validation set (20%) containing 640 images and Testing set (20%) consisting of 640 images. The main aim of this dataset is to help in multi disease prediction using the retinal scans.

IV. METHODOLOGY

Figure 1. Shows the flow of the entire experiment. The pipeline starts with image basic preprocessing followed by image augmentation, which is done in order to reduce the imbalance in the dataset. This is followed by model training using transfer learning, each model requires different preprocessing before the images are fed to the network. This is taken care of using an image pre-processing function specific to the model. Model training is followed by model evaluation where 4 metrics are tracked, Accuracy, F1-Score, Precision and Recall. Based on these metrics model selection is performed, 3 models with the highest metric values are selected. These models are finally put into an ensemble learning system to get the final output. 2 different ensemble techniques are used, Majority Voting and Weighted Average.

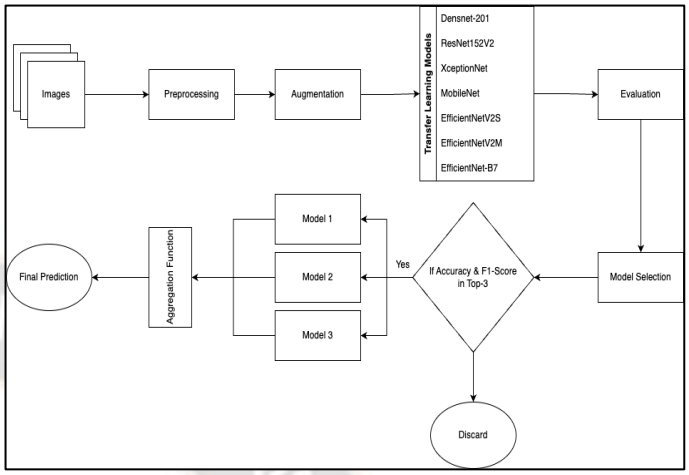

Figure 1. Methodology Used for Creating the Hybrid Model

Figure 1. describes the methodology and approach used for creating the hybrid model. The following concepts and techniques were used for implementing the models.

V. EXPERIMENTATION

To evaluate the effectiveness of retinal scans in detecting diseases, two types of experiments were conducted: binary classification and multi-class classification. For both types of experiments, image size and type of augmentation were experimented with to determine their impact on classification accuracy.

For the binary and multi-class classification experiments, retinal scans of patients with and without the disease were divided into training, validation, and testing sets with a 60:20:20 ratio. Three types of augmentations were applied to the images: random rotations, flips, and brightness adjustments. Image sizes were also changed, using 256x256, and 512x512 pixel sizes. The classification accuracy, f1-score, precision, recall, and AUC-ROC were calculated for each model, and the results were compared for different image sizes and types of augmentations. All the experiments were performed using the TensorFlow framework on NVIDIA V100 Tensor Core GPUs.

A. Dealing with Data Imbalance - Augmentation

During the sorting and segregation of data based on the different diseases, it was found that there were 8 different diseases that did not have enough images to be classified, so the images were combined into the "others" category. Now the dataset consists of 20 different diseases, compared to 28 diseases before. But during the exploration and training of the models, it was observed that the data was still

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7031
***Article Received**: 01 April 2023 **Revised:** 18 May 2023 **Accepted:** 08 June 2023*

imbalanced, and the models were biased. So, the technique of under sampling was implemented, but this technique did not give accurate results, so the implementation of oversampling the images using different augmentations like random rotations, flips, and brightness adjustments were applied as seen in Figure 2.

These augmentations were implemented with the help of the Augmentor library in Python [23]. The oversampling of the images helped improve the results as well as the accuracy of the models.

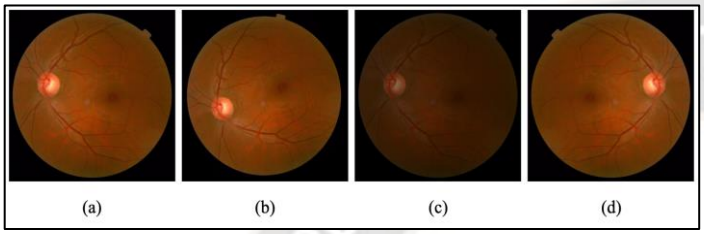

Figure 2. Different Augmentations used (a) original image, (b) rotated by 30 degrees, (c) lower brightness, and (d) horizontal flip.

The final dataset had the following categories: Age-related macular degeneration (ARMD), Asteroid Hyalosis (AH), Branch retinal vein occlusion (BRVO), Chorioretinitis (CRS), Central retinal vein occlusion (CRVO), Central serous retinopathy (CSR), Drusen (DN), Diabetic retinopathy (DR), Epiretinal membrane (ERM), Laser scars (LS), Media Haze (MH), Macular scar (MS), Myopia (MYA), Optic disc cupping (ODC), Optic disc edema (ODE), Optic disc pallor (ODP), Retinitis (RS), Retinal traction (RT), Tessellation (TSLN) and OTHER.

B. Model Training and Fine Tuning

For this research, transfer learning is used to train the SOTA CNN models. This technique has been used for both risk classification and multi disease classification. The models used during training were DenseNet 201 [24], ResNet152V2 [25], XceptionNet [26], EfficientNetV2 (S, M, L)[27], EfficientNetB7 [28], MobileNetV2 [29]. Each of these models were trained for 30 epochs. For the first 20 epochs the top layers were frozen and for the next 10 epochs the last few layers of the CNN were unfrozen to fine-tuning the model further. This technique helped in achieving a good performance for each of the models. For training the models Adam optimizer was used with an initial learning rate of 10e-4 and ReduceLROnPlateau was used to change the learning rate in case the model performance became stagnant. For Risk classification binary cross entropy loss was used and for multiclass classification categorical cross entropy loss was used. A batch size of 16 was used for all experiments.

Binary Cross Entropy is defined by the following expression:

$$L(y, \hat{y}) = -y\log(\hat{y}) - (1-y)\log(1-\hat{y}) \quad (1)$$

Here, y is the true label and $\hat{y}$ is the predicted label.

Categorical Cross Entropy is defined using the following expression:

$$L(y, \hat{y}) = -\sum_{i=1}^{C} yi \log(\hat{y}_i) \quad (2)$$

Here, y is the true label, $\hat{y}$ is the predicted probability distribution over C classes, and y_i and $\hat{y}_i$ are the i^{th} elements of the true and predicted distributions, respectively.

C. Types of ensemble learning:

Ensemble learning can be used to combine the predictions of multiple models to improve the accuracy of disease detection. There are several types of ensembles learning techniques, including majority voting and averaging.

1) Majority Voting

Majority voting is a technique that involves aggregating the predictions of multiple models and selecting the most predicted class label. For example, if three models predict that a retinal scan belongs to class A and two models predict that it belongs to class B, then the ensemble prediction would be class A. It can be represented using the following equation:

$$\hat{y} = arg\ max_{i \in \{1,2,\ldots,K\}} \left(\sum_{j=1}^{N} I(y_j^{(i)} = k)\right) \quad (3)$$

Here, $\hat{y}$ is the predicted label, K is the number of classes, N is the number of models or classifiers, $y_j^{(i)}$ is the predicted label for the j^{th} example of the i^{th} model.

2) Weighted Averaging

Weighted Averaging is a technique that involves aggregating the predictions of multiple models by taking their weighted average probability scores for each class label. It can be represented using the following equation:

$$\hat{y} = \frac{1}{\sum_{i=1}^{N} wi} \sum_{i=1}^{N} wiyi \quad (4)$$

Here, $\hat{y}$ is the predicted label, N is the number of models or classifiers and y_i is the predicted label for the i^{th} example by the i^{th} model and w_i is the accuracy or any other equivalent metric that denotes the importance of the i^{th} model prediction.

Overall, ensemble learning can improve the accuracy of disease detection in retinal scans by leveraging the predictions of multiple models. Both majority voting and averaging are effective ensemble techniques that can be used to combine the predictions of multiple models. Similarly, based on the above information, the following model is

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7031
***Article Received**: 01 April 2023 **Revised:** 18 May 2023 **Accepted:** 08 June 2023*

proposed for risk classification and disease detection via such methods.

D. Metrics:

In all, 5 different metrics were used to compare and evaluate the models. They are Accuracy, F1-Score, Precision, Recall and AUC-ROC. These metrics can be defined in terms of the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) as given below:

Accuracy:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

Accuracy measures the proportion of correctly classified examples out of all the examples in the dataset. A high accuracy indicates that the model is performing well overall.

F1-Score:

$$F1\text{-}Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (6)$$

The F1-score is the harmonic mean of precision and recall, and provides a balance between the two metrics. A high F1-score indicates that the model is performing well in both precision and recall.

Precision:

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

Precision is a metric that quantifies the ratio of correctly predicted positive instances among all the instances that were predicted as positive. A high precision value suggests that the model is adept at accurately predicting the positive class.

Recall:

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

Recall measures the proportion of correctly predicted positive examples out of all the positive examples in the dataset. A high recall indicates that the model is good at identifying the positive class.

AUC-ROC:

$$AUC\text{-}ROC = \frac{TPR \cdot (1 - FPR) + 0.5 \cdot FPR \cdot (1 - TPR)}{TPR + (1 - FPR)} \quad (9)$$

where,

$$TPR = \frac{TP}{TP + FN} \qquad FPR = \frac{FP}{FP + TN} \quad (10)$$

The AUC-ROC (Area Under the Receiver Operating Characteristic Curve) is a widely utilized metric for assessing the effectiveness of binary classification models. It represents the area under the curve of the ROC curve, which illustrates the true positive rate (TPR) plotted against the false positive rate (FPR) at various classification thresholds.

VI. RESULTS

The results for all the experiments are discussed in this section. The results for risk classification and multi-disease detection can be seen below.

TABLE I. RESULTS OF RISK CLASSIFICATION

Model Name	Image Size	Accuracy (%)	F1-Score	Precision	Recall	AUC-ROC
Drive Dataset (Multi-Class Classification)						
DenseNet 201 [24]	256, 256	94.09	0.8627	0.8197	0.9	0.8950
ResNet15 2V2 [25]	256, 256	94.40	0.8670	0.8227	0.8859	0.8943
Xception Net [26]	512, 512	95.50	0.9097	0.8699	0.9131	0.9259
EfficientN etv2m [27]	512, 512	94.71	0.8813	0.8817	0.8807	0.9127
EfficientN et-B7 [28]	256, 256	94.40	0.8714	0.8846	0.8583	0.9025
MobileNe tV2 [29]	256, 256	93.78	0.8522	0.7738	0.8410	0.8848
Ensemble 1	NA	96.78	0.9197	0.8657	0.9035	0.9102
Ensemble 2	NA	96.98	0.9207	0.8757	0.9135	0.9209

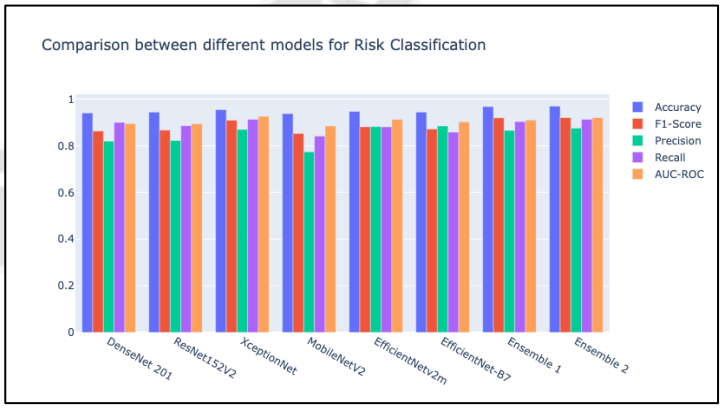

Figure 3. Comparison between all the models for risk classification

Table 1. Shows the results for binary classification. The Ensemble 1 method is the majority voting method and Ensemble Learning 2 method is the weighted average method. As seen in Table 1. For binary classification the Ensemble 2 performs best with metric scores of (accuracy =

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7031
***Article Received**: 01 April 2023 **Revised:** 18 May 2023 **Accepted:** 08 June 2023*

96.98%, f1-score = 0.9207, precision = 0.8757, recall = 0.9135). The models taken for ensemble systems are: XceptionNet [26], Resnet152V2 [25] and EfficientNetV2M [27], as they are the top performers. Both the ensemble methods show an improvement over the performance of individual models. Figure 3. gives a visual comparison between the models.

TABLE II. RESULTS FOR MULTI-CLASS CLASSIFICATION.

Model Name	Image Size	Accuracy (%)	F1-Score	Precision	Recall	AUC-ROC
Drive Dataset (Multi-Class Classification)						
DenseNet 201 [24]	256, 256	71.34	0.6624	0.6673	0.6582	0.6952
ResNet152 V2 [25]	256, 256	74.76	0.6812	0.6932	0.6724	0.7235
XceptionNet [26]	512, 512	75.72	0.7084	0.7243	0.6934	0.7402
EfficientNetv2m [27]	512, 512	75.12	0.7134	0.7440	0.6852	0.7396
EfficientNet-B7 [28]	256, 256	67.67	0.6323	0.6390	0.6261	0.6512
MobileNet V2 [29]	256, 256	68.43	0.6447	0.6473	0.6421	0.6581
Ensemble 1	NA	73.28	0.6932	0.7152	0.6735	0.7128
Ensemble 2	NA	76.92	0.7128	0.7191	0.7063	0.7808

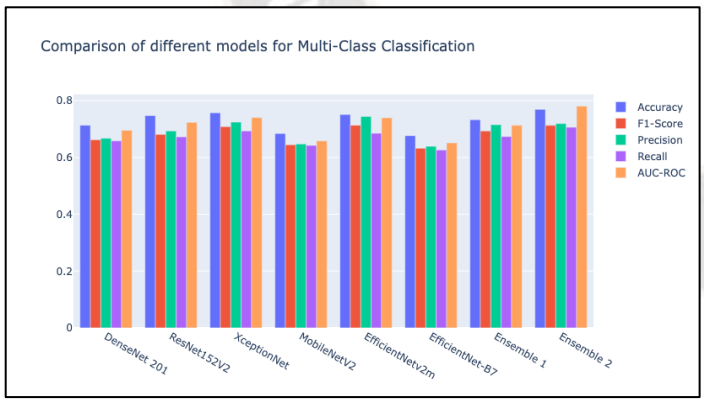

Figure 4. Comparison between all the models for multi-disease classification

Table 2. Shows the results for multi-disease classification. The Ensemble 1 method is the majority voting method and Ensemble Learning 2 method is the weighted average method. As seen in Table 2. Similar to risk classification for multi-disease classification the Ensemble 2 performs best with metric scores of (accuracy = 76.92%, f1-score = 0.7128, precision = 0.5579, recall = 0.7063). The models taken for ensemble systems are: XceptionNet [26], Resnet152V2 [25] and EfficientNetV2M [27] as they are the top performers. Here, it is noticed that only the Ensemble 2 method performs better than individual models and not Ensemble 1. Figure 4. gives a visual comparison between the models.

VII. FINAL PROPOSED MODEL

Based on the results of the experiment, the final proposed model is as shown in Figure 5. The scan is first passed to a preprocessing function specific to the model. The scan is then passed to the selected models. The models used in the final ensemble system are XceptionNet [26], ResNet152V2 [25] and EfficientNetV2M [27]. These models are selected based on their metric scores (accuracy, f1-score, precision, recall). The output of these models is then passed to a weighted average function that aggregates the predictions of the models and gives the final output.

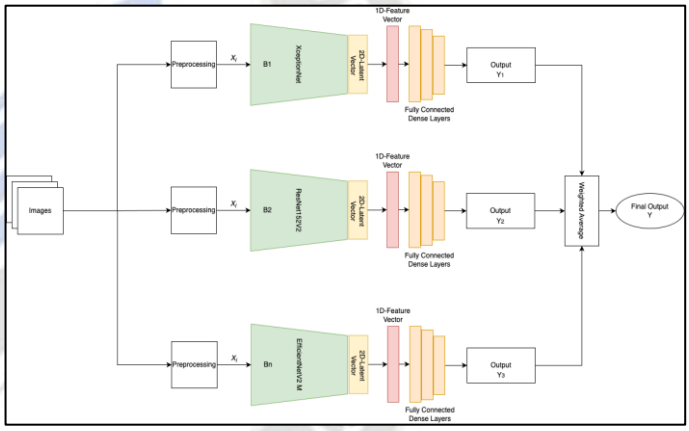

Figure 5. Proposed Ensemble Learning System Representation

VIII. CONCLUSION AND FUTURE SCOPE

In conclusion, this research paper provides an in-depth analysis of the application and performance of deep learning algorithms for retinal scan classification, with a focus on risk classification and disease detection. Additionally, the handling of data imbalance and fine-tuning of the models with appropriate epoch and batch sizes further enhanced the accuracy of the classification results. The results of the study demonstrate the effectiveness of transfer learning and ensemble learning in achieving high accuracy and performance in both use cases. The ensemble learning technique was employed to further enhance the accuracy of the classification results. Overall, the findings of this study demonstrate the potential of deep learning algorithms in retinal scan classification and their usefulness in the diagnosis of ocular diseases. The use of these techniques can

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7031
***Article Received**: 01 April 2023 **Revised:** 18 May 2023 **Accepted:** 08 June 2023*

provide an efficient and accurate method for diagnosis and can help in the early detection of ocular diseases.

Predicting ophthalmic diseases using retinal scans and deep learning models is a growing area of research with several potential future scopes, including Incorporating multimodal imaging: Future research could focus on incorporating other imaging modalities, such as OCT and visual field testing, to improve accuracy and reliability of predictions. Integration of IoT based portable devices: Future research could focus on development of IoT and cloud-based applications for faster computation and model training as well as maintenance [30]. Developing personalized medicine: Future research could focus on developing personalized medicine that tailors' treatment plans based on individual patient characteristics and disease status. Exploring ethical implications: The use of deep learning models for predicting ophthalmic diseases raises important ethical concerns around privacy, data security, and algorithmic bias [30]. Future research could explore these issues and develop ethical guidelines to ensure that these models are used safely and responsibly.

REFERENCES

[1] Gu, H., Guo, Y., Gu, L. et al. Deep learning for identifying corneal diseases from ocular surface slit-lamp photographs. Sci Rep 10, 17851 (2020). https://doi.org/10.1038/s41598-020-75027-3

[2] Cen, LP., Ji, J., Lin, JW. et al. Automatic detection of 39 fundus diseases and conditions in retinal photographs using deep neural networks. Nat Commun 12, 4828 (2021). https://doi.org/10.1038/s41467-021-25138-w

[3] Nuzzi R, Boscia G, Marolo P and Ricardi F (2021) The Impact of Artificial Intelligence and Deep Learning in Eye Diseases: A Review. Front. Med. 8:710329. doi: 10.3389/fmed.2021.710329

[4] Md Shakib Khan, Nafisa Tafshir, Kazi Nabiul Alam, Abdur Rab Dhruba, Mohammad Monirujjaman Khan, Amani Abdulrahman Albraikan, Faris A. Almalki, "Deep Learning for Ocular Disease Recognition: An Inner-Class Balance", Computational Intelligence and Neuroscience, vol. 2022, Article ID 5007111, 12 pages, 2022. https://doi.org/10.1155/2022/5007111

[5] Ting DSW, Pasquale LR, Peng L, et alArtificial intelligence and deep learning in ophthalmologyBritish Journal of Ophthalmology 2019;103:167-175.

[6] Tan, C.; Sun, F.; Kong, T.; Zhang, W.; Yang, C. & Liu, C. (2018), 'A Survey on Deep Transfer Learning' , cite arxiv:1808.01974Comment: The 27th International Conference on Artificial Neural Networks (ICANN 2018)

[7] Wang, K.,Gao, X.,Zhao, Y.,et al. Pay Attention to Features, Transfer Learn Faster CNNsC],2020.

[8] Rezaei, Shahbaz & Liu, Xin. (2019). A Target-Agnostic Attack on Deep Models: Exploiting Security Vulnerabilities of Transfer Learning.

[9] Ammar Mohammed, Rania Kora, A comprehensive review on ensemble deep learning: Opportunities and challenges, Journal of King Saud University - Computer and Information Sciences, Volume 35, Issue 2, 2023, Pages 757-774, ISSN 1319-1578, https://doi.org/10.1016/j.jksuci.2023.01.014.

[10] Su, K.; Wu, J.; Gu, D.; Yang, S.; Deng, S.; Khakimova, A.K. An Adaptive Deep Ensemble Learning Method for Dynamic Evolving Diagnostic Task Scenarios. Diagnostics 2021, 11, 2288. https://doi.org/10.3390/diagnostics11122288

[11] Müller, Dominik & Soto-Rey, Iñaki & Kramer, Frank. (2022). An Analysis on Ensemble Learning optimized Medical Image Classification with Deep Convolutional Neural Networks.

[12] E. TaSci and A. Ugur, "Image classification using ensemble algorithms with deep learning and hand-crafted features," 2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, Turkey, 2018, pp. 1-4, doi: 10.1109/SIU.2018.8404179.

[13] Naseer I, Akram S, Masood T, Jaffar A, Khan MA, Mosavi A. Performance Analysis of State-of-the-Art CNN Architectures for LUNA16. Sensors (Basel). 2022 Jun 11;22(12):4426. doi: 10.3390/s22124426. PMID: 35746208; PMCID: PMC9227226.

[14] A. A. Ardakani, A. R. Kanafi, U. R. Acharya, N. Khadem, and A. Mohammadi, "Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: results of 10 convolutional neural networks," Computers in Biology and Medicine, vol. 121, article 103795, 2020.

[15] Porag, Al Mohidur Rahman, Md. Mahedi Hasan and Dr. Md Taimur Ahad. "A Comparison Study of Deep CNN Architecture in Detecting Pneumonia." ArXiv abs/2212.14744 (2022): n. pag.

[16] K. Prasad, P. S. Sajith, M. Neema, L. Madhu and P. N. Priya, "Multiple eye disease detection using Deep Neural Network," TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON), Kochi, India, 2019, pp. 2148-2153, doi: 10.1109/TENCON.2019.8929666

[17] Topaloglu, Ismail. (2022). Deep Learning Based Convolutional Neural Network Structured New Image Classification Approach for Eye Disease Identification. Scientia Iranica. 10.24200/SCI.2022.58049.5537.

[18] Nazir, Tahira, Aun Irtaza, Ali Javed, Hafiz Malik, Dildar Hussain, and Rizwan Ali Naqvi. 2020. "Retinal Image Analysis for Diabetes-Based Eye Disease Detection Using Deep Learning," Applied Sciences 10, no. 18: 6185. https://doi.org/10.3390/app10186185

[19] Liu, J., Deng, Y., Wang, J., et al. (2020). Automated detection and classification of four common types of ocular diseases using deep learning. Frontiers in Bioengineering and Biotechnology, 8, 62.

[20] Rajalakshmi, R., Subashini, R., Arunprasath, V., et al. (2018). Performance of deep learning algorithms for detection of diabetic retinopathy using retinal images: A systematic review and meta-analysis. Eye, 32, 1138-1158.

[21] Zebin, T., Minaee, S., Abdolrahimzadeh, S., et al. (2020). Early detection of ocular diseases using deep learning on retinal OCT images. Medical Image Analysis, 65, 101765.

[22] Samiksha Pachade, Prasanna Porwal, Dhanshree Thulkar, Manesh Kokare, Girish Deshmukh, Vivek Sahasrabuddhe, Luca Giancardo, Gwenolé Quellec, Fabrice Mériaudeau, November 25, 2020, "Retinal Fundus Multi-disease Image Dataset (RFMiD)", IEEE Dataport, doi: https://dx.doi.org/10.21227/s3g7-st65.

[23] Marcus D Bloice, Peter M Roth, Andreas Holzinger, Biomedical image augmentation using Augmentor, Bioinformatics, Volume 35, Issue 21, 1 November 2019, Pages 4522–4524, https://doi.org/10.1093/bioinformatics/btz259

[24] Wang, Shuihua and Yudong Zhang. "DenseNet-201-Based Deep Neural Network with Composite Learning Factor and Precomputation for Multiple Sclerosis Classification." ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 16 (2020): 1 - 19.

[25] Dr. Naveen Jain. (2020). Artificial Neural Network Models for Material Classification by Photon Scattering Analysis. International Journal of New Practices in Management and Engineering, 9(03), 01 - 04. https://doi.org/10.17762/ijnpme.v9i03.88

[26] Nur Aziz Thohari, Afandi & Triyono, Liliek & Hestiningsih, Idhawati & Suyanto, Budi & Yobioktobera, Amran. (2022). Performance Evaluation of Pre-Trained Convolutional Neural Network Model for Skin Disease Classification. JUITA: Jurnal Informatika. 10. 9. 10.30595/juita.v10i1.12041.

[27] D. Shah, D. Shah, D. Jodhawat, J. Parekh and K. Srivastava, "Xception Net & Vision Transformer: A comparative study for Deepfake Detection," 2022 International Conference on Machine Learning, Computer Systems and Security (MLCSS), Bhubaneswar, India, 2022, pp. 393-398, doi: 10.1109/MLCSS57186.2022.00077.

[28] R. S. S. Devi, V. R. V. Kumar and P. Sivakumar, "Efficientnetv2 model for plant disease classification and pest recognition," Computer Systems Science and Engineering, vol. 45, no.2, pp. 2249–2263, 2023.

[29] Tan, Mingxing & Le, Quoc. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks.

[30] Sandler, Mark & Howard, Andrew & Zhu, Menglong & Zhmoginov, Andrey & Chen, Liang-Chieh. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. 4510-4520. 10.1109/CVPR.2018.00474.

[31] Paul Garcia, Ian Martin, Laura López, Sigurðsson Ólafur, Matti Virtanen. Automated Grading Systems: Advancements and Challenges. Kuwait Journal of Machine Learning, 2(1). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/165

[32] Ajitkumar Shitole, Dr. Manoj Devare, "Optimization of IoT Enabled Physical Location Monitoring using DT and VAR", International Journal of Cognitive Informatics and Natural Intelligence (IJCINI), ISSN: 1557-3958, Vol: 15, Issue: 4, Oct 2022.