International Science and Technology Journal
المجلة الدولية للعلوم والتقنية

عدد خاص بالمؤتمر الليبي الدولي للعلوم التطبيقية و الهندسية دورته الثانية
LICASE -2
2024 / 10 / 30-29

المجلة الدولية للعلوم والتقنية
ISTJ

تم استلام الورقة بتاريخ:27/ 9/2024م
وتم نشرها على الموقع بتاريخ: 30/ 10/2024م

Advancing Eye Disease Diagnosis with Deep Learning: A ResNet18 Model Approach

http://www.doi.org/10.62341/ayms2390

Ayyah Mustafa Salih

Electrical and Electronic Engineering, University of Benghazi, Libya
ayaalfagih98@gmail.com

Abstract

This paper addresses the classification and prediction of eye diseases, specifically Glaucoma, Diabetic Retinopathy, Cataract, and normal cases, utilizing advanced deep learning techniques. The primary objective is to assess the performance of the ResNet18 model, enhanced with transfer learning and various optimizer methods, in accurately identifying these conditions. The dataset used in this study consists of 1757 retinal fundus images, including 453 normal cases, 441 cases of Diabetic Retinopathy, 476 cases of Glaucoma, and 387 cases of Cataract. The images were resized to 224x224 pixels and normalized for training.

The ResNet18 model, pre-trained on the ImageNet dataset, is fine-tuned through transfer learning to specialize in eye disease classification. Three optimization methods—SGDM, ADAM, and RMSProp—are employed to optimize the learning process. Cross-validation, specifically a 5-fold method, is utilized during training to enhance the model's reliability and robustness. The performance metrics, which include accuracy, precision, recall, and the Area Under the Curve (AUC), demonstrate a high accuracy rate of 91.4% and an exceptional AUC value of 99.99%.

While the results confirm the effectiveness of the model in identifying various eye diseases, there were challenges such as overfitting and the need for further with external data. Future research could explore the performance of the model on larger, independent datasets to validate its generalization capabilities and expand the range of eye diseases considered.

Keywords: Deep learning, Eye diseases, Glaucoma, Diabetic Retinopathy, Cataract, transfer learning, ResNet18, SGDM, ADAM, RMSProp, classification, cross-validation.

International Science and Technology Journal
المجلة الدولية للعلوم والتقنية

عدد خاص بالمؤتمر الليبي الدولي للعلوم التطبيقية و الهندسية دورته الثانية
LICASE -2
2024 / 10 / 30-29

المجلة الدولية للعلوم والتقنية
ISTJ

تم استلام الورقة بتاريخ:27/ 9/2024م وتم نشرها على الموقع بتاريخ: 30/ 10/2024م

تطوير تشخيص أمراض العيون باستخدام التعلم العميق: نهج يعتمد على نموذج ResNet18

آية مصطفى صالح

قسم الهندسة الكهربائية والإلكترونية، جامعة بنغازي، ليبيا

ayaalfagih98@gmail.com

الملخص

تتناول هذه الورقة التصنيف والتنبؤ بأمراض العيون، وتحديدًا الجلوكوما واعتلال الشبكية السكري وإعتام عدسة العين والحالات الطبيعية، باستخدام تقنيات التعلم العميق المتقدمة. والهدف الأساسي هو تقييم أداء نموذج ***ResNet18*** ، المعزز بالتعلم الانتقالي وطرق التحسين المختلفة، في تحديد هذه الحالات بدقة. تتكون مجموعة البيانات المستخدمة في هذه الدراسة من 1757 صورة لقاع الشبكية، بما في ذلك 453 حالة طبيعية، و441 حالة من اعتلال الشبكية السكري، و476 حالة من الجلوكوما، و387 حالة من إعتام عدسة العين. تم تغيير حجم الصور إلى 224 × 224 بكسل وتطبيعها للتدريب.

تم ضبط نموذج ***ResNet18*** ، المدرب مسبقًا على مجموعة بيانات***ImageNet*** ، من خلال التعلم الانتقالي للتخصص في تصنيف أمراض العيون. تم استخدام ثلاث طرق تحسين ***SGDM*** – و ***ADAM***و – ***RMSProp*** لتحسين عملية التعلم. يتم استخدام التحقق المتبادل، وخاصة طريقة 5 أضعاف، أثناء التدريب لتعزيز موثوقية النموذج وقوته. تُظهر مقاييس الأداء، التي تشمل الدقة والدقة والتذكير والمساحة تحت المنحنى ***(AUC)*** ، معدل دقة مرتفع بنسبة 91.4٪ وقيمة ***AUC*** استثنائية بنسبة 99.99٪.

في حين تؤكد النتائج على فعالية النموذج في تحديد أمراض العيون المختلفة، فقد كانت هناك تحديات مثل الإفراط في الملاءمة والحاجة إلى مزيد من البيانات الخارجية. ويمكن للبحوث المستقبلية استكشاف أداء النموذج على مجموعات بيانات أكبر ومستقلة للتحقق من قدراته على التعميم وتوسيع نطاق أمراض العيون التي تم النظر فيها.

الكلمات المفتاحية *:* التعلم العميق، أمراض العيون، الجلوكوما، اعتلال الشبكية السكري، إعتام عدسة العين، التعلم بالنقل، ***ResNet18***، ***SGDM***، ***ADAM***، ***RMSProp***، التصنيف، التحقق المتبادل.

1. INTRODUCTION

Eye diseases such as glaucoma, diabetic retinopathy, and cataracts are among the leading causes of visual impairment and blindness globally. If these conditions are not detected or treated in a timely manner, they can lead to severe vision loss or even permanent blindness. According to estimates from the World Health Organization (WHO), over a billion people worldwide suffer from vision impairment, with a significant portion of these cases attributed to preventable or treatable eye conditions like glaucoma and

diabetic retinopathy [1]. Early detection and accurate diagnosis of these diseases are critical for preventing their progression and minimizing the risk of permanent blindness.

Glaucoma is a group of eye conditions that cause damage to the optic nerve, often due to elevated intraocular pressure. It is one of the leading causes of blindness, particularly among older adults. The vision loss caused by glaucoma is irreversible, and the disease often progresses without noticeable symptoms until it reaches an advanced stage. Thus, early detection is crucial for managing the disease and preventing further vision loss. However, traditional diagnostic methods for glaucoma usually require highly trained specialists and sophisticated equipment, which can be expensive and inaccessible, especially in low-resource settings[2]. This underscores the need for automated diagnostic systems that can assist in the early detection of glaucoma.

Diabetic retinopathy (DR) is a major cause of blindness, particularly in individuals with diabetes. High blood sugar levels can cause damage to the blood vessels in the retina, leading to swelling, leakage, and in severe cases, the formation of abnormal new blood vessels that can result in significant vision problems. Diabetic retinopathy progresses through different stages, and early detection and timely intervention are essential to prevent vision loss. However, similar to glaucoma, the diagnosis of diabetic retinopathy often relies on the manual examination of retinal images by ophthalmologists, which can be time-consuming and subject to human error [3]. This highlights the need for automated systems capable of accurately classifying and diagnosing various stages of diabetic retinopathy.

Cataracts, characterized by the clouding of the eye's natural lens, are the leading cause of blindness worldwide, particularly among the elderly. Cataracts cause blurred vision, reduced contrast sensitivity, and difficulty seeing in low-light conditions. While cataract surgery is an effective treatment to restore vision, early detection is important to ensure timely surgical intervention. Although diagnosing cataracts is generally straightforward, many individuals, especially in regions with limited access to healthcare, suffer from untreated cataracts, emphasizing the need for cost-effective and scalable diagnostic tools [4].

1.1. Economic and Social Impact

The widespread prevalence of these eye diseases not only affects individuals on a personal level but also places a significant economic burden on healthcare systems globally. The costs associated with treating advanced cases of glaucoma, diabetic retinopathy, and cataracts are substantial, and the associated loss of productivity further exacerbates the economic impact. According to studies, the global economic burden of visual impairment due to these conditions is estimated in the hundreds of billions of dollars annually [5]. Automated diagnostic systems have the potential to alleviate much of this burden by

تم استلام الورقة بتاريخ:27/ 9/2024م وتم نشرها على الموقع بتاريخ: 30/ 10/2024م

enabling early detection and intervention, thereby improving patient outcomes and reducing healthcare costs.

1.2. The Role of Technology in Diagnosis

Recent advancements in medical imaging and artificial intelligence, particularly deep learning, have spurred interest in developing automated systems for detecting and classifying eye diseases. Convolutional Neural Networks (CNNs), a type of deep learning model, have demonstrated significant success in image classification tasks, making them well-suited for analyzing retinal images and diagnosing various eye diseases. These models can process large volumes of visual data, identifying patterns and features that signal the presence of specific diseases. By automating the diagnostic process, these systems can assist healthcare professionals in making faster, more accurate diagnoses, reducing the workload on medical staff, and making ophthalmic care more accessible, especially in underserved regions [6].

1.3. The Need for an Effective Diagnostic System

Despite the potential of deep learning models, several challenges remain in developing an effective diagnostic system for eye diseases. One of the key challenges is ensuring that the model generalizes well across different datasets, which is essential for real-world clinical deployment. Moreover, optimizing the model to achieve a balance between accuracy and computational efficiency is critical, especially in medical settings that require rapid results without compromising diagnostic accuracy. Additionally, while previous research has made progress in developing models for individual diseases like glaucoma or diabetic retinopathy, there is a need for a comprehensive system that can diagnose multiple eye diseases simultaneously, providing a more holistic approach to ophthalmic care.

2. Study Objective

This study aims to address these challenges by developing a robust, deep learning-based diagnostic system capable of accurately classifying multiple eye diseases, including glaucoma, diabetic retinopathy, and cataracts. The system is based on the ResNet18 architecture, which has shown excellent performance in image classification tasks. Through transfer learning, the model is fine-tuned on a dataset of retinal images, allowing it to specialize in eye disease classification while benefiting from pre-trained features learned on a larger, more diverse dataset [7]. The model is optimized using various techniques, including SGDM (Stochastic Gradient Descent with Momentum), Adam (Adaptive Moment Estimation), and RMSProp, to improve performance and generalization capabilities.

3. Problem Statement

Despite the advances in medical imaging technologies, the diagnosis of eye diseases remains highly dependent on human expertise, which can be prone to errors and time-

consuming. This reliance creates a critical challenge in handling the large volume of data generated daily in clinics and hospitals. Automated diagnostic systems, powered by deep learning, offer a promising solution to this problem by providing fast, accurate, and scalable methods for disease detection. However, much of the existing research lacks a comprehensive and efficient model that combines high accuracy with generalization across multiple eye diseases.

4. Research Gap

While several studies have explored the application of deep learning in diagnosing individual eye diseases such as glaucoma or diabetic retinopathy, there is a notable gap in models that can generalize across multiple diseases with high accuracy. Furthermore, most of the current research focuses on specific datasets or optimization methods without exploring the full potential of transfer learning and advanced optimizers. This study aims to fill this gap by leveraging the ResNet18 architecture, alongside transfer learning techniques and multiple optimization methods, to create a robust diagnostic system capable of classifying several eye diseases.

5. LITERATURE SURVEY

Several research efforts have successfully employed deep learning models, particularly convolutional neural networks (CNNs), for the classification of eye diseases [6, 8, 9]

In the field of glaucoma detection, numerous studies have employed deep learning algorithms, particularly convolutional neural networks (CNNs), to automate the detection process [10]. For instance,Chen et al.(2015) created a CNN architecture with hierarchical feature representation to distinguish between glaucoma and non-glaucoma patterns. Their model, which comprises six multilayers, achieved AUC values of 0.8321 and 0.887 on the ORIGA and SCES datasets, respectively [11] .

Another approach by Asaoka et al. (2016) integrated deep feed-forward neural networks with other machine learning classifiers to form a deep ensemble solution. They achieved an AUC value of 92.5% with the deep FNN classifier. These studies illustrate the effectiveness of deep learning algorithms in detecting glaucoma. [12]

Deep learning algorithms have found applications in diabetic retinopathy for feature extraction and classification of retinal images[13].

Several experiments employing various convolutional architectures were conducted, and their performance was assessed on the MESSIDOR database using cross-validation. Among the architectures tested, a ResNet50-based model delivered the best results, achieving an AUC of 0.93 for grades 0 and 1, 0.81 for grade 2, and 0.92 for grade 3 labeling. Additionally, when tackling a binary classification problem, the AUCs exceeded 0.97[14].

Another study advocated for the use of transfer learning and fine-tuning techniques in the diagnosis of diabetic retinopathy (DR). The researchers specifically focused on applying ResNet18, a widely recognized deep learning architecture, using the publicly accessible Kaggle dataset. They approached the DR diagnosis as a multi-class classification task, categorizing the disease into five severity levels ranging from 0 (No DR) to 4 (Proliferative DR). By leveraging transfer learning and fine-tuning, the researchers achieved promising results. The final ResNet18 model recorded an accuracy of 70%, a recall of 50%, and a specificity of 88%. Notably, this model outperformed others developed from scratch, illustrating the effectiveness of transfer learning in this context. Additionally, the approach demonstrated shorter training times, further highlighting its efficiency.[15]

Deep learning algorithms have also been utilized in the detection and diagnosis of cataracts. For instance, researchers developed a deep learning framework using the ResNet18 model to classify cataract severity levels based on slit-lamp images. Their model achieved high accuracy in differentiating between various stages of cataracts. [16] Additionally, another study introduced a deep learning approach for cataract diagnosis that integrated ResNet18 with attention mechanisms. This model demonstrated superior performance compared to other methods in terms of accuracy and AUC values. These studies underscore the effectiveness of deep learning, especially the ResNet18 model, in detecting cataracts and classifying their severity. [17]**.**

6. METHODOLOGY

The primary motivation for utilizing ResNet18 in the study is that it possesses distinct constraints, features, architecture, and trainable variables compared to other networks.

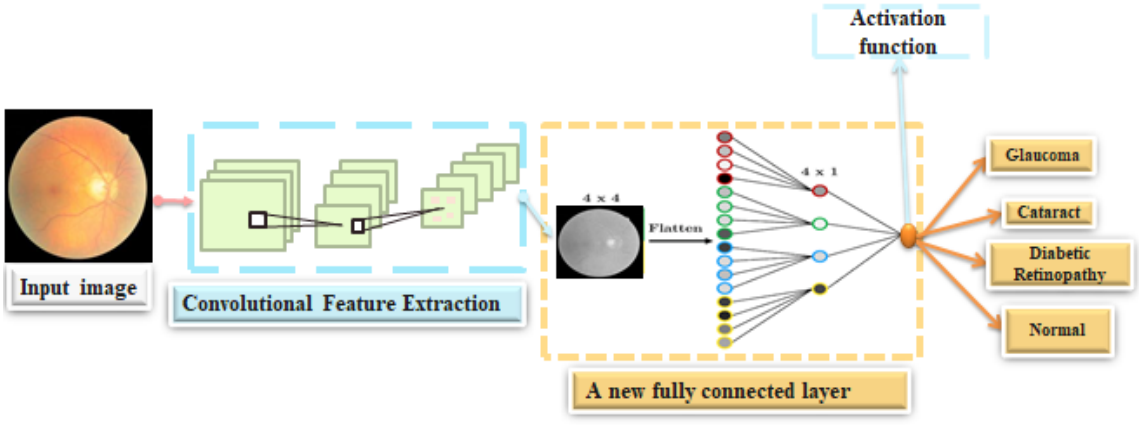

Figure 1. Shown the steps of the methodology

6.1. Data Collection and Preprocessing

In this paper, the dataset used for training the model was collected from **Kaggle [18]** The dataset, titled "**eye_diseases_classification,**" consists of 1757 retinal fundus images, which include 453 **normal** images, 441 images with **diabetic retinopathy**, 476 images with **glaucoma**, and 387 images with **cataract**. These images are in RGB format and have varying sizes, providing a diverse set of samples for the training and evaluation of the model.A sample of this dataset is presented in Figure 2.

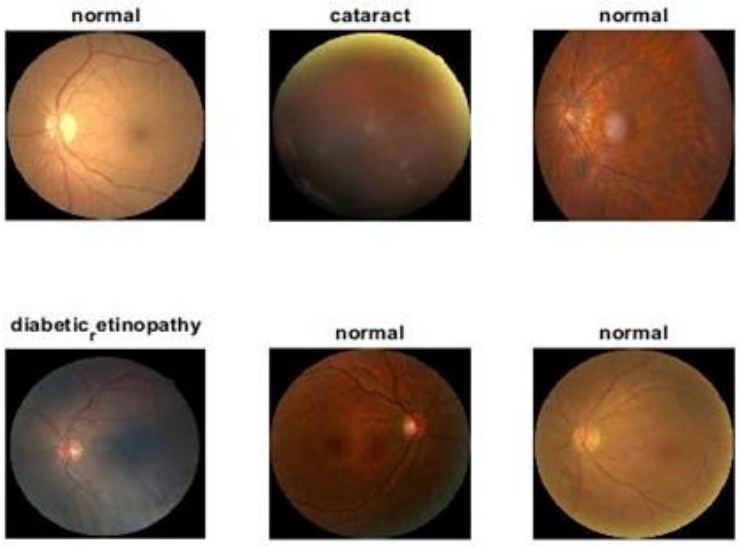

Figure2. some samples of "eye_ diseases _classification" dataset images.

After loading each image from the dataset, resizes the images to a fixed size, 224x224 pixels, and normalizes the pixel values of the resized images. A common approach for normalization is to scale the pixel values to the range of [0, 1][19].

6.2. Enhancing the ResNet18 Model Using Transfer Learning Techniques

The **ResNet18** model, a well-known deep neural network recognized for its high performance in image classification, served as the foundation for developing a diagnostic system for eye diseases. To adapt the model for the task of classifying eye diseases, transfer learning techniques were applied. This approach involves using a pre-trained model on a large and diverse dataset [10], such as ImageNet, and then fine-tuning the last layers of the model on the specific dataset of eye diseases.

Key Information about ResNet18:

1) Model Architecture:

ResNet-18 is composed of 18 layers as shown in Figure 3, featuring convolutional

layers, batch normalization, ReLU activation functions, and residual blocks. The model incorporates skip connections to facilitate information flow from earlier to later layers, effectively addressing the vanishing gradient problem. Each residual block contains two convolutional layers with a shortcut connection bypassing them. The model concludes with a fully connected layer for classification.[7, 20]

2) Pretrained Models:

- **ResNet-18** is accessible as a pretrained model in popular deep learning frameworks such as **MATLAB** [21].
- Pretrained models are trained on large-scale image datasets, such as ImageNet, and can be used for transfer learning or as a starting point for training on specific tasks.[10]
- The output of **ResNet-18** is a probability distribution over the classes in the dataset.

3) Usage:

- ResNet-18 can be utilized for various computer vision tasks, including image classification, object detection, and image segmentation.
- The model requires input images to be of size 224x224 pixels and necessitates normalization of the input data.[7, 22]

4) Performance:

- **ResNet-18** has achieved impressive performance on benchmark datasets like ImageNet.
- The top-1 error rate of ResNet-18 on ImageNet is 30.24%, and the top-5 error rate is10.92%. (mailto:(https://pytorch.org/hub/pytorch_vision_resnet/).

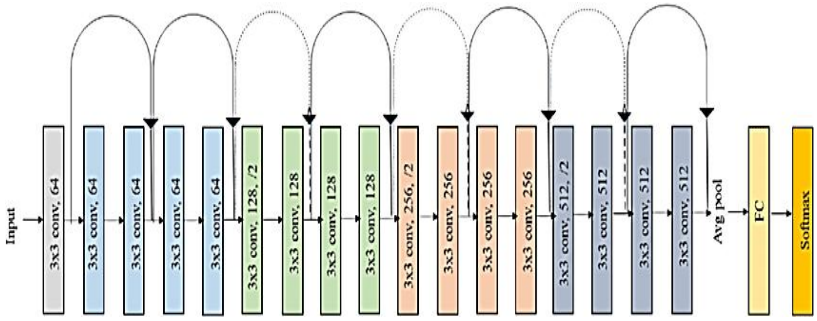

Figure 3. ResNet-18 is a convolution neural network architecture[23]

6.3.Optimizations on Optimization Methods

3.3.1 Comparative Analysis of Adam, SGDM, and Gradient Descent

In deep learning, the choice of optimization algorithm significantly impacts training efficiency and model performance. This paper employs three optimizers—Adam, Stochastic Gradient Descent with Momentum (SGDM), and Gradient Descent—to

enhance the model's accuracy and speed, thereby improving its performance in classifying eye diseases. Below is a brief description of each optimizer:

1) **SGDM** (Stochastic Gradient Descent with Momentum): This method combines the traditional stochastic gradient descent algorithm with the concept of momentum. The momentum helps in accelerating the convergence of the training process and prevents the model from getting stuck in local minima the potential bias that could arise from a single train-test split [24, 25].
2) **Adam** (Adaptive Moment Estimation): This method combines the advantages of both SGDM and RMSProp. It uses different learning rates for each parameter based on the first and second moment estimates, making it effective and efficient in reaching optimal solutions[25].
3) **RMSProp** (Root Mean Square Propagation):This method focuses on dynamically adjusting the learning rate based on gradient oscillations, helping to control high learning rates and prevent model oscillations during training.

The experiments in this article evaluate the performance of these optimizers on the eye_ diseases dataset, using binary cross-entropy loss as the evaluation metric. Specific details about the hyperparameters, such as learning rate (0.00001), epochs (20), and batch size (30), are provided to ensure a comprehensive and reproducible analysis.

6.4. Cross-Validation

Cross-validation is a widely adopted technique in machine learning for assessing model performance and generalization ability. I utilized this technique in the training process of the model used for classifying eye diseases. The 5-fold cross-validation method was employed, which is a well-established and effective approach for model validation.

By splitting the dataset into five subsets and iteratively training the model on four subsets while evaluating on the remaining one, a more robust and unbiased estimate of the model's performance is obtained. This method provides a more comprehensive evaluation, as shown in Figure 3.The main advantages of using 5-fold cross-validation in this research are[26]:

- **Improved Generalization:** Cross-validation helps in estimating the model's performance on unseen data, giving a better indication of its generalization capabilities [27].
- **Reliable Performance Metrics:** By reporting the average performance metrics across all five folds, a more reliable and representative assessment of the model's

effectiveness is obtained [28].

- **Mitigation of Overfitting**: Cross-validation helps identify and address potential overfitting issues by evaluating the model's performance on held-out data during each iteration [29].

The choice of a 5-fold approach is common and well-justified, as it strikes a balance between the number of iterations and the size of the validation sets. This ensures that the model is thoroughly evaluated while maintaining a reasonable computational cost.

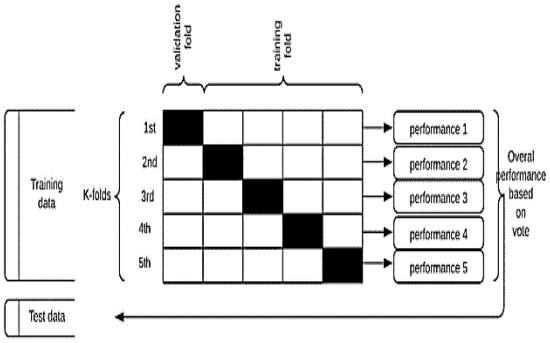

Figure 3. Repeated stratified k-fold cross validation procedure applied to each classification algorithm.[30].

6.5. Model evaluation

To evaluate the model's performance on the test set and obtain a confusion matrix along with other performance metrics, this study follows these steps:

1. Load the trained model that has been previously trained on the training and validation sets.
2. Iterate through the test set and input the test samples into the model to obtain predictions.
3. Compare the predicted labels with the true labels of the test set to construct a confusion matrix.

Calculate additional performance metrics such as accuracy, precision, recall, F1-score, and the Area Under the Receiver Operating Characteristic curve (AUROC).

6.5.1. Confusion matrix

A confusion matrix is a widely used tool for evaluating the performance of classification models. It provides a detailed breakdown of the model's predictions compared to the actual labels, allowing for a more granular analysis of classification accuracy. In this study, the confusion matrix is used to assess the model's ability to classify four eye disease categories: **Glaucoma**, **Diabetic Retinopathy**, **Cataract**, and **Normal.**

TABLE I. Confusion matrix for a classification

	Predicted			
	Glaucoma	Diabetic Retinopathy	Cataract	Normal
Glaucoma	TP	FN	FN	FN
Diabetic Retinopathy	FN	TP	FN	FN
Cataract	FN	FN	TP	FN
Normal	FN	FN	FN	TP

6.5.2. Key Terms in the Confusion Matrix from TABLE I [31]:

- **TP (True Positive)**: The number of samples correctly predicted as a specific class[32]. For example, the value in the (1,1) cell represents the number of **Glaucoma** samples that were correctly predicted as **Glaucoma**.
- **FN (False Negative):** The number of samples incorrectly predicted as not belonging to a specific class when they actually do[33]. For example, the value in the (1,2) cell represents the number of **Glaucoma** samples that were incorrectly classified as **Diabetic Retinopathy**.
- **FP (False Positive):** The number of samples incorrectly predicted as belonging to a specific class when they actually do not [33]. For example, **Glaucoma** samples predicted as **Diabetic Retinopathy** would be counted as false positives for **Diabetic Retinopathy**.
- **TN (True Negative):** The number of samples correctly predicted as not belonging to a specific class[33]. **TN** is not explicitly shown in the confusion matrix but can be calculated from the total number of samples minus the TPs, FPs, and FNs.

6.5.3. Performance Metrics Derived from the Confusion Matrix:

From the confusion matrix, several important performance metrics can be derived to assess the model's classification performance. The equations provided below are the mathematical representations of the statistical measures used to analyze the performance of the eye disease classification deep learning algorithm.

1) Accuracy:

$$Accuracy = \frac{(TN + TP)}{(TN + TP + FN + FP)} \qquad (1)$$

Accuracy quantifies the overall correctness of the model's predictions compared to the actual labels [34].Accuracy ranges from 0 to 1, with 1 representing perfect accuracy.

2) Precision:

$$Precision = \frac{TP}{(TP + FP)} \qquad (2)$$

Precision indicates how many of the predicted positive samples were actually correct. It is particularly useful in situations where false positives are costly [35].

3) Recall (Sensitivity):

$$Recall = \frac{TP}{(TP\ +\ FN)} \qquad (3)$$

Recall measures how well the model identifies true positive samples. It is crucial in medical diagnostics where missing a true positive can have serious consequences [36].

4) F1-Score (AUC) :

$$F1_{Score} = 2 \times \frac{PrecisionxRecall}{(Precision + Recall)} \qquad (4)$$

The F1-Score is the harmonic mean of precision and recall and provides a balanced measure of both, especially when the dataset is imbalanced [37].

5) **AUC (Area Under the Curve): AUC** is commonly used as an evaluation metric for binary classification tasks, particularly in **ROC** (Receiver Operating Characteristic) curve analysis. The **AUC** represents the area under the **ROC** curve, which plots the true positive rate (TPR) against the false positive rate (FPR) at various classification thresholds. The **AUC** value ranges from 0 to 1, with 1 indicating a perfect classifier.[37]

7. Results

The eye-diseases-classification-Deep system was tested on a personal laptop. The eye-diseases-classification-Deep process was implemented in **MATLAB®** and the experiments were conducted using MATLAB-2021a. The workstation used for the experiments was equipped with an Intel Core i7 HP M4000 CPU, 8 GB RAM, and a 64-bit operating system.

7.1. Experimental Results

- **Image Specifications:**

– The dataset from kaggle dataset [18] consisted of **1757 retinal fundus images,** divided into four categories: **453 normal** images**, 441** images with **Diabetic Retinopathy**, **476 images with Glaucoma,** and **387 images** with **Cataracts**.

– All images were resized to a uniform size of **224x224 pixels** and converted to **RGB** format. To normalize the pixel values, they were scaled to the [0, 1] range.

- Basic **data augmentation** techniques were applied, including random rotations, flips, and contrast adjustments, to help prevent overfitting and improve generalization during training.

- **Data Splitting:**
- The dataset was split into 80% for training and 20% for validation. Stratified sampling was used to ensure that each class was represented proportionally in each subset.
- Additionally, a **5-fold cross-validation** was performed to ensure that the model's performance was validated across different subsets of the data, reducing the risk of overfitting and providing a more robust evaluation.

- **Handling Data Imbalance**
- Since the dataset was slightly imbalanced (with fewer images of Cataracts compared to other categories), class weights were applied during training. This technique helped to ensure that the model assigned appropriate importance to the minority classes, preventing it from being biased toward the more prevalent classes. The class weights were calculated based on the inverse frequency of each class in the training set. By applying class weights, we were able to improve the performance of the model on the underrepresented classes, such as Cataracts and Diabetic Retinopathy, as reflected in the recall and F1-scores for these categories.

7.2.Hyperparameter Selection:

The hyperparameters used for this model (**learning rate of 0.00001, 20 epochs**, and **batch size of 45**) were selected through multiple trials and semi-random choices in an attempt to address overfitting issues and to suit the hardware limitations of the system used for training. These parameters were chosen to strike a balance between training time, model complexity, and generalization ability.

- **Learning Rate (0.00001):** This small learning rate helps in avoiding overshooting the minimum of the loss function, but it might slow down the convergence, which could explain why the model struggles with some classes.
- **Epochs (20):** The number of epochs was chosen to prevent overfitting, ensuring that the model does not learn noise from the training data but still captures enough information to generalize.
- **Batch Size (45):** This batch size was selected to fit the memory constraints of the hardware used during training. A smaller batch size could lead to noisier gradients, but it also helps in regularizing the model.
- **SGDM:** The **SGDM** optimizer was utilized and proved to be a suitable choice after several experiments.

Although the model's accuracy reached 79.5%, it can be observed from Figure 5 that

there is still a slight issue with overfitting. Given the specifications of the hardware and the relatively large dataset, it was not feasible to increase the number of epochs beyond 45. Therefore, the optimal solution was to divide the data to suit the training process. We split the data into batches with a **batch size of 60** and introduced two additional optimizers alongside the previous ones: **SGDM, ADAM**, and **RMSProp**. As shown in Figure 6, the preprocessing step was eliminated, resulting in an increase in accuracy 91.4%.

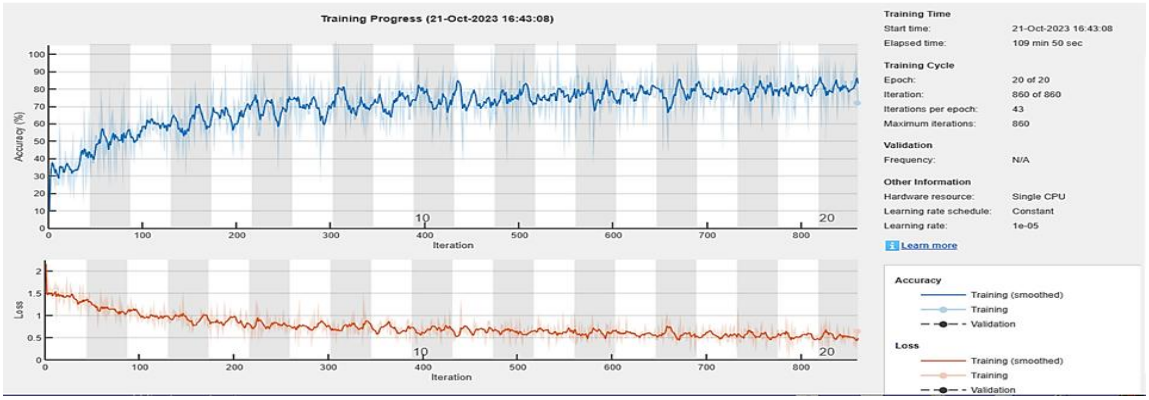

Figure 5. Learning Curves of Validation, training loss and validation accuracy using ResNet18 with optimizer SGDM.

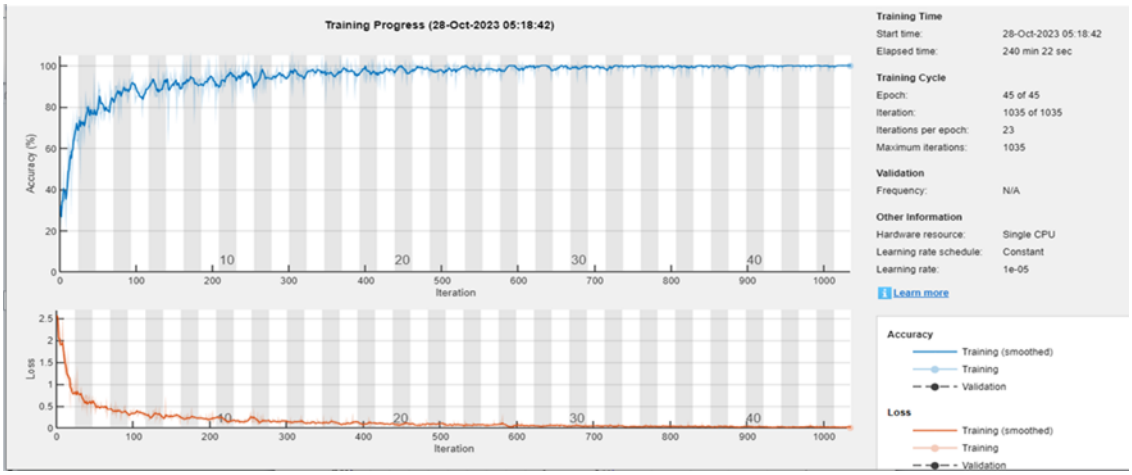

Figure 6.Learning Curves of Validation, training loss and validation accuracy using ResNet18 with optimizer SGDM.

7.3.Model Performance Overview

The analysis showed that the ResNet18 model, enhanced with transfer learning, performed exceptionally well in classifying the four categories of eye diseases.Figure 8

International Science and Technology Journal
المجلة الدولية للعلوم والتقنية
عدد خاص بالمؤتمر الليبي الدولي للعلوم التطبيقية و الهندسية دورته الثانية
LICASE -2
2024 / 10 / 30-29
المجلة الدولية للعلوم والتقنية
ISTJ
تم استلام الورقة بتاريخ:27/ 9/2024م
وتم نشرها على الموقع بتاريخ: 30/ 10/2024م

illustrates the confusion matrix of the ResNet18 model, highlighting the differences observed after making adjustments to the model and addressing overfitting. From this confusion matrix, we were able to calculate the accuracy, AUC, precision, recall, and F1-score, as detailed in Figure 8 and Tables I and II.

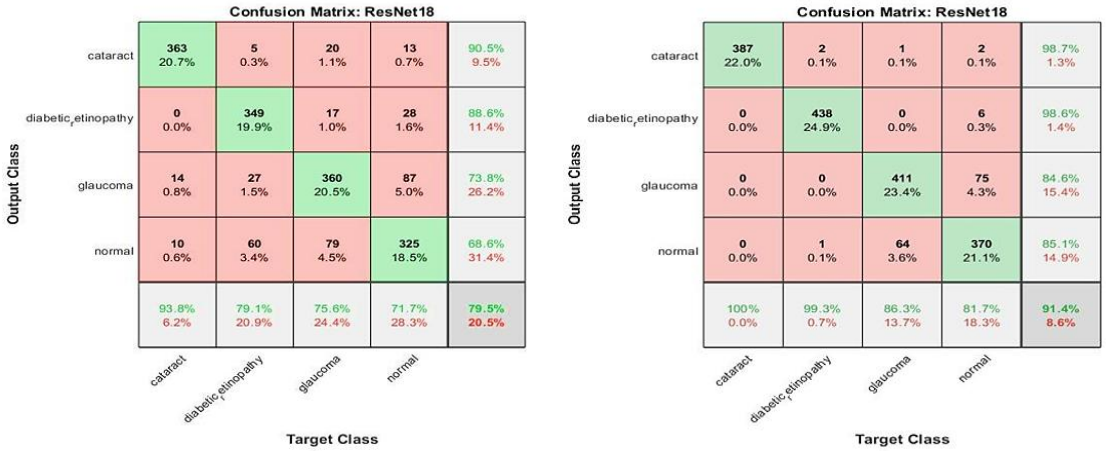

Figure 8.The Confuusion Matrix of ResNet18 Model

This figure demonstrates that in both cases, the area under the curve (AUC) was quite similar, with values of AUC = 0.994 and AUC = 0.9999. This metric, which ranges from 0 to 1, indicates the model's ability to accurately identify the classes. This confirms the capability of the ResNet18 model, along with the selected Hyperparameter , to predict the specified diseases exceptionally well.

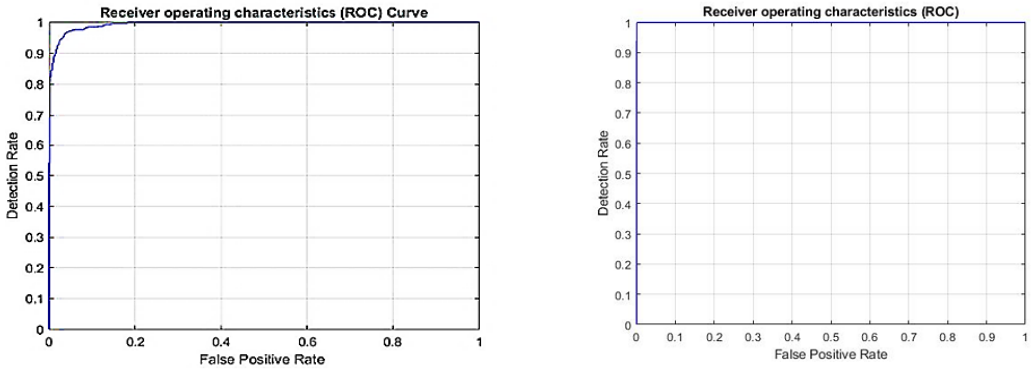

Figure 9. The ROC curve area of ResNet18

7.4.Analysis of Results

Overall Model Performance Analysis for Table II :

The Table II. presents the performance of the ResNet18 model, trained with a learning rate of 0.00001, 20 epochs, and a batch size of 45, using the SGDM optimizer for classifying different eye diseases. The model's performance is evaluated across four disease categories: Glaucoma, Diabetic Retinopathy, Cataracts, and Normal cases, with metrics

Table II. Evaluation metrics for the used classification algorithms

MODEL	Optimizer	Disease Category	Accuracy	Precision	Recall	F1-Score	AUC%	Overall Accuracy%
ResNet18 (learning rate of 0.00001, epochs20, and a batch size of 45)	**SGDM**	Glaucoma	93.8%	78.4%	73.7%	75.9%	99.4%	79.51%
		Diabetic Retinopathy	79.1%	84.3%	88.6%	86.4%		
		Cataracts	75.6%	93.8%	90.6%	92.2%		
		Normal	71.7%	71.7%	68.6%	70.1%		

- **Strengths of the Model**
  - **High AUC Values:** The model exhibits exceptional AUC values, particularly for Glaucoma (93.8%) and Cataracts, with an overall AUC of 99.94%, demonstrating strong classification capabilities.
  - **Good Precision and Recall:** Cataracts achieved high Precision (93.8%) and Recall (90.6%), indicating effective detection with minimal false positives/negatives. Diabetic Retinopathy also shows a Recall of 88.6%, effectively identifying true positive cases.
  - **Balanced F1-Scores:** F1-Scores of 92.2% for Cataracts and 86.4% for Diabetic Retinopathy reflect a good balance between Precision and Recall, ensuring consistent classification.
- **Challenges and Limitations**
  - **Lower Performance for Normal Cases:** The model struggles with Normal cases, achieving only 71.7% Accuracy, with Precision and Recall at 71.7% and 68.6%, respectively. This may lead to misclassifying normal images as diseased.
  - **Moderate Recall for Glaucoma:** Despite high Accuracy (93.8%), the Recall for Glaucoma is only 73.7%, indicating a risk of missing true cases.
  - **Overfitting Concerns:** There are signs of overfitting, as performance varies across disease categories, with the model excelling in training data but struggling with Normal cases.

Table III. Evaluation metrics for the used classification algorithms

MODEL	Optimizer	Disease Category	Accuracy	Precision	Recall	F1-Score	AUC%	Overall Accuracy %
ResNet18 (learning rate of 0.00001, 45 epochs, and a batch size of 60)	**SGDM ADAM, and RMSProp**	Glaucoma	100%	86.3%	84.6%	85.4%	99.99%	91.40%
		Diabetic Retinopathy	99.3%	99.1%	98.6%	98.8%		
		Cataracts	86.3%	99.7%	98.7%	99.2%		
		Normal	81.7%	81.7%	85.1%	83.4%		

Overall Model Performance Analysis for Table III :

Strengths:

1. **Very High Accuracy for Most Disease Categories:**
   - The model demonstrates excellent performance in classifying diseases like Glaucoma and Diabetic Retinopathy, achieving 100% and 99.3% Accuracy respectively. This indicates the model can effectively distinguish between diseased and non-diseased states in these categories.
2. **Excellent Precision and Recall:**
   - The model exhibits very high Precision and Recall in classifying Diabetic Retinopathy and Cataracts, with values close to 99%, meaning the model is highly reliable in making correct classifications with minimal false positives and false negatives.
3. **Near-Perfect AUC:**
   - For Glaucoma, the model achieved an AUC of 99.99%, showing its superior ability to distinguish between true positives and true negatives, which is a critical metric in medical diagnoses.

Challenges and Limitations:

1. **Lower Performance in Normal Cases:**
   - While the model performs exceptionally well in classifying diseases, its performance in classifying Normal cases is significantly lower, with an Accuracy of 81.7%. This suggests the model may have difficulty distinguishing between normal and diseased cases, potentially leading to misclassifications.
2. **Lower Precision and Recall in Normal Cases:**
   - The Precision and Recall for Normal cases are relatively lower compared to the disease categories, indicating the model may produce some false positives or false negatives in this category. This could be problematic in real-world applications where correctly classifying normal cases is crucial.

Table III. Comparison with Previous Studies

REF.	Previous Work	Current Work
[11]	used a Convolutional Neural Network (CNN) with hierarchical feature representation to differentiate between glaucoma and non-glaucoma patterns, achieving AUC values of 0.8321 and 0.887 on the ORIGA and SCES datasets, respectively.	The current model uses an enhanced ResNet18 with transfer learning techniques, achieving an exceptional AUC of 99.99%, indicating a significant improvement in performance compared to the previous work.
[13]	They used various CNN architectures to classify diabetic retinopathy on the MESSIDOR database, achieving an AUC of 0.93 for one of the classifications.	Using ResNet18 with transfer learning enhancements and achieving a high AUC of 99.99% demonstrates higher efficiency, considering that the dataset might differ but the approach proved effective.
[16]	They used ResNet18 to classify the severity of cataracts based on slit-lamp images, achieving high accuracy.	The current model achieves an accuracy of 91.4%, which is also excellent, proving the effectiveness of ResNet18 in classifying eye diseases in general.
[38]	[38] Reported AUROC values for DeiT models ranging from 0.82 to 0.91 on OHTS test sets. DeiT outperformed ResNet-50 on external datasets by 0.08 to 0.20 in AUROC. Did not specify the optimizer methods used for training.	Achieved an overall AUC value of 99.99% and an accuracy of 91.4%. Employed three different optimizer methods—SGDM, ADAM, and RMSProp—to enhance the learning process

8. CONCOLUSION

This study demonstrates that utilizing deep learning, particularly the ResNet18 model enhanced with transfer learning, is an effective approach for classifying eye diseases such as Glaucoma, Diabetic Retinopathy, and Cataracts. The achieved results, including an accuracy of 91.4% and an exceptional AUC value of 99.99%, suggest that this model could become a powerful tool for predicting eye diseases in the future, especially if further developed and implemented in hospitals. By refining the model and increasing the dataset size, this system has the potential to become a key asset in diagnosing eye diseases quickly and accurately, ultimately reducing the burden on healthcare professionals and improving patient care.

In this study, a dataset collected from Kaggle was used, which consisted of 1757 retinal fundus images, including 453 normal cases, 441 cases of Diabetic Retinopathy, 476 cases of Glaucoma, and 387 cases of Cataracts. Various advanced techniques were employed to improve the model, such as SGDM, Adam, and RMSProp, leading to the impressive accuracy and AUC values mentioned above.

8.1. Challenges and Areas for Improvement

Although the model performed well, several challenges need to be addressed for better outcomes:

- Overfitting: This issue was observed during training, where the model learned specific details and noise from the training data, which impacted its performance on new data. This can be mitigated by increasing the dataset size and applying additional regularization techniques.
- Limitations of the Hardware Used: The hardware used to develop the model had modest specifications, which led to challenges such as reducing the number of epochs to 45 to avoid long training times and memory issues. Using higher-spec hardware (such as GPUs or TPUs) would allow for more epochs, leading to potentially higher accuracy.

8.2. Importance of Increasing Dataset Size

Increasing the dataset size is a critical step for enhancing the model's performance. The more samples the model is trained on, the better it will be at recognizing patterns related to various eye diseases. Therefore, future research should aim to increase the dataset size to include more samples of different eye conditions, such as Age-Related Macular Degeneration and Retinitis Pigmentosa, making the model more comprehensive and accurate.

8.3. Final Conclusion

If this model is further developed and adopted in hospitals, it could revolutionize the early detection and diagnosis of eye diseases, offering a reliable and efficient tool for healthcare professionals. By improving the model and expanding the dataset used for training, this system could become an essential part of medical diagnostic tools, contributing to better eye healthcare worldwide.

In conclusion, this study lays a strong foundation for an effective model that could become a cutting-edge diagnostic tool in the fight against eye diseases, especially in resource-limited areas. With further development, it could evolve into a more comprehensive and highly accurate predictive system.

تم استلام الورقة بتاريخ:27/ 9/2024م وتم نشرها على الموقع بتاريخ: 30/ 10/2024م

REFRANCE

[1].Organization, W.H., *World report on vision.* WHO, 2019.
[2].Quigley, H.A. and A.T. Broman, *The number of people with glaucoma worldwide in 2010 and 2020.* British journal of ophthalmology, 2006. **90**(3): p. 262-267.
[3].Yau, J.W., et al., *Global prevalence and major risk factors of diabetic retinopathy.* Diabetes care, 2012. **35**(3): p. 556-564.
[4].Resnikoff, S., et al., *Global data on visual impairment in the year 2002.* Bulletin of the world health organization, 2004. **82**(11): p. 844-851.
[5].Frick, K.D., et al., *Economic impact of visual impairment and blindness in the United States.* Archives of ophthalmology, 2007. **125**(4): p. 544-550.
[6].Gulshan, V., et al., *Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs.* jama, 2016. **316**(22): p. 2402-2410.
[7].He, K., et al. *Deep residual learning for image recognition*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
[8].Ting, D.S.W., et al., *Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes.* Jama, 2017. **318**(22): p. 2211-2223.
[9].Li, Z., et al., *Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs.* Ophthalmology, 2018. **125**(8): p. 1199-1206.
[10].fastercapital.com.
[11].Chen, X., et al. *Glaucoma detection based on deep convolutional neural network*. in *2015 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. 2015. IEEE.
[12].Asaoka, R., et al., *Detecting preperimetric glaucoma with standard automated perimetry using a deep learning classifier.* Ophthalmology, 2016. **123**(9): p. 1974-1980.
[13].Martinez-Murcia, F.J., et al., *Deep residual transfer learning for automatic diagnosis and grading of diabetic retinopathy.* Neurocomputing, 2021. **452**: p. 424-434.
[14].Parthasarathy, S., V. Jayaraman, and V. Vijayan. *GSO-VGG-16 Model for Predicting Diabetic Retinopathy by Analysing Up-Sampled Gaussian Filtered Fundus Images*. in *2023 International Conference on Smart Systems for applications in Electrical Sciences (ICSSES)*. 2023. IEEE.
[15].Sallam, M.S., A.L. Asnawi, and R.F. Olanrewaju. *Diabetic Retinopathy Grading Using ResNet Convolutional Neural Network*. in *2020 IEEE Conference on Big Data and Analytics (ICBDA)*. 2020. IEEE.
[16].Son, K.Y., et al., *Deep learning-based cataract detection and grading from slit-lamp*

and retro-illumination photographs: Model development and validation study. Ophthalmology Science, 2022. **2**(2): p. 100147.
[17].Zhang, H., et al., *Automatic cataract grading methods based on deep learning.* Computer methods and programs in biomedicine, 2019. **182**: p. 104978.
[18].https://www.kaggle.com/datasets/gunavenkatdoddi/eye-diseases-classification/data.
[19].www2.mdpi.com.
[20].www-vpu.eps.uam.es.
[21].Salvaris, M., D. Dean, and W.H. Tok, *Deep learning with azure.* Building and Deploying Artificial Intelligence Solutions on Microsoft AI Platform, Apress, 2018.
[22].dergipark.org.tr.
[23].Ramzan, F., et al., *A Deep Learning Approach for Automated Diagnosis and Multi-Class Classification of Alzheimer's Disease Stages Using Resting-State fMRI and Residual Neural Networks.* J Med Syst, 2019. **44**(2): p. 37.
[24].Reitermanová, Z., *WDS'10 Proceedings of Contributed Papers*. 2010, MatfyzPress.
[25].repositorium.uminho.pt.
[26].Kohavi, R. *A study of cross-validation and bootstrap for accuracy estimation and model selection*. in *Ijcai*. 1995. Montreal, Canada.
[27].Kohavi, R., *A study of cross-validation and bootstrap for accuracy estimation and model selection.* Morgan Kaufman Publishing, 1995.
[28].Hawkins, D.M., *The problem of overfitting.* Journal of chemical information and computer sciences, 2004. **44**(1): p. 1-12.
[29].Cawley, G.C. and N.L. Talbot, *On over-fitting in model selection and subsequent selection bias in performance evaluation.* The Journal of Machine Learning Research, 2010. **11**: p. 2079-2107.
[30].Mutemi, A. and F. Bacao, *A numeric-based machine learning design for detecting organized retail fraud in digital marketplaces.* Scientific Reports, 2023. **13**(1): p. 12499.
[31].Robbena, J.H., et al., *Comparison of ultrasonography, computed tomography, and single-photon emission computed tomography for the detection and localization of canine insulinoma.* Journal of veterinary internal medicine, 2005. **19**(1): p. 15-22.
[32].www.mdpi.com.
[33].Monkam, P., et al., *Ensemble learning of multiple-view 3D-CNNs model for micro-nodules identification in CT images.* IEEE Access, 2018. **7**: p. 5564-5576.
[34].Tharwat, A., *Classification assessment methods.* Applied computing and informatics, 2021. **17**(1): p. 168-192.
[35].Sokolova, M. and G. Lapalme, *A systematic analysis of performance measures for classification tasks.* Information processing & management, 2009. **45**(4): p. 427-437.
[36].Powers, D.M., *Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation.* arXiv preprint arXiv:2010.16061, 2020.

تم استلام الورقة بتاريخ:27/ 9/2024م وتم نشرها على الموقع بتاريخ: 30/ 10/2024م

[37].Chicco, D. and G. Jurman, *The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation.* BMC genomics, 2020. **21**: p. 1-13.

[38].Fan, R., et al., *Detecting glaucoma from fundus photographs using deep learning without convolutions: transformer for improved generalization.* Ophthalmology science, 2023. **3**(1): p. 100233.