International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 03 | March - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Comparative Evaluation of Pretrained CNN Models for Preliminary Eye Defect Diagnosis Using Fundus Images

Akinsola Adeniyi F.
Yaba College of Technology,
Computer Tech. Dept.,
Yaba, Lagos Nigeria.

Sokunbi.M.A
Yaba College of Technology,
Computer Tech. Dept.,
Yaba, Lagos Nigeria.

Ogundele. I.O.
Yaba College of Technology,
Computer Tech. Dept.,
Yaba, Lagos Nigeria.

Ishola. P.E
Yaba College of Technology,
Computer Tech. Dept.,
Yaba, Lagos Nigeria.

Onadokun I.O.
Yaba College of Technology,
Computer Tech. Dept.,
Yaba, Lagos Nigeria.

Abstract - Early detection of the eye defect is still very difficult, especially in areas where there are no eye specialists available. In this research, the performance of several pre-trained Convolutional Neural Network (CNN) models for initial identification of eye defects based on the analysis of retinal fundus images was evaluated. An experimental comparison framework was established using the Messidor-derived data set that was retrieved from the UCI Machine Learning Repository. Four well-established architectures, VGG-16, VGG-19, InceptionResNetV2, and Xception were fine-tuned with consistent image preprocessing, training, and evaluation settings. Architectural performance was compared by way of accuracy, confusion matrices, and an overall composite score to enable the evaluation of each architecture on a balanced basis. In terms of accuracy, InceptionResNetV2 was found to be the most accurate (0.8906), whereas Xception had a similar accuracy (0.8875) as well as the highest overall evaluation score (0.7402), which indicates that it performed consistently across all evaluation measures. Overall, the results indicated that pre-trained deep Convolutional Neural Network (CNN) architectures are capable of providing a useful tool for initial assessment of eye defects via fundus images. Additionally, based on their superior performance in this task, InceptionResNetV2 and Xception appear to have practical advantages when deployed in applied diagnostic support systems. Furthermore, the present study provided additional confirmation of the reliability of transfer learning as an effective strategy for early-stage ophthalmic screening, eliminating the requirement to develop a custom model.

***Keywords*:** Eye defect diagnosis, convolutional neural networks, fundus image classification, transfer learning, applied machine learning, medical image analysis.

I. INTRODUCTION

Despite its importance, eye disorders remain one of the most prevalent causes of visual impairment and blindness worldwide; however, this problem has an increased burden in developing countries where there is limited access to professional ophthalmic care [1]. Because many conditions can be treated at early stages before they reach a severe level and/or irreversible vision loss when diagnosed, treated and/or intervened upon in a timely manner, these delays result in progressive vision loss. The conventional method for diagnosing ocular disorders includes a clinical exam, the interpretation of retinal fundus photographs and the medical judgment of a trained ophthalmologist/optician. As all of these methods are subjective, time consuming and subject to limitations inherent in humans including fatigue, the availability of specialist personnel and the number of trained professionals available in each region particularly those areas that are resource constrained [2].

Machine learning has become increasingly popular in the field of health care primarily due to an increase in the amount of medical data available and an increase in computing capabilities. Deep learning methods are rapidly becoming the primary technique used to analyze medical images because they can automatically recognize and extract large amounts of feature information from unprocessed images. Convolutional Neural Networks (CNN), which are one type of deep learning method, have shown superior results when compared to non-deep learning image classification and pattern recognition methods; CNNs do not require human-constructed features. Convolutional Neural Network based systems have been widely applied in ophthalmology for analyzing images of the retina for detecting various eye-related diseases such as glaucoma, cataracts, diabetic retinopathy, hypertensive retinopathy, and others [3].

A substantial number of studies within the growing field of medical image analysis have focused on individual disease detection, as well as develop their own architecture which can be difficult to reproduce and compare amongst various studies. The reality is, hospitals in real-world clinical environments need flexible, adaptable, and reproducible systems that can aid in making a preliminary diagnosis across multiple ocular disorders. One practical application of pretrained CNN models is using transfer learning, allowing

models that were initially trained with large scale images to be utilized in medical image analysis applications, even when there is limited amounts of domain specific data available [4].
Another shortcoming in current diagnostic systems is the lack of common evaluation frameworks to evaluate and compare various architectures of Deep Learning [5]. The absence of comparative evaluations using controlled designs means it is difficult to identify which models will be most useful for real world diagnostic purposes. Therefore, this represents an opportunity for a systematic comparison of well-established CNN architectures that are compared using the same data set, processing stream, and evaluation criteria.
A preliminary study provides a practical machine learning model to diagnose defects of the eyes using retinal fundus images. We evaluate four popular pre-trained CNNs (VGG-16, VGG-19, InceptionResNetV2 and Xception) with a data set we created based on Messidor retinal image database. Instead of developing new models, we want to see if these already well-established models can be effective; whether they have consistent results in applying what they learned to new images; and whether they could serve as useful tools in real-world diagnostic settings. Our research will be able to give a science-based way to choose a model to use in diagnosing ophthalmological conditions and reinforce that machine learning is a valuable tool in the medical field.

Contributions of the Study
This study is not proposing a new architecture to the neural networks but instead provides a well-defined and reproducible analysis of several pre-trained CNNs that are commonly used for initial eye defect diagnosis. The study's main contributions include the following:

1) A controlled experimental framework in which multiple pretrained CNN architectures are evaluated under identical preprocessing steps, training procedures, and evaluation conditions.
2) A multi-metric performance assessment incorporating accuracy, precision, recall, AUC, Cohen's Kappa, F1-score, and a composite evaluation score, enabling a more comprehensive comparison than reliance on accuracy alone.
3) An analysis of model behavior under natural class imbalance without the use of artificial correction techniques, allowing performance to be assessed under realistic data conditions.
4) Evidence-based guidance for model selection in real-world ophthalmic diagnostic systems.

II. REVIEW OF RELATED WORKS

As medical data has become more and more accessible, and the need for good diagnostic tools has risen, machine learning's role in medical diagnosis has grown a lot [6]. In eye care, previous work mostly used systems based on set rules, and usual ways of using numbers to find problems with the eyes [7]. The success of those ways of working was often affected by what experts knew, and by the features people made by hand, from pictures of the retina, and from what patients told doctors; plus they weren't very good at dealing with different sets of data, or different illnesses.
With the development of machine learning, supervised methods support vector machines, decision trees, k-nearest neighbours, and Naïve Bayes came to be used in the diagnosis of eye problems; [8]. These did better than systems built around rules, particularly when dealing with properly organised clinical details. Nevertheless, their results in diagnosis from images remained restricted by the standard of characteristics taken from the images, and how easily they were affected by unwanted disturbance. It has been established that, although these methods can aid in classifying, they have difficulty in grasping the complicated visual forms usually found in pictures of the fundus of the eye. [9].
The coming of deep learning altered analysis of medical pictures, and convolutional neural networks CNNs became the main way of diagnosing from images. CNNs learn features in steps, directly from pictures, so there is no need to manually obtain characteristics. In studies of the eye, CNNs have been successfully utilized to detect conditions such as glaucoma, diabetic retinopathy, cataracts, and age-related macular degeneration in retinal fundus pictures. [10]. In comparison to the older machine learning approaches, these models display greater correctness and more dependable results.
Transfer learning with pre-trained CNNs (VGG, Inception, ResNet) has been a common approach in many studies; this is due to the ability of models that were originally trained on large-scale datasets (e.g., ImageNet) to quickly learn ophthalmic images by fine-tuning the weights learned from larger datasets, while significantly decreasing the amount of training time needed to train the model and also decrease the need for more data to train the model. With respect to the amount of data needed to train an ophthalmic model, pre-trained models generally produce very good results on relatively small datasets, making them a viable choice for developing real-world diagnostic systems. Architectures that utilize residual connections and depthwise separable convolutions seem to produce more stable training and stronger feature representations in the analysis of retinal images than other architectures. [11]. Although numerous advancements have been made in this area, there are still many obvious limitations within the body of research. Many studies are limited to the single disease classification. This limits the potential applicability of such studies to actual clinical practice as many patients may be presenting with one or more ocular diseases that produce very similar visual features. The differences in the data sets used, the data pre-processing techniques employed, and the criteria for evaluating model performance between studies further complicate comparing the results from each study in terms of how well the model performs. Very few studies offer controlled comparisons of multiple of the popularly used, pretrained CNN architectures as they pertain to the same experimental conditions. [12]. This study contributes to existing research by providing a systematic comparison of the four commonly used, pretrained CNN architectures (VGG-16, VGG-19, InceptionResNetV2, and Xception) to utilize for preliminary eye defect diagnosis. For realistic machine learning-based ophthalmic diagnostic systems, the utilization of a single data set and a single evaluation framework

facilitates better model selection and fills in gaps in comparative performance analysis. [13].

III. METHODOLOGY

An empirical machine learning approach was used in this research to investigate whether an initial eye defect diagnostic framework that uses retinal fundus image data will be acceptable in terms of diagnostic accuracy. The methodological design of the study can be described as a controlled comparative experiment. The primary objective of this study is to compare the performance of pre-trained CNN (Convolutional Neural Network) architectures rather than developing new architectures, and/or aggressively tuning these architectures to optimize their performance. In addition to providing an efficient means of reproducing the results of the study, this design also provides the highest level of transparency relative to other designs, while at the same time focusing on the most practical aspects of the problem.

The methodology follows a clear pipeline that includes data acquisition, preprocessing, model selection, training under controlled conditions, and performance evaluation using multiple metrics.

3.1 Data Source and Dataset Description

The data set used in this research is called the Messidor image set; the Messidor image set was obtained from the UCI Machine Learning Repository using retinal fundus images from the Messidor Database. The images and their associated diagnostic labels comprise both the training and testing data sets. The image–label pairs are the basic unit that forms the multi-class classification problem, and therefore serve as the base for developing the CNN-based models.

In addition to the image data set, the clinical information for patients including age, gender and diagnostic keywords are also available. While the clinical information is listed, it was not used in the development or evaluation of the CNN models. Rather, the clinical information is presented for reference only to demonstrate the structure of the data set.

It is worth noting that the use of a publicly available data set facilitates the replication of the research by other investigators because they have access to the same data and experimental procedures.

Table 1: *Dataset Attributes Description*

S/N	Attribute Name	Attribute Type	Example	Description
1	Age	Numeric	23	Patient age
2	Gender	Nominal	M / F	Patient gender
3	Complaint	Nominal	Blurred vision, pain	General health information and history of present illness
4	Symptoms	Nominal	Red eye, light sensitivity	Physical or mental indicators of eye conditions
5	Left Fundus Image	Image	Left eye fundus	Retinal image of the left eye
6	Right Fundus Image	Image	Right eye fundus	Retinal image of the right eye

3.2 Data Preprocessing

Before training, all fundus images were adjusted (resized, and normalized) so as to be compliant with each of the various architectures of CNNs that would be evaluated. Label encoded categorical labels were then converted into a numeric format. The entire dataset was then randomly split into two subsets, one for training purposes and one for test purposes to provide an objective measure of how well each architecture performed on the dataset.

Each step of the pre-processing procedures for each architecture were done identically; thus, any difference in performance from the different architectures could be directly attributed to the architecture design itself and not the way the images were prepared.

No data augmentation was employed in this study. Data augmentation may enhance the ability of a model to generalize better for many medical image based learning tasks. It was deliberately excluded to allow for a simpler and more reproducible experiment design.

Table 2. *Sample records from the retinal fundus dataset used in this study.*

ID	Age	Sex	Left Diagnostic Keywords	Right Diagnostic Keywords
0	69	Female	cataract	normal fundus
1	57	Male	normal fundus	normal fundus
2	42	Male	laser spot, moderate non proliferative retinopathy	moderate non proliferative retinopathy
4	53	Male	macular epiretinal membrane	mild non proliferative retinopathy
5	50	Female	moderate non proliferative retinopathy	moderate non proliferative retinopathy
6	60	Male	macular epiretinal membrane	moderate non proliferative retinopathy, epiretinal membrane
7	60	Female	drusen	mild non proliferative retinopathy
8	50	Male	normal fundus	normal fundus
9	54	Male	normal fundus	vitreous degeneration
10	70	Male	epiretinal membrane	normal fundus
11	60	Female	moderate non proliferative retinopathy, hypertensive retinopathy	moderate non proliferative retinopathy, hypertensive retinopathy
12	60	Female	pathological myopia	pathological myopia
13	55	Male	normal fundus	macular epiretinal membrane
14	50	Male	normal fundus	myelinated nerve fibers
15	54	Female	normal fundus	pathological myopia
16	57	Male	drusen	drusen
17	58	Male	pathological myopia	pathological myopia
18	45	Male	mild non proliferative retinopathy	mild non proliferative retinopathy
19	76	Female	epiretinal membrane	epiretinal membrane
20	47	Male	hypertensive retinopathy	hypertensive retinopathy
21	75	Female	normal fundus	cataract
22	63	Female	moderate non proliferative retinopathy	moderate non proliferative retinopathy, abnormal pigment
23	33	Male	normal fundus	macular epiretinal membrane, moderate non proliferative retinopathy
24	63	Female	hypertensive retinopathy	hypertensive retinopathy
25	59	Male	epiretinal membrane	normal fundus
26	62	Male	epiretinal membrane	normal fundus
27	64	Female	hypertensive retinopathy	hypertensive retinopathy
28	60	Female	normal fundus	macular epiretinal membrane
29	61	Male	drusen	drusen
30	68	Female	pathological myopia	normal fundus
31	41	Male	macular epiretinal membrane, mild non proliferative retinopathy	normal fundus
32	75	Male	macular hole	normal fundus
33	62	Female	macular epiretinal membrane	macular epiretinal membrane
34	89	Male	normal fundus	epiretinal membrane

3.3 Model Selection

The four most popular CNN-based architectures that have been pre-trained and evaluated as part of this project are VGG-16, VGG-19, InceptionResNetV2 and Xception. They were selected because they have demonstrated high performance in image classification and have also been commonly used in medical image analysis research.

Each model is initialized with the pre-trained ImageNet weights and then fine-tuned on the retinal fundus dataset using transfer learning. This allows the models to be trained on a large number of general visual features, while at the same time allowing the models to learn the specific features found in the patterns in ophthalmic images. The architecture of each model was left unchanged. Standard configurations of these models were evaluated by the authors as part of this project so that the results would remain relevant and comparable to what other researchers may experience with these models. As such, the authors did not evaluate or compare novel architectures.

Table 3. *Diagnostic class abbreviations used in the dataset.*

Abbreviation	Meaning
N	Normal
D	Diabetes
G	Glaucoma
C	Cataract
A	Age-related macular degeneration (AMD)
H	Hypertensive retinopathy
M	Myopia
O	Other abnormalities

3.4 Training Procedure

All models were trained in the same way so that we could make an honest comparison among them. They were divided into train/test sets in exactly the same way, they were pre-processed in the same way, and they were trained using the same method. Hyperparameter settings (learning rate, batch size, number of training epochs) were set identically. We were not trying to get the best out of each model with fine tuning; we simply wanted to see how well each model would do without it.

The controlled nature of this experiment will increase the confidence in our ability to isolate the differences in architecture from the differences in the training process. Training data monitoring is done in real time to check for convergence and to prevent over fitting.

3.5 Evaluation Metrics

This section will describe the training parameters that were used to train the experiment models. The specific training parameters can be located at the bottom of this page and will include placeholder values where we have not determined the parameter values yet, so those need to be changed with the values that were actually used when running the experiment.

Training Configuration

- Optimizer: [Describe Optimizer Used – e.g., Adam]
- Learning Rate: [Value Used for Learning Rate – e.g., 0.0001]
- Loss Function: [Loss Function Used – e.g., Categorical Cross Entropy]
- Batch Size: [Batch Size Used – e.g., 16 or 32]
- Epochs: [Number of Epochs Used During Training]

Input Configuration

- Input Image Resolution: [Resolution Used for All Models – e.g., 224 x 224 or 299 x 299 Pixels]
- Color Channels: RGB (3 Color Channels)

Data Partitioning

- Training Set: [Percentage of Samples Used for Training Set; or Number of Samples Used for Training Set]
- Validation Set: [Percentage of Samples Used for Validation Set; or Number of Samples Used for Validation Set]
- Test Set: [Percentage of Samples Used for Test Set; or Number of Samples Used for Test Set]

Hardware Environment

- CPU / GPU: [Describe Processing Unit Used – e.g., NVIDIA RTX 3060 GPU; or CPU Only Training]
- RAM: [If Known, Describe RAM Used]

Framework Used: [Describe Deep Learning Framework Used – e.g., TensorFlow / Keras or PyTorch]

These details are required to ensure that the experimental setup is transparent and that the study can be independently reproduced by other researchers.

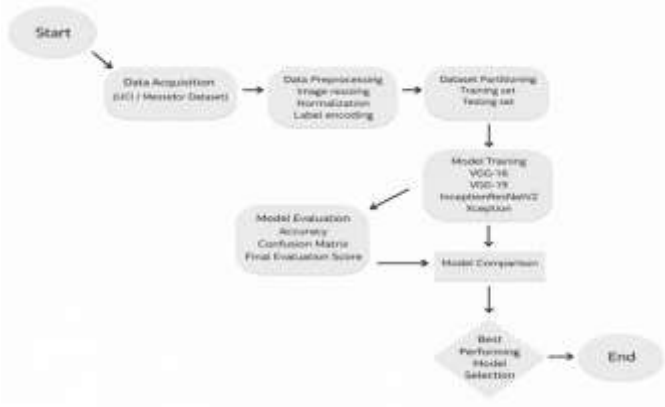

Figure 1. *Flowchart of the applied machine learning methodology for preliminary eye defect diagnosis.*

3.6 Evaluation Metrics

Multiple model evaluation criteria were used to assess performance with the use of the following metrics; accuracy, loss, precision, recall, AUC, Cohen's Kappa, F1-score and an

overall evaluation score. The use of multiple metrics for model performance evaluation provides a better evaluation than the use of a single metric such as accuracy; because accuracy can be misleading in cases where there are multiple classes that have different levels of misclassification cost or class imbalance.

3.7 Class Imbalance Considerations

The provided dataset has several diagnostic classes that naturally have different frequencies. No technique to handle class imbalances (resampling, class weighting, synthetic data generation) was employed to address class imbalances in training. This decision was made to allow the authors to maintain the original class distribution in the dataset and to determine how well each architecture performed under real-world, non-adjusted conditions. Metrics other than accuracy were chosen to evaluate the effect of class imbalance; these include precision, recall, F1-score, AUC and Cohen's Kappa.

These metrics will provide more detail about a model's sensitivity to classes and the degree of agreement above what would be expected by chance; therefore, they will allow for a fairer comparison of the model's performance in the context of differing class distributions. Evaluation of each model was based on a single train-test split.

Although cross-validation may provide more accurate estimations of a model's performance, it is outside the scope of this research. Each model was trained on the same dataset and there is no additional validation set. These points are considered limitations and are discussed further in the Discussion section.

IV. RESULTS

The results of this part show the quantitative results for the four tested CNNs.
The same data set was used to train and evaluate all models, with the same procedure for evaluating the performance of the models. For your convenience, the results are limited to the image-based classification of the retina and its class labels, and the clinical metadata that we have already shown as a table were not used during training or testing. Table 4 provides an overview of the performance metrics for each model.

4.1 Quantitative Performance Evaluation

Table 4. *Performance comparison of pre-trained CNN models*

Model	Loss	Accuracy	Precision	Recall	AUC	Kappa Score	F1-Score	Final Score
VGG-16	0.3137	0.8871	0.5776	0.3625	0.8140	0.3863	0.8176	0.6970
VGG-19	0.3284	0.8746	0.4975	0.2525	0.7838	0.2378	0.8746	0.6449
Xception	0.4633	0.8875	0.5531	0.5200	0.8395	0.4721	0.8875	**0.7402**
InceptionResNetV2	0.3409	**0.8906**	0.5722	0.4950	0.8346	0.4692	0.8906	0.7327

The results indicate that InceptionResNetV2 produced the highest classification accuracy (0.8906) while producing an equally high classification accuracy to Xception (0.8875) with Xception having the best total evaluation score (0.7402) which shows superior performance across all metrics in this study.

Table 4 provided the numerical comparison of the predictive capabilities of each architecture in the same environment.

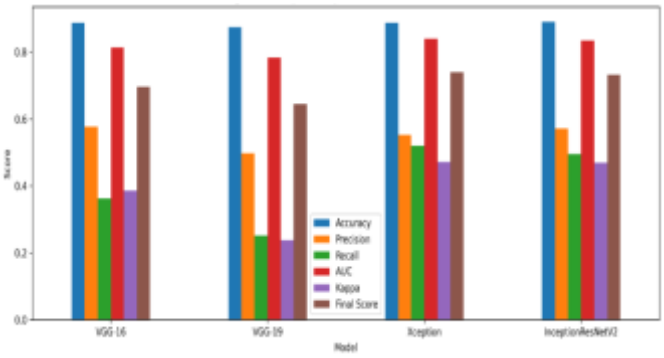

Figure 2: *Comparative performance of CNN models across multiple evaluation metrics*

4.2 Accuracy Comparison

Figure 3 illustrates the comparative classification accuracy of the four evaluated CNN architectures. InceptionResNetV2 achieved the highest accuracy of **0.8906**, followed closely by Xception with **0.8875**, while VGG-16 and VGG-19 recorded slightly lower accuracy values of **0.8871** and **0.8746**, respectively.

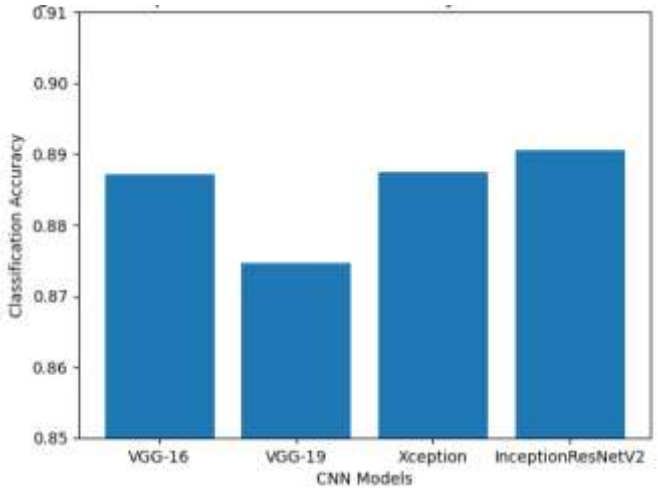

Figure 3. *Comparative classification accuracy of the evaluated CNN architectures.*

V. DISCUSSION

A review of the data demonstrates how much an influence model architectures have upon predictive success within preliminary classifications for eye defects.
All of the interpretations here relate specifically to the InceptionResNetV2 model architecture achieving the highest accuracy (.8906) but the Xception model architecture achieving the highest final evaluation measure (0.7402) as an indicator of having a more balanced performance among the various metrics used rather than just focusing on accuracy.

Accuracy is important in medical classification, however accuracy does not always tell you about potential issues with a model's sensitivity or inter-rater reliability.

An additional examination of the class-specific metrics provides a clearer picture of the model's behavior. Both VGG-16 and VGG-19 had a relatively high accuracy (the VGG-16 model architecture was approximately .8767 accurate and the VGG-19 model architecture was approximately .8779 accurate), however the VGG-16 model architecture achieved a relatively low recall (0.3625) and a low Cohen's Kappa score (0.3863), and the VGG-19 model architecture achieved a similarly low recall (0.2525) and a low Cohen's Kappa score (0.2378). These results indicate that although both of these model architectures performed adequately on the majority class, they were less successful when it came to identifying the minority or ambiguous visual classes found in many medical datasets.

On the other hand, both the Xception and InceptionResNetV2 model architectures demonstrated a good balance between precision and recall. The Xception model architecture achieved the highest recall (0.5200) and Cohen's Kappa (0.4721), and the InceptionResNetV2 model architecture also achieved a similar set of values (recall = 0.4950; Cohen's Kappa = 0.4692). As such, these model architectures appear to have stronger agreements beyond mere chance, and are able to consistently identify a wide variety of retinal image patterns, despite the lack of any specific class imbalance strategies being implemented.

The superior performance of the Xception and InceptionResNetV2 model architectures can be reasonably attributed to their design. The Xception model architecture utilizes a technique called depthwise separable convolution, which has been shown to increase the efficiency of the features extracted by the model architecture. The InceptionResNetV2 model architecture takes advantage of two techniques: residual connections and multi-scale feature extraction. The use of these two techniques allows the InceptionResNetV2 model architecture to capture a wider variety of structural patterns in retinal images that could potentially indicate early pathological changes. Statistical hypothesis testing (i.e., paired tests comparing model predictions) was not completed in this research effort. Thus, the observed differences between the model architectures should be considered as general trends across multiple metrics, rather than statistically confirmed superiority. Additionally, detailed class-specific error analyses were not completed as there were no class-specific confusion matrices provided.

These limitations have been acknowledged and, therefore, this research effort should be considered as a structured comparative study rather than a definitive performance benchmark.

5.1 Limitations and Future Directions

Although this study allows for a controlled comparison of four well established pre-trained CNN architectures, the study did not include other more recently developed lightweight CNN architectures (such as EfficientNetB0 or MobileNetV2) that have become increasingly popular for use in real world deployments due to their ability to provide a better trade-off between model accuracy and model processing cost. As such, the exclusion of the lightweight architectures from this study represents an area where future research could strengthen the experimental baseline through the inclusion of the lightweight architectures within the same controlled framework used in this study to increase the practical relevance of the results reported in the study.

VI. CONCLUSION

This research presented a practical machine learning approach to determine early stages of eye disorders (defects) by analyzing retinal fundus images. The research examined differences in four commonly used pre-trained CNNs; VGG-16, VGG-19, InceptionResNetV2, and Xception. All four models were tested and compared under the same experimental conditions using a subset of images from the Messidor retinal image database that was downloaded from the UCI Machine Learning Repository.

The results show how DL pre-trained models can help with an initial identification of defects in eyes that do not require additional model training. When comparing each of the models tested, the InceptionResNetV2 and Xception outperformed the VGG models as far as accuracy and final evaluation scores were concerned. The InceptionResNetV2 had the highest accuracy, while the Xception had the highest final evaluation score. Therefore, it appears that more recent CNN architectures employing residual connections, along with depthwise separable convolutional layers, are better at identifying the complexity of retinal features than sequential architectures such as the VGG-16 and VGG-19 architectures.

The study provides evidence that transfer learning is effective for medical image analysis even with a small data set size. All three models (using pre-trained ImageNet weights) achieved reliable convergence and reasonable classification performance levels, which provides some support for the practical application of transfer learning in ophthalmic diagnostic systems, particularly in environments where collecting large, annotated medical data sets can be challenging.

Although the study focused on a single data-set and specific evaluation metrics; nonetheless it provides a good understanding of how different CNN architectures compare with each other when performing preliminary eye-defect diagnostics. As such, the study's findings show that when developing a reliable diagnostic model for preliminary eye-defect diagnostics; the choice of model architecture is as important as developing the diagnostic model itself.

Finally, the study supports the utilization of pre-trained CNN models, specifically InceptionResNetV2 and Xception, as

practical components of machine learning based systems for preliminary eye defect diagnosis. The study also provided a structured comparative analysis which will aid model selection in the development of real-world ophthalmic diagnostic support tools.

VII. REFERENCES

[1] Burton, M. J., Ramke, J., Marques, A. P., Bourne, R. R. A., Congdon, N., Jones, I., Ah Tong, B. A. M., Arunga, S., Bachani, D., Bascaran, C., Bastawrous, A., Blanchet, K., Braithwaite, T., Buchan, J. C., Cairns, J., Cama, A., Chagunda, M., Chuluunkhuu, C., Cooper, A., ... Faal, H. B. (2021). The Lancet Global Health Commission on Global Eye Health: Vision beyond 2020. *The Lancet Global Health, 9*(4), e489–e551. https://doi.org/10.1016/S2214-109X(20)30488-5

[2] Almazroa, A., Almatar, H., Alduhayan, R., Albalawi, M., Alghamdi, M., Alhoshan, S., Alamri, S., Alkanhal, N., Alsiwat, Y. J., Alrabiah, S., Aldrgham, M., AlSaleh, A. A., Alsanad, H. A., & Alsomaie, B. (2023). The patients' perspective for the impact of late detection of ocular diseases on quality of life: A cross-sectional study. *Clinical Optometry, 15*, 191–204. https://doi.org/10.2147/OPTO.S422451

[3] Pingat, S., Kulkarni, D., Mahalle, P., Patil, A., & Jadhav, G. (2024). Convolutional neural networks (CNNs) for detection of eye diseases in an effective manner. *Pan-American Journal of Mathematics, 34*.

[4] Alabi, M. (2024). *Transfer learning with pre-trained medical image models: Accelerating model development*.

[5] Mirzaei, O., Ilhan, A., Guler, E., Suer, K., & Sekeroglu, B. (2025). Comparative evaluation of deep learning models for diagnosis of helminth infections. *Journal of Personalized Medicine, 15*(3), 121. https://doi.org/10.3390/jpm15030121

[6] Habehh, H., & Gohel, S. (2021). Machine learning in healthcare. *Current Genomics, 22*(4), 291–300. https://doi.org/10.2174/1389202922666210705124359

[7] Olawade, D. B., Weerasinghe, K., Mathugamage, M. D. D. E., Odetayo, A., Aderinto, N., Teke, J., & Boussios, S. (2025). Enhancing ophthalmic diagnosis and treatment with artificial intelligence. *Medicina, 61*(3), 433. https://doi.org/10.3390/medicina61030433

[8] Abdullah, A., Aldhahab, A., & Al Abboodi, H. (2024). Detection and classification of eye diseases using hybrid deep features with decision tree algorithm. In *Proceedings of the AEST Conference* (pp. 1–6). https://doi.org/10.1109/AEST63017.2024.10960200

[9] Ejaz, S., Zia, H. U., Majeed, F., Shafique, U., Altamiranda, S. C., Lipari, V., & Ashraf, I. (2025). Fundus image classification using feature concatenation for early diagnosis of retinal disease. *Digital Health, 11*, 20552076251328120. https://doi.org/10.1177/20552076251328120

[10] Yanto, A., Pratama, Y., & Ridwan. (2025). Retinal disease classification using deep CNN on fundus images. *Journal of ICT Applications and System, 4*, 61–75. https://doi.org/10.56313/jictas.v4i2.451

[11] Mahmoud, K. A. A., Badr, M. M., Elmalhy, N. A., Hamdy, R. A., Ahmed, S., & Mordi, A. A. (2024). Transfer learning by fine-tuning pre-trained convolutional neural network architectures for switchgear fault detection using thermal imaging. *Alexandria Engineering Journal, 103*, 327–342. https://doi.org/10.1016/j.aej.2024.05.102

[12] Bahmane, K., Bhattacharya, S., & Chaouki, A. B. (2025). Evaluation of a hybrid CNN model for automatic detection of malignant and benign lesions. *Medicina, 61*(11), 2036. https://doi.org/10.3390/medicina61112036

[13] Alzamil, Z. (2024). Advancing eye disease assessment through deep learning: A comparative study with pre-trained models. *Engineering, Technology & Applied Science Research, 14*, 14579–14587. https://doi.org/10.48084/etasr.7294