International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

A Newly Developed Deep Learning-Based Xception Model for Classification and Detection of Eye Disease Using Fundus Images

Himanshu Sharma*[1], Javed Wasim[2], Pankaj Sharma[3]

Submitted:15/03/2024 **Revised**: 25/04/2024 **Accepted**: 07/05/2024

Abstract: A condition affecting the retina, Diabetic Retinopathy (DR) develops in patients with long-term diabetes mellitus (>20 years). In many parts of the globe, DR is a leading cause of avoidable blindness. People with diabetes are at increased risk for developing diabetic retinopathy (DR), a serious eye disease. Treatment of eye diseases and prevention of irreversible vision loss depends on early diagnosis and prompt treatment. Fundus photos aid medical professionals in identifying disorders affecting patients' eyes. A fundus picture could show several eye diseases. One frequent method for identifying eye diseases is screening retinal fundus images, although manual identification is labor-intensive and takes a lot of time. Many scholars have thus resorted to deep learning (DL) methods in an effort to automate the diagnosis of retinal eye disorders. To classify and detect eye diseases using fundus images from the DDR Dataset, images are preprocessed, resized, and filtered before being labelled into three classes. The dataset undergoes 5-fold cross-validation. A fine-tuned Xception model with additional fully connected layers is trained using the Adam optimiser. The experimental findings show that the suggested system achieves 92.78% and 98.98% train and validation accuracy with Xception architecture and surpasses existing approaches for categorising eye diseases. The model's robustness is shown by its accuracy, recall, and F1-score metrics, which demonstrate its ability to produce correct predictions for each class. The model's training and validation accuracy, as well as loss curves, show that it avoids overfitting and has high generalisation capabilities. The suggested model has great potential as a helpful resource for doctors to use in the early and accurate identification of eye disorders, which would lead to better care for patients and better ocular health management overall.

Keywords: *Diabetic Retinopathy, Fundus images, DDR Dataset, Deep Learning, Xception, CNN, Cross-Validation.*

1. Introduction

When a person has diabetes, their blood glucose levels are consistently higher than usual [1]. Hyperglycemia and abnormalities in the metabolism of carbohydrates, lipids, and proteins, along with either an increase or decrease in insulin secretion or activity, define diabetes. Diabetes affected 9.3% of the global population in 2019 (463 million people), and experts predict that number will rise to 10.9% by 2045[2]. Damage to the retina may occur in people with diabetes as a result of excessive blood glucose levels. This results in the inability to see. Several eye problems, including DR, diabetic macular edoema (DME), cataracts, and glaucoma, may affect people with diabetes. DR is a condition affecting the retina's blood vessels; symptoms include microaneurysms, haemorrhages, and both soft and hard exudates [3].

The most frequent kind of eye illness that may cause blindness is diabetic eye disease [4]. Damage to the retina's blood vessels, which may lead to blindness, results from this. Among the over 4.1 million individuals afflicted with this particular kind of eye disease, a quarter have some degree of visual impairment [5] [6]. Bleeding from the eyes, distorted or double vision, or even total blindness may result from retinopathy. A disorder when the optic nerve is damaged due to a total blockage of blood flow. Despite this, keep in mind that anybody might have blindness at any point throughout their lives. The likelihood of developing a visual impairment or becoming blind rises as one ages. Lesions or other anomalies that can suggest a disease might be seen during an imaging study of the retinal fundus[7][8]. The utilisation of a fundus image allows for the detection of certain retinal disorders. One of the numerous medical disorders that may cause cottonwool spots is diabetes mellitus. Other possible causes include systemic hypertension, leukaemia, AIDS, and many more[9]. Figure 1 presents the fundus images.

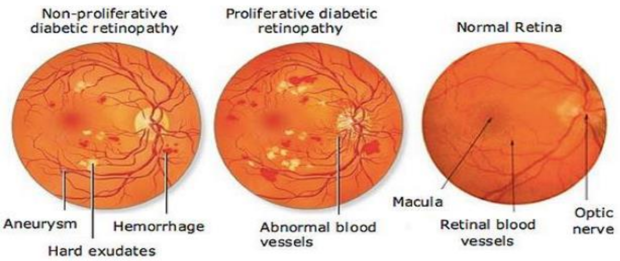

Fig. 1. Fundus images[10].

Medical screening systems based on artificial intelligence (AI) are on the rise, providing practical and affordable options for the automated detection of retinal illnesses, made possible by the recent exponential expansion of data-driven technologies and digital processors[11]. This century

[1] *Mangalayatan University, INDIA*
ORCID ID : 0000-0000-0000-000X
[2] *Mangalayatan University, INDIA*
ORCID ID : 0000-0000-0000-000X
[3] *Eshan College of Engineering, INDIA*
ORCID ID : 0000-0000-0000-000X
** Corresponding Author Email: Sharma_himanshu86@rediffmail.com*

has been greatly influenced by machine learning, AI, and deep learning algorithms. Through accurate diagnosis, these technological methods have helped people with a range of eye problems. It has been observed, nevertheless, that both DL and conventional ML provide the impetus for machine learning. Fundus image analysis is one area where computer vision and DL methods have shown very impressive development and potential [12][13].

The two basic types of DL tasks used for diagnosing retinal diseases are classification and segmentation tasks [14]. The input photos are categorised into different diseases directly as part of the classification process. The use of deep learning-based approaches to medical image processing has recently shown enormous potential. This research presents a new method for diagnosing ocular diseases utilising CNN and other deep learning-based approaches. A CNN which can learn the characteristics from retinal pictures and categorise the images into different eye disorders makes up the suggested model. To optimise training time or transfer knowledge from one job to another using a small number of pictures, deep learning makes use of pre-trained CNN networks [15]. Fine-tuning the pre-trained network is the go-to approach for transfer learning. Whether it's a pre-trained model or a brand new one, before training CNN architectures, picture data sets are usually preprocessed in several methods, including resizing, amount, standardisation, and augmentation. Compare the recommended technique to other models: The study shows that Xception categorises fundus images better than DenseNet121 and Inception-ResNetV2.

As a result of developments in DL algorithms, automated early diagnosis of diabetic eye disease provides substantial benefits over human detection. There have been a number of recent publications of high-quality research on the topic of diabetic eye disease detection. Automatic methods for detecting diabetic eye illness are thoroughly examined in this work from several angles, such as (i) existing datasets, (ii) picture preprocessing methodologies, (iii) DL models, and (iv) metrics for evaluating performance. Research communities, healthcare providers, and diabetic patients may benefit from the survey's thorough summary of methods for detecting diabetic eye illness, which includes methods from the cutting edge of the field.

1.1. Contribution of the Research

The objective of this study was to provide a solid system for DR categorisation in fundus images. "Preprocessing, extraction, and classification" describes the method that was shown. Here is a rundown of the practical benefits of this study:

- To leverage the Xception architecture for high accuracy and reliability: The goal of proposed study is to use the Xception model originally developed as a deep and quite effective in feature extraction that will provide for high performance in the classification of eye diseases from fundus images.
- To enhance preprocessing techniques for better model performance: In image preprocessing, the analysis includes image resizing, Gaussian blur filter, and cropping of the fundus images before inputting them into the model to maximise the quality of images for better classification results.
- To employ 5-fold cross-validation for robust evaluation: By applying 5-fold cross-validation guarantees the model's performance is evaluated and tested hence coming up with more accurate and general results as per the diagnosis of eye diseases.
- To fine-tune the Xception model with additional fully connected layers: The study involves defining extra fully connected layers to the Xception model, adjusting hyperparameters and employing an Improved Adam optimiser to improve the chances of the model classifying fundus image correctly.
- Perform well in classification tasks: The model's effectiveness may be assessed by its accuracy, precision, recall, and F1-score, which measure DR phase recognition.
- Compare the recommended technique to other models: The study shows that Xception categorises fundus images better than DenseNet121 and Inception-ResNetV2.

1.2. Organization of the paper

The remaining portion of the paper follows this pattern. Section 2 covers ML, DL, and CNN-based DR research. Section 3 presents model, flowchart, and hyperparameter-trained eye fundus image categorisation methodology. In Section 4, results and discussions with suggested and current models and performance measures are given. Section 5 concludes with limits and research ideas.

2. Related Work

Most of the research on DR detection has focused on both older and more recent methods of image processing, deep learning, and ML. Fundus images were traditionally pre-processed and feature-extracted from big image collections using a variety of image processing techniques. The identification of eye diseases using various methods has been the subject of several groundbreaking investigations.

Nakhim Chea et al. (2021) improved our knowledge of fundus image multi-category classification using data augmentation, efficient image preparation methods, and optimal residual deep neural networks. Using these to classify three eye diseases using publicly available information produced average accuracies of 85.79% and peak accuracies of 91.16%. Results showed specificities of 90.06%, 99.63%, 99.82%, and 91.90% when comparing images of healthy eyes, GLC eyes, AMD eyes, and DR eyes, respectively. It is possible that improved specificity performances can aid in the early detection of ocular diseases, signalling to patients the possibility of vision loss prevention strategies[16].

Researchers Samira Ortiz et al. (2023) suggest a multi-class classification approach that uses deep learning to accurately identify various eye diseases in fundus pictures. Using a dataset of 1096 fundus pictures for training and testing, the model employs a pre-trained ResNet-50 architecture. Our findings show that the model is effective, with a 97% classification accuracy. If the suggested model can help doctors make accurate diagnoses of eye diseases earlier on, it would greatly benefit patient care and the management of eye health[17].

According to Abini M.A. et al. (2023), there has been a surge of curiosity in CNNs, a type of deep learning system, and its possible use in medical picture analysis. A primary goal of this research is to use the ResNet-50-CNN architecture to categorise fundus pictures into various severity levels. The APTOS 2019 dataset from Kaggle may be used to test the algorithm's efficacy. The two main types of DR classifications that will be covered in the study are binary and multiclass. Two-class model accuracy on the APTOS dataset is 96% and 95%, respectively, throughout training and validation. The enlarged APTOS dataset has training and validation accuracy for multiple class classification of 77% and SO%, respectively[18].

A number of powerful models have made important contributions to medical image analysis; they include Rupali Chavan et al. (2023), VGG 16, and Inception V3. They used the retinal fundus image collection to test these two theories. Area under the curve was calculated by running the trained model on the test dataset. The Inception V3 model has an AUROC of 0.942, and VGG 16 has an AUROC of 0.913. Compared to VGG 16, Inception V3 performs better on the RFMiD dataset. Retinal colour fundus pictures make up RFMiD[19].

For reliable DR and GL classification, G R S Naga Kumar et al. (2023) suggest a multi-scale assisted DL model. To reduce computational burden, the DL model-based classifier is enhanced with a novel ASPP layer that transfers features. Incorporating Soft F1-Loss and CCE Loss into the revised Loss function enhances accuracy and recall while simultaneously increasing the F1-score. To test the suggested approach with several DL models, we utilise a dataset including 10,500 images from various public and private sources. Comparing both models, there is a better result of the multi-scale RESNET50 with F1-score = 0. 89 and accuracy = 87. 5%. This method provides a viable option for reliable DR and GL detection, which will allow for the early identification and treatment of these potentially blinding diseases[20].

According to Hoe Yean Sam et al. (2022), the dataset DenseNet was trained using a CNN deep learning model architecture. The training photos have gone through various photo enhancement techniques in order to enhance the result of the recognition. The outcome of this was to create an algorithm that predicts and classifies the severity label on the fundus images. On the Messidor-2 dataset, the prediction model produced a quadratic weighted kappa score of 0.9308 and 65% accuracy, which was considered reasonably accurate[21].

In this study, Kashish Chauhan et al. (2022) provide an ensemble approach that utilises VGG-19, ResNet101V2, and InceptionV3 to diagnose cataracts in images of the eye's fundus. Soft vote determined the final classification. Ensemble model F-1 Score was 95.90 for test dataset. Ensemble networks perform better than individual networks, according to our research.[22].

Originally, Maneesha Vadduri et al. (2023) used a raw retinal fundus dataset in their study. They tried four pretrained models, which are Xception, ResNet50, VGG-16, EfficientNetB7 and concluded that EfficientNetB7 is the best model. Later on, these raw images were processed by various image enhancement mechanisms, including green colour channel extraction, CLAHE and lighting enhancement. These pictures, after preprocessing, defeated the suggested customised DCNN model along with the four pre-trained models mentioned above. With these images, the model was trained. The DCNN outlined in this paper holds some potential for the NORMAL detection test and the detection tasks of glaucoma (GL), diabetic retinopathy (DR), and cataract (CA), which yielded 96. 43%, 97%, and 96% accuracy, respectively[14].

However, it has been observed that there are some research gaps in the detection of DR through traditional and modern ML, DL, and image processing methods. Firstly, the transferability of models to different populations or different imaging environments remains relatively poor, implying a weakness in highly generalisable data. Secondly, many studies focus on achieving high accuracy without adequately addressing model interpretability, which is crucial for clinical acceptance. Additionally, the integration of multi-modal data, such as combining fundus images with other clinical data, is underexplored, potentially limiting the comprehensiveness of the diagnostic models. Real-world applicability and scalability of these models in clinical settings also remain challenging, often due to computational constraints and the necessity for real-time processing. Furthermore, while high accuracy in multi-class classification is reported, the ability to reliably detect early-stage disease and differentiate between closely related conditions requires further improvement. Possible future directions for this study include adding additional NPDR classification images (in mild, moderate, and severe classes) to our dataset and tackling more difficult classification challenges.

Table 1 provides an overview of different studies, their methodologies, datasets used, results obtained, limitations noted, and suggested future work, offering a comprehensive look at the state of research in fundus photograph classification.

Table 1. Summary of the related work for eye disease detection using different tools and techniques

Reference	***Methods***	***Dataset***	***Results***	***Limitations***	***Future Work***
Nakhim Chea et al. (2021) [16]	Optimal residual deep neural networks, image preprocessing, data augmentation	Public datasets	Peak accuracy: 91.16%, Average accuracy: 85.79%, Specificities: Healthy: 90.06%, GLC: 99.63%, AMD: 99.82%, DR: 91.90%	Focused on specificities, less on other metrics	Explore additional diseases, improve generalisation
Samira Ortiz et al. (2023) [17]	Pre-trained ResNet-50	1096 fundus images	Accuracy: 97%	Limited dataset size	Validate on larger and more diverse datasets
Abini M.A et al. (2023) [18]	ResNet-50-CNN	APTOS 2019	Binary classification accuracy: Train: 96%, Val: 95%; Multiclass accuracy: Train: 50%, Val: 77%	Lower multiclass performance	Enhance multiclass accuracy, integrate more preprocessing
Rupali Chavan et al. (2023) [19]	VGG 16, Inception V3 (transfer learning)	RFMiD dataset	VGG 16 AUROC: 0.913, Inception V3 AUROC: 0.942	Comparison limited to two models	Explore other models and hybrid approaches
G R S Naga Kumar et al. (2023) [20]	Multi-scale RESNET50, Atrous Spatial Pyramid Pooling (ASPP), Soft F1-Loss, Categorical Cross Entropy (CCE) Loss	10,500 images from various databases	Accuracy: 87.5%, F1-score: 0.89	Computational complexity	Optimise model for efficiency, extend to other diseases
Hoe Yean Sam et al. (2022) [21]	DenseNet, image preprocessing	Messidor-2 dataset	Quadratic weighted kappa: 0.9308, Accuracy: 65%	Moderate accuracy	Improve preprocessing techniques, enhance accuracy
Kashish Chauhan et al. (2022) [22]	Ensemble technique (VGG-19, ResNet101V2, InceptionV3), Soft voting	Eye fundus images	F1-Score: 95.90	Dependence on multiple models	Test on larger datasets, refine ensemble approach
Maneesha Vadduri et al. (2023) [14]	EfficientNetB7, image enhancement (CLAHE, illumination correction), image segmentation	Raw retinal fundus dataset	Accuracies: CA: 96.43%, DR: 98.33%, GL: 97%, NORMAL: 96%	Complex preprocessing pipeline	Simplify preprocessing, test on additional datasets

3. Methodology

This project aims to construct a deep-learning model that can effectively classify and identify eye illnesses using fundus images. The model will use the Xception architecture to achieve high reliability and accuracy in medical image analysis. To classify and detect eye diseases using fundus images from the DDR Dataset, images are preprocessed, resized, and filtered before being labelled into three classes. Five-fold cross-validation separates data. Adam optimisers train Xception models with more completely connected layers. Confusion matrix, recall, precision, accuracy, and F1-score evaluate the model. Automatic DR categorisation utilising fundus picture data is successfully accomplished by our suggested deep-learning method, as shown in this experiment.

3.1. Dataset Pre-Processing and Visualisation (DDR Dataset[1])

In this work, used DDR dataset that collected from the Kaggle. Gathered from a variety of hospitals in China between 2016 and 2018, this new dataset contains 13,673 fundus images from 9598 individuals. Figure 2 displays the example images from the collection. The five categories consist of normal (6266 data points), mild (630 data points), moderate (4477 data points), severe (236 data points), and PDR (913 data points). Figure 4 displays the data distribution graph. On this dataset, various data preprocessing techniques for enhance performance of proposed classification model with better dataset.

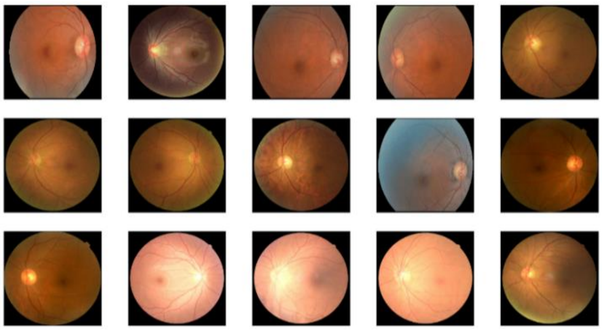

Fig. 2. Sample images of DDR

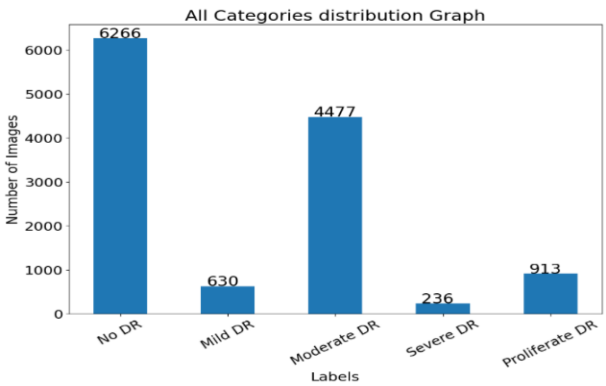

Fig. 3. Data distribution graph of fundas image classes

The above Figure 3 shows the DDR dataset distribution graph. The figure's x-axis displays the labels, while the y-axis counts the number of range images that comprise each of the five classes— Severe DR, No DR, Mild DR, Moderate DR, and Proliferate DR—with the corresponding numbers of images being 6266, 630, 4477, 236, and 913.

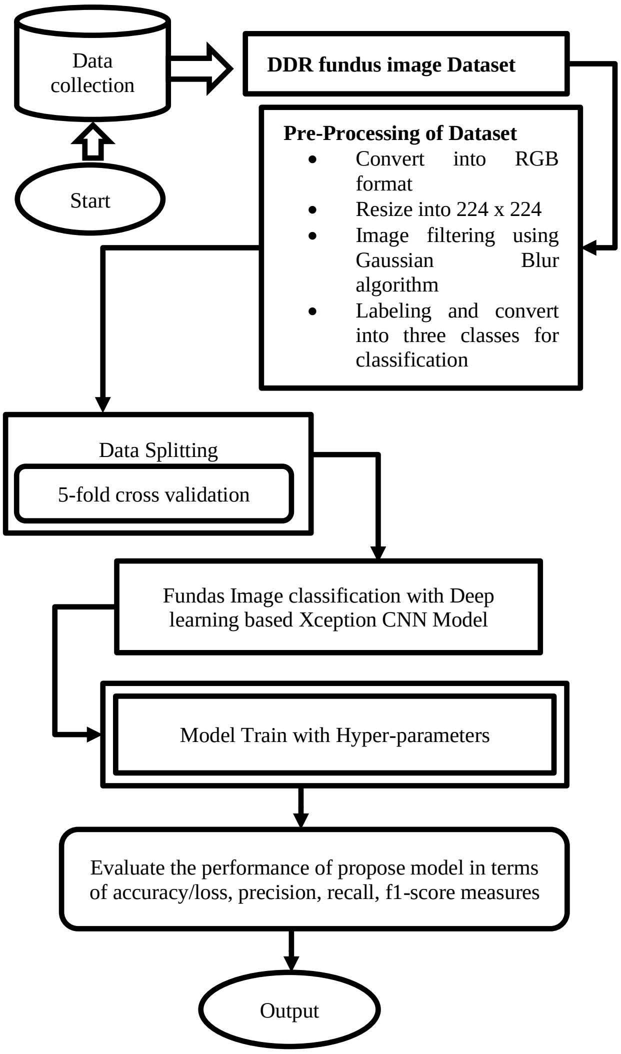

Fig. 4. Flow chart of proposed methodology for fundas image classification

The whole process of methodology process shown in Figure 4, also their deep discussion of each phase is provided below:

The colour fundus image in the DDR fundus image collection has the RGB channel, which stands for red, green, and blue. In order for the created model to fit, the images were then reduced to 224 × 224 pixels. After this, applied Image filtering using Gaussian Blur algorithm. Last, Labeling and converting images into three classes for classification. After that, the black border was clipped from each retinal fundus image so that it would not be too noticeable. The OpenCV package, a popular choice for

[1] https://www.kaggle.com/datasets/mariaherrerot/ddrdataset

image processing in Python, is used to do the cropping and resizing stages. Image scaling and data filtering are the two primary stages of the preprocessing step. The sections that follow elaborate on these stages.

- **Image Resizing:** A common pixel size is used to resize all of the images. To conserve space and memory and shorten the training period, it is necessary to resize the input images, even if utilising the real size might aid in learning.
- **Image filtering using Gaussian Blur algorithm:** The classic approach for blurring images is Gaussian Blur, which is sometimes called Gaussian Smoothing. Executing a weighted average operation on the whole image is the simplest way to describe the Gaussian blur method. Each pixel's value is obtained by taking a weighted average of its own value and all the other values in the field. A low pass filter known as a Gaussian Filter is used to blur parts of an image and lower the volume of background noise.

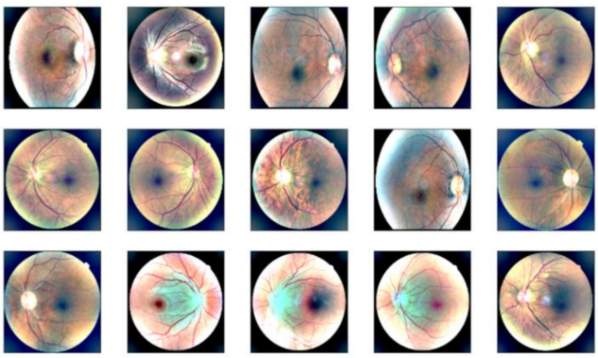

Fig. 5. After dataset preprocessed images of DDR dataset

Figure 5 shows an example of a preprocessed image. DR lesions are absent from the normal retina, as may be seen.

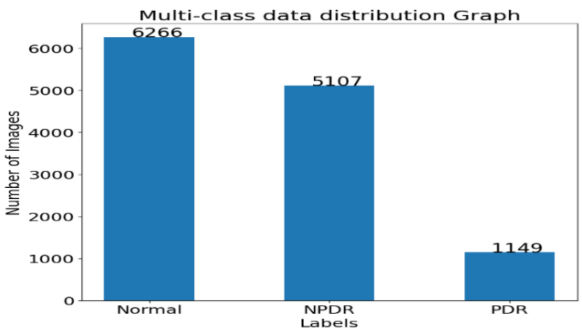

Fig. 6. Multiclass data distribution graph

The above Figure 6 shows the Multiclas data distribution graph of DDR dataset. In graph x-axis displays a labels and y-axis displayed a number of range images. The normal class contain 6266 images, NPDR hold 5107 images and PDR contain 1149 images.

3.2. Data Spiiting

Divide a dataset into a testing and training set after preprocessing. To assess model performance and generalisability, machine learning often uses dataset splitting and 5-fold cross-validation. The dataset is divided into five folds for training and validation. The K-fold cross-validation approach creates five distinct models.

each of which employs a speaker from both the training and test datasets.

3.3. Classification DL-based pre-trained model (Xception)

A variety of retinal diseases may be identified via retinal fundus images. A lot of people are interested in using CNNs to predict eye problems from fundus images because of the recent developments in DL and ML in general. Here, we demonstrate how diabetic retinopathy (DR) may be identified using deep learning models that were trained on external images of the eyes instead. Classification using a traditional neural network model is carried out using computed features. The categorisation process is carried out to estimate various eye disorders, including diabetic retinopathy (DR). The suggested Xception paradigm placed a heavy emphasis on the detection and treatment of ocular disorders. F. Chollet first suggested Xception in 2017. Xception, short for "Extreme Inception," is an Inception-V3 variant with several major modifications. The Xception architecture is made up of fourteen modules that are organised by 36 convolutional layers. Xception is a deep CNN of 134 layers. A 22.9 million parameters are processed by Xception. Two standard convolutional layers (conv) with 32 and 64 filters, respectively, form the basis of the model. Both layers use a 3×3 pixel filter. What follows are five blocks. With the exception of the fourth block, all inputs to such a block go via two separable convolution layers, a maxpooling (MP) layer, and a pointwise convolution via a shortcut link. The fourth block is eight iterations long and has a single separable convolution layer. Two separable convolution layers and a GAP layer follow a final block. Then, an output layer is inserted last, followed by the fully connected (FC) layers[23]. Figure 7 depicts the Conceptual Model of Xception.

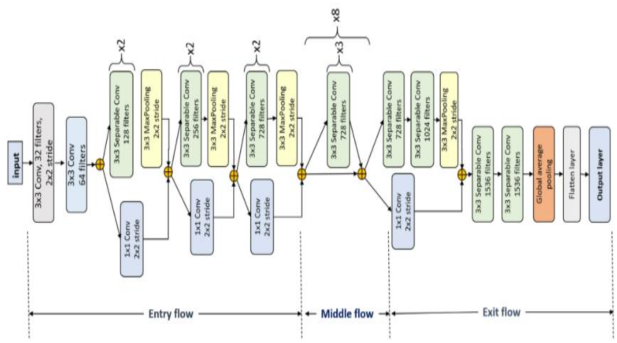

Fig. 7. Architecture of the Xception deep CNN model[24].

Xception, an acronym for "Extreme Inception," is François Chollet's 2017 proposal for a cutting-edge architecture for deep learning [10]. The design of CNNs, especially those for image classification applications, has come a long way with this innovation.[11]. Xception introduces a new way of doing convolution operations, which is at the heart of a radical break from traditional CNN designs [25].

Feature extraction from input images is handled by traditional CNNs using typical convolutional layers. By applying a collection of trainable filters to the whole input volume, these layers generate feature maps that depict spatial patterns. The downside of this method is that it often produces models with too many parameters, which increases the likelihood of overfitting and causes computational inefficiencies.

whereas Xception uses depth-wise separable convolutions like Inception. [26]. Thanks to depth-wise separable convolutions, the traditional convolution process may be divided into two parts: the depth-wise and the point-wise.

S is the stride length, K is the kernel (or filter), and F is the input feature map. One may describe the depth-wise convolution process using Equation (1):

$$F'_{ij} = \sum_{m,n} F_{(i.S+m)(j.S+n)\times K_{mn}} \qquad (1)$$

in where Kmn stands for the matching kernel element and F ′ ij is the output feature map. Through the use of depth-wise convolutions, which are executed independently across each channel of the input feature map, computational complexity is significantly reduced without sacrificing spatial information.

To fuse data coming from various channels, point-wise convolutions are initiated following a depth-wise convolution. Equation (2) expresses this operation:

$$F'_{ij} = \sum_{k} F_{ij\ \times K_k} \qquad (2)$$

in where Kk stands for the point-wise kernel's k-th element. Xception achieves a remarkable equilibrium between expressive power and computing economy via the use of depth-wise and point-wise convolutions.

Xception comes with many techniques that help in stabilising and accelerating the convergence of its models, which are architectural enhancement, batch normalisation, and the ReLU activation [27]. Particularly in medical image analysis applications specifically, Xception has been well received with high accuracy scoring in image categorisation.

3.4. Training with Hyperprameters to fine tunned model

Training the fine-tuned Xception model requires the selection of hyperparameters that will result in the best positive outcome. Using this model, the Adam optimiser is applied together with 5 epochs 32 batches, 0.0001 learning rate. These decisions are an attempt to strike a balance between the time complexity and optimisation that gradient descent undergoes during the training process. The categorical cross-entropy loss function, which computes the disparity between the actual and projected class distributions, aids in training the model with accurate classification findings. An all-encompassing strategy, hyperparameter tweaking increases model generalisation and picture categorisation, which is crucial for diagnosing and monitoring eye illnesses using fundus images. A loss function used in problems involving multi-class categorisation is known as ***categorical cross-entropy***. It computes to equ. 3.

$$Loss = -\sum_{i=1}^{\substack{output\\ size}} y_i\ .log\ \hat{y}_i \qquad (3)$$

here $\hat{y}_i$ represents the i-th scalar value in the model's output, where $\hat{y}_i$ stands for the number of scalar values in the model's output and is the same as the objective value.

Through five iterations of the Adam Optimizer, our model is trained with a learning rate of 0.0001. The Adam optimiser operates according to equations 4 and 5 shown below:

$$\widehat{m}_{t=} \frac{m_t}{1-\beta_1^t}, \qquad (4)$$

$$v_{t=} \frac{v_t}{1-\beta_2^t}, \qquad (5)$$

Consider the following methods: mt estimates gradient mean, and vt¬ calculates second moment (uncentered variance). Because mt and vt are vectors of 0s, Adam's investigators see that they lean towards 0 at early time steps and particularly when decay rates are minimal, meaning that β1 and β2 are close to 1.

Model: "sequential"

Layer (type)	Output Shape	Param #
xception (Functional)	(None, 4, 4, 2048)	20861480
flatten (Flatten)	(None, 32768)	0
dense (Dense)	(None, 512)	16777728
dropout (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 512)	262656
dropout_1 (Dropout)	(None, 512)	0
dense_2 (Dense)	(None, 3)	1539

Total params: 37,903,403
Trainable params: 37,848,875
Non-trainable params: 54,528

Fig. 8. Trained Xception model summary for image classification

The trained Xception model, designed for image classification, incorporates a sophisticated architecture with a total of 37,903,403 parameters, of which 37,848,875 are trainable, as detailed in Figure 8. The model begins with the Xception base, a functional layer extracting high-level features, producing an output shape of (4, 4, 2048) with 20,861,480 parameters. This is followed by a flattened layer, transforming the multi-dimensional output into a 1-

dimensional vector of 32,768 elements. Subsequent dense layers with 512 neurons each, and corresponding dropout layers, are included to enhance learning capacity and prevent overfitting, contributing 16,777,728 and 262,656 parameters, respectively. The final dense layer, with 3 neurons, completes the classification task by mapping to the three output classes, adding 1,539 parameters. This meticulously structured model leverages deep learning to effectively classify images, emphasising both precision in feature extraction and robustness through dropout regularisation.

4. Results and Discussion

Here, we detail the process of analysing the simulation findings and key performance metrics. The free cloud service "Google Colab" was used to create the suggested framework using python 3.7. We made use of the well-known machine-learning packages Matprolib, Numpy, Pandas, TensorFlow, Keras, and Sealborn. For the preprocessing processes, we used the open-source Python library OpenCV. For the classification, we used the DL and other Python packages. Tests were done on a local machine with a core i5/2.4 GHz, 16 GB RAM, and 1 GB VRAM. Splitting DDR datasets uses 5-fold cross-validation. Finish with the performance measures used to assess the recommended Xception model. Suggested model outcomes and assessment matrix:

4.1. Evaluation Performance indicators

The performance of a model may be assessed using a variety of indicators. If you want to choose an assessment metric that will help you evaluate the model, you need to know how each metric computes. The fundamental purpose of this research was to examine all available performance measures after assessing the efficiency of various ML approaches. Confusion matrices and truth tables are the foundations upon which classification performance metrics are built. When evaluating ML algorithms, one particular table is the confusion matrix. We may see a sample generic confusion matrix in Table 2. On one side of the matrix, you'll see examples in real classes, and on the other, you'll see instances in forecasted classes. Several ideas are introduced, including terms true negative (TN), true positive (TP), false positive (FP), and false negative (FN). [28].

Table 2. Confusion Matrix

Classes	*Negative (Actual)*	*Positive* **(Actual)**
Negative (Predict)	TN	FN
Positive (Predict)	FP	TP

- **Accuracy:** Accuracy (ACC) is the ratio of true outcomes (TP and TN) to cases investigated. The formula for its calculation is eq.6:

$$Accuracy = \frac{TP+TN}{N} \qquad (6)$$

- **Precision:** Precision is defined as the ratio of correctly predicted positive outcomes to all positive predictions. The equation may be expressed as euq.7:

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

- **Recall:** Counting the number of correct forecasts divided by the overall number of correct predictions is how recall is determined. Recall ranges between 0 and 1. The best recall is 1. Recall is often computed using the formula euq.8:

$$Recall = \frac{TP}{TP+FP} \qquad (8)$$

- **F1 score:** The F-measure is a weighted harmonic mean estimate of recall and accuracy which is employed to compare various machine learning approaches. Eq.9 provides a F-measure formula.

$$F_1score = \frac{2\times(Precision-Recall)}{(Precision+Recall)} \qquad (9)$$

- **Loss:** In deep learning, loss is more of a function than an accuracy metric; it represents the total of the machine's mistakes made throughout training and validation and is not given as a percentage. Changing the number of epochs or the batch size of training images are two examples of hyperparameters that may be tweaked in a deep-learning model in an effort to lower the loss.

4.2. Results of Xception

The implementation work's experimental findings are presented in the section entitled The Positive Experimental Findings from the Proposed Xception Model for Fundas Image Classification.

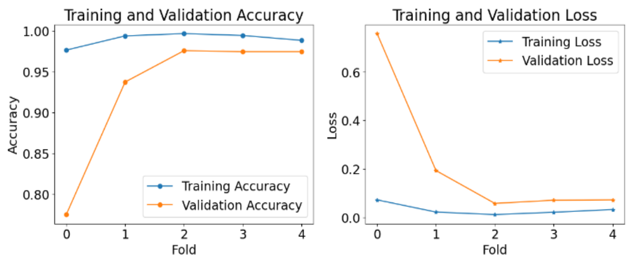

Fig. 9. Ploting Train and Val Accuracy/Loss plot of Xception model for fundas image classification

The above Figure 9 shows the training and validation accuracy/loss with 5 folds for fundas image classification Number of folds is on the x-axis, accuracy, loss, and percentage on the y-axis. Training curves are blue and validation curves are red in figures showing Xception architecture's training and validation accuracy and loss.

With a validation accuracy of 92.74%, the training accuracy is 98.98%. With respect to this, training loss is 3.20% and validation loss is 7.23%, respectively.

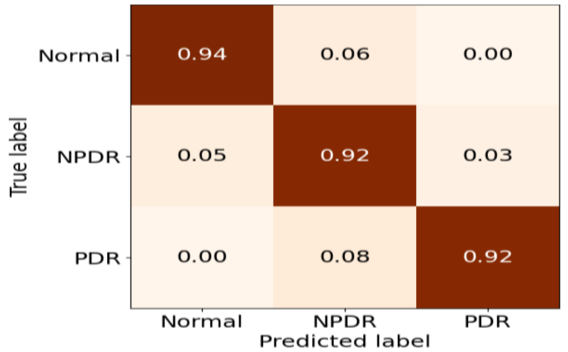

Fig. 10. Confusion matrix of Xception model for fundas image classification

The above Figure 10 shows the confusion matrix of proposed Xception model with different classes such that diagonally representation of different classes shows the positive correlation and other variable show the negative correlation for fundas image classification. In Figure x-axis shows the predicated label and y-axis shows the true labels of input data classes that are PDR, NPDR and normal. The highly predicated images of normal dataset that contain 94% accuracy. NPDR get 92% accuracy and PDR class obtain 92% accuracy with help of Xception model for fundas image classification.

Table 3. Parameters Performance of Xception model for fundas image classification

Model	*Training accuracy*	*Validation Accuracy*	*F1-score*	*Recall*	*Precision*
Xception	98.98	92.74	92.78	92.74	92.87

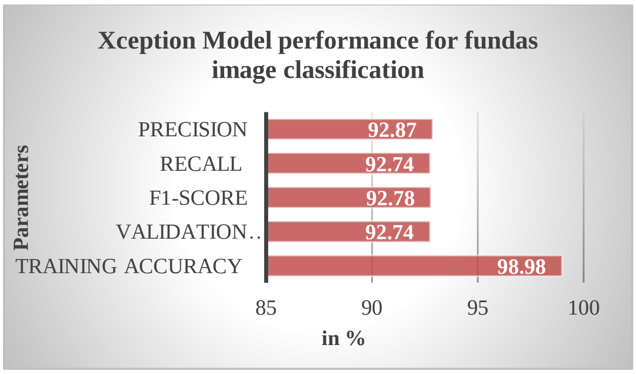

Fig. 11. Xception Model performance of parameter for fundas image classification

As shown in Table 3 and Figure 11, the model's fundas image classification performance characteristics are 98.98% training accuracy, 92.74% validation accuracy, 92.78% F1-score, 92.74% recall, and 92.87% precision.

Table 4. 5-Fold cross-validation performance of Xception model with classification of fundus images

K-Folds	*Training accuracy*	*Validation Accuracy*	*F1-score*	*Recall*	*Precision*
Fold 1	97.63	77.52	77.66	77.52	77.95
Fold 2	99.37	93.73	93.76	93.73	93.82
Fold 3	99.65	97.56	97.56	97.56	97.59
Fold 4	99.42	96.44	97.45	97.44	97.48
Fold 5	98.82	97.44	97.45	97.44	97.50

The 5-fold cross-validation results for the Xception model demonstrate its consistency and robustness in classifying fundus images. Each fold shows high training accuracy, ranging from 97.63% to 99.65%, indicating effective learning across different subsets of data. Validation accuracy varies, with the first fold at 77.52% and subsequent folds significantly higher, reaching up to 97.56%. This suggests that while initial splits might face challenges, the model overall generalises well. High F1 scores, recall, and precision in folds 2 to 5 (all above 93%) further affirm the model's reliability and precision in predicting positive cases accurately. The substantial performance in later folds, especially with validation accuracies above 96%, underscores the model's robustness and suitability for reliable fundus image classification.

We conducted the following trials twice to compare the three models' performance. Table 5 displays the findings of the test dataset, which were obtained after each model was trained on the training dataset.

Table 5. Comparison between base and proposed models for fundas image classification of eye disease

Model	*Accuracy*	*F1-score*	*Recall*	*Precision*
DenseNet121 and Inception-ResNetV2	91	90	91	91
Xception	98.98	92.78	92.74	92.87

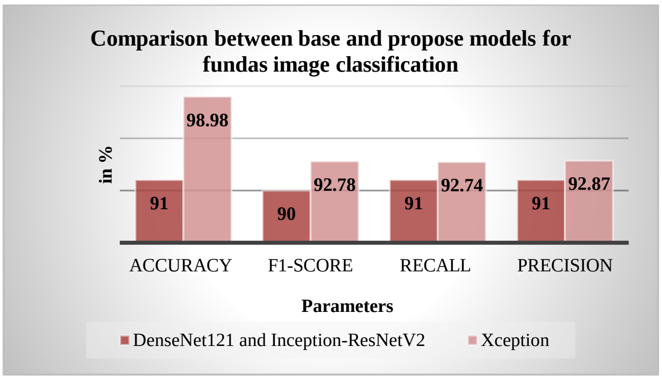

Fig. 12. Bar graph of base and proposed models' comparison for fundas image classification of eye disease

Table 4 and Figure 12 compare base and suggested models for fundas image categorisation of eye disease. suggested models exhibit 98.98% accuracy, 92.78% f1 score, 92.74 recall, and 92.87% precision. The accuracy, recall, F1 score, and precision for this simple model are 91%, 90%, and 91% when dense net 121 and inception resnetv2 are combined. It demonstrates that the proposed model (Xception) has produced satisfactory results for fundas image classification of eye diseases.

The study develops a deep learning model using the Xception architecture for automated classification of eye diseases from fundus images. With 92.87% precision, 98.98% accuracy, 92.74% recall, and 92.78% F1-score, the model performs well across all assessment criteria. It effectively distinguishes between normal fundus images and those indicative of diabetic retinopathy, spanning various severity levels. It uses modern preprocessing methods and applies 5-fold cross-validation to improve the stability and, at the same time, the robustness of the models built. These findings show that the Xception model can accurately diagnose and treat eye diseases by analysing medical images. This can improve healthcare technology and clinical medicine.

5. Conclusion and Future Work

At present, fundus photography of the retina is still relied on by doctors to diagnose DR through clinical examination. Compared to manual techniques, automated methods are more dependable, economical, and efficient in terms of efficiency. The authors of this study provide a deep learning (DL) model that can automatically identify DR images by dividing retinal fundus images into three distinct categories. Using fundus images and the Xception architecture, this study set out to create a deep-learning model that could effectively identify and classify eye illnesses. The DDR dataset was used using a 5-fold cross-validation technique; it included preprocessed, scaled, and filtered images that were labelled into three groups. Training accuracy was 98.98%, validation accuracy was 92.74%, precision was 92.87%, recall was 92.74%, and the F1-score was 92.78% for the fine-tuned Xception model that was optimised using the Adam optimiser. The confusion matrix showed high accuracy for the normal (94%), NPDR (92%), and PDR (92%) classes. Comparatively, the Xception model outperformed the DenseNet121 and Inception-ResNetV2 base models, which achieved a maximum accuracy of 91%. These findings highlight the enhanced efficacy and potential of the suggested model as a trustworthy diagnostic tool for eye illnesses, improving patient outcomes and ocular health care. The suggested method performs competitively on the DR fundus image categorisation. All three classes in the dataset we gathered, however, had an equal amount of images. Prospective developments for this research include expanding our dataset and including NPDR classification images, categorised into mild, moderate, and severe classes, to tackle more intricate classification tasks.

While the proposed Xception model demonstrates high accuracy and reliability in classifying eye diseases from fundus images, it has limitations, including the potential lack of generalizability to diverse populations and imaging conditions, as the DDR dataset may not encompass all demographic and clinical variations. Additionally, the model's interpretability remains a challenge, which can hinder clinical adoption. Future work should focus on enhancing model generalizability by incorporating more diverse and comprehensive datasets, improving interpretability through the development of explainable AI techniques, and exploring the integration of multi-modal data to provide a more holistic approach to eye disease diagnosis. Furthermore, addressing real-world applicability and scalability in clinical settings, along with ensuring ethical considerations like data privacy and mitigating algorithmic bias, are essential steps for broader implementation.

References

[1] T. Nazir, A. Irtaza, A. Javed, H. Malik, D. Hussain, and R. A. Naqvi, "Retinal image analysis for diabetes-based eye disease detection using deep learning," *Appl. Sci.*, 2020, doi: 10.3390/APP10186185.

[2] P. Saeedi *et al.*, "Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition," *Diabetes Res. Clin. Pract.*, 2019, doi: 10.1016/j.diabres.2019.107843.

[3] "Retinal Image Analysis for Diabetes-Based Eye Disease Detection Using Deep Learning," *Int. J. Progress. Res. Eng. Manag. Sci.*, 2023, doi: 10.58257/ijprems31685.

[4] H. Mustafa, S. F. Ali, M. Bilal, and M. S. Hanif, "Multi-Stream Deep Neural Network for Diabetic Retinopathy Severity Classification under a Boosting Framework," *IEEE Access*, 2022, doi: 10.1109/ACCESS.2022.3217216.

[5] P. Porwal *et al.*, "Indian diabetic retinopathy image dataset (IDRiD): A database for diabetic retinopathy screening research," *Data*, 2018, doi: 10.3390/data3030025.

[6] S. Patil *et al.*, "Artificial Intelligence in the Diagnosis of Oral Diseases: Applications and Pitfalls," *Diagnostics*. 2022. doi: 10.3390/diagnostics12051029.

[7] S. Sengupta, A. Singh, H. A. Leopold, T. Gulati, and V. Lakshminarayanan, "Ophthalmic diagnosis using deep learning with fundus images – A critical review," *Artificial Intelligence in Medicine*. 2020. doi: 10.1016/j.artmed.2019.101758.

[8] J. H. Tan, U. R. Acharya, S. V. Bhandary, K. C. Chua, and S. Sivaprasad, "Segmentation of optic disc, fovea and retinal vasculature using a single convolutional neural network," *J. Comput. Sci.*, 2017, doi: 10.1016/j.jocs.2017.02.006.

[9] R. Chavan and D. Pete, "Automatic multi-disease classification on retinal images using multilevel glowworm swarm convolutional neural network," *J. Eng. Appl. Sci.*, 2024, doi: 10.1186/s44147-023-00335-0.

[10] R. H. Paradisa, A. Bustamam, W. Mangunwardoyo, A. A. Victor, A. R. Yudantha, and P. Anki, "Deep feature vectors concatenation for eye disease detection using fundus image," *Electron.*, 2022, doi: 10.3390/electronics11010023.

[11] J. Colcombe, R. Mundae, A. Kaiser, J. Bijon, and Y. Modi, "Retinal Findings and Cardiovascular Risk: Prognostic Conditions, Novel Biomarkers, and Emerging Image Analysis Techniques," *Journal of Personalized Medicine*. 2023. doi: 10.3390/jpm13111564.

[12] B. Goutam, M. F. Hashmi, Z. W. Geem, and N. D. Bokde, "A Comprehensive Review of Deep Learning Strategies in Retinal Disease Diagnosis Using Fundus Images," *IEEE Access*. 2022. doi: 10.1109/ACCESS.2022.3178372.

[13] S. Dubey and M. Dixit, "Recent developments on computer aided systems for diagnosis of diabetic retinopathy: a review," *Multimed. Tools Appl.*, 2023, doi: 10.1007/s11042-022-13841-9.

[14] M. Vadduri and P. Kuppusamy, "Enhancing Ocular Healthcare: Deep Learning-Based Multi-Class Diabetic Eye Disease Segmentation and Classification," *IEEE Access*, 2023, doi: 10.1109/ACCESS.2023.3339574.

[15] R. Sarki, K. Ahmed, H. Wang, Y. Zhang, J. Ma, and K. Wang, "Image Preprocessing in Classification and Identification of Diabetic Eye Diseases," *Data Sci. Eng.*, 2021, doi: 10.1007/s41019-021-00167-z.

[16] N. Chea and Y. Nam, "Classification of Fundus Images Based on Deep Learning for Detecting Eye Diseases," *Comput. Mater. Contin.*, 2021, doi: 10.32604/cmc.2021.013390.

[17] S. Ortiz and M. A. Goenaga Jimenez, "Deep Learning-Based Ocular Disease Classification in Fundus Images," in *1st IEEE Colombian Caribbean Conference, C3 2023*, 2023. doi: 10.1109/C358072.2023.10436234.

[18] M. A. Abini and S. S. S. Priya, "A Deep Learning Framework for Detection and Classification of Diabetic Retinopathy in Fundus Images Using Residual Neural Networks," in *9th International Conference on Smart Computing and Communications: Intelligent Technologies and Applications, ICSCC 2023*, 2023. doi: 10.1109/ICSCC59169.2023.10335079.

[19] R. Chavan and D. Pete, "Classification of Retinal Fundus Images using VGG16 and Inception V3," in *2023 14th International Conference on Computing Communication and Networking Technologies, ICCCNT 2023*, 2023. doi: 10.1109/ICCCNT56998.2023.10306593.

[20] G. R. S. N. Kumar, R. S. Sankuri, and S. P. K. Karri, "Multi Scale aided Deep Learning model for High F1-score classification of fundus images based Diabetic Retinopathy and Glaucoma," in *2023 International Conference on Computer, Electronics and Electrical Engineering and their Applications, IC2E3 2023*, 2023. doi: 10.1109/IC2E357697.2023.10262434.

[21] H. Y. Sam, S. A. Zikri B Sayed Aluwee, N. S. Bt Che Lah, C. M. Goh, C. M. Tyng, and H. Ikhwan B Murat, "Preliminary Study of Diabetic Retinopathy Classification from Fundus Images Using Deep Learning Model," in *2022 3rd International Conference on Artificial Intelligence and Data Sciences: Championing Innovations in Artificial Intelligence and Data Sciences for Sustainable Future, AiDAS 2022 - Proceedings*, 2022. doi: 10.1109/AiDAS56890.2022.9918681.

[22] K. Chauhan, Kashish, K. Dagar, and R. K. Yadav, "Cataract detection from eye fundus image using an ensemble of transfer learning models," in *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering, ICACITE 2022*, 2022. doi: 10.1109/ICACITE53722.2022.9823638.

[23] A. A. Mukhlif, B. Al-Khateeb, and M. A. Mohammed, "Breast cancer images Classification using a new transfer learning technique," *Iraqi J. Comput. Sci. Math.*, 2023, doi: 10.52866/ijcsm.2023.01.01.0014.

[24] M. M. A. Shibly, T. A. Tisha, T. A. Tani, and S. Ripon, "Convolutional neural network-based ensemble methods to recognize Bangla handwritten character," *PeerJ Comput. Sci.*, 2021, doi: 10.7717/peerj-cs.565.

[25] R. E. Mandiya, H. M. Kongo, S. K. Kasereka, K. Kyandoghere, P. M. Tshakwanda, and N. M. Kasoro, "Enhancing COVID-19 Detection: An Xception-Based Model with Advanced Transfer Learning from

X-ray Thorax Images," *J. Imaging*, vol. 10, no. 3, 2024, doi: 10.3390/jimaging10030063.
[26] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2015. doi: 10.1109/CVPR.2015.7298594.
[27] S. Ioffe and C. Szegedy, "Batch normalisation: Accelerating deep network training by reducing internal covariate shift," in *32nd International Conference on Machine Learning, ICML 2015*, 2015.
[28] A. Tharwat, "Classification assessment methods," *Appl. Comput. Informatics*, 2018, doi: 10.1016/j.aci.2018.08.003.