Intelligent Methods in Engineering Sciences 3(1): 037-045, 2024

IMIENS

INTELLIGENT METHODS
IN ENGINEERING SCIENCES

International
Open Access

March, 2024

https://www.imiens.org

e-ISSN 2979-9236

Research Article

https://doi.org/10.58190/imiens.2024.85

Evaluation of Machine Learning and Deep Learning Approaches for Automatic Detection of Eye Diseases

Ibrahim KAYA [a], ***Ilkay CINAR*** [b,*]

[a] *Department of Electric and Electronic Engineering, Selcuk University, Konya, TÜRKİYE*
[b] *Department of Computer Engineering, Selcuk University, Konya, TÜRKİYE*

ARTICLE INFO

Article history:
Received 20 January 2024
Accepted 28 March 2024

Keywords:
Classification,
Deep features,
Deep learning,
Eye diseases,
Machine learning

ABSTRACT

There are many ocular diseases present in the world. These diseases may arise from factors such as genetic predisposition, environmental influences, and aging. In recent years, advancements in technology have facilitated the detection of ocular pathologies through machine learning techniques. Machine learning models can serve as decision support mechanisms in diagnostic scenarios. In this study, the aim is to detect ocular diseases using machine learning and deep learning techniques. To enhance the results obtained from classification with 4,217 images in the study, 705 images were added to the glaucoma class and 370 images were added to the Diabetic Retinopathy class. The supplemented dataset with additional images comprises a total of four classes. One class represents the control group and is labeled as "normal." The remaining three classes represent disease categories: Diabetic Retinopathy, Cataract, and Glaucoma. To extract deep features from the images, a pre-trained InceptionV3 model was utilized, resulting in 2048 features extracted. These extracted features were then classified using Neural Network (NN), Logistic Regression (LR), k-Nearest Neighbors (k-NN), and Random Forest (RF) machine learning models. Before the dataset supplemented with additional images, the classification accuracies of the machine learning models were as NN 89.2%, LR 87.3%, k-NN 81.2%, and Random Forest 76.9%. Upon examining the classification accuracies after dataset supplemented with additional images, the following improvements were observed: NN 90.9% with a 1.7% increase, LR 90.2% with a 2.9% increase, k-NN 84.6% with a 3.4% increase, and Random Forest 82% with a 5.1% increase. Performance evaluation was conducted using recall, precision, and F-1 score metrics. Additionally, the learning performance of the machine learning models was assessed through Receiver Operating Characteristic (ROC) curves and Area Under Curve (AUC) values.

1. INTRODUCTION

A delicate organ of our body which converts incoming light into nerve impulses is eye. These generated optic nerve signals are transmitted to the brain, where the process of vision is executed [1]. Numerous ocular diseases exist worldwide, stemming from various causes such as genetic predisposition, development due to any other disease, environmental factors, and aging. Conditions such as Diabetic Retinopathy (DR) can occur due to an increase in blood sugar levels, increasing age may lead to the development of Cataracts, and Glaucoma can arise from imbalances in intraocular pressure regulation. In recent years, advancements in various methods and technological techniques have enabled the analysis of images of diseased eyes, providing insights into the respective diseases. Following the analysis, structures were developed using machine learning techniques to enable the identification of potential small diagnostic points in the images examined by doctors, which might otherwise go unnoticed. The developed analyses serve as decision support mechanisms in diagnostic scenarios [2]. Disease detection algorithms play a significant role in early diagnosis of illnesses. The aim is to establish a disease detection framework using machine learning and deep learning methods to diagnose eye diseases early. Initially, the classification accuracies of a dataset containing 4,217 images were examined, and subsequently, supplementation of the dataset with additional images was employed to enhance the existing classification accuracies.

Literature on the detection of eye diseases using deep learning and machine learning methods is provided below:

Pan et al., reported focusing on two different diseases in their study: Macular Degeneration and Mosaic Fundus Disease. They indicated that they also ensured control by representing the Normal Fundus class. They reported

*** Corresponding Author:** ilkay.cinar@selcuk.edu.tr

classification accuracies of 93.81% and 91.76% using ResNet-50 and InceptionV3 architectures, respectively [3].

Bitto et al., focused on two different diseases in their study: Conjunctivitis and Cataract. They ensured control by representing the Normal class. They reported a classification accuracy of 97.08% using InceptionV3 [4].

Hameed et al., conducted a study on the classification of eye diseases using an 8-class dataset comprising a total of 590 samples. They reported a classification accuracy of 89.83% using backpropagation with a parabolic function based on a linear cyclic learning rate [5].

Smaida et al., obtained an accuracy of 76.42% using a dataset containing 1200 images in their study. After utilizing the Deep Convolutional Generative Adversarial Network (DCGAN) method to augment the image data, they reported an increased accuracy of 83.74% [6].

Ahmad et al., conducted a study using data consisting of 360 labeled images to classify diseases on the outer part of the eye. They divided this dataset into four classes based on the affected part of the eye. They applied a hierarchical classification technique and reported a classification accuracy of 75.71% [7].

Wang et al., classified Macular Degeneration (MD) disease using a dual-stream CNN architecture. They utilized 2,379 images collected from the Department of Ophthalmology at Peking Union Medical College Hospital and reported a classification accuracy of 86% [8].

Hirota et al., used CNN models for glaucoma detection in 260 retinal images. Following classification, they reported an accuracy of 56% [9].

Sayres et al., conducted their study in two categories: eyes with DR (Diabetic Retinopathy) disease and eyes without DR disease. They utilized the EYEPACS dataset, which consists of 5 classes and a total of 1796 DR images. Using InceptionV4, they achieved a classification accuracy of 88.4% [10].

Serener et al., utilized GoogleNet and ResNet to work on four different stages of Glaucoma. They conducted their study using a dataset consisting of 1544 images and reported a classification accuracy of 86% [11].

Islam et al., proposed a CNN model to classify eight ocular diseases by applying preprocessing to the images, including contrast-limiting histogram equalization. As a result of these processes, they reported a classification accuracy of 85% on the ODIR-19 dataset [12].

Lam et al., proposed the detection of Diabetic Retinopathy (DR) using GoogleNet and AlexNet. In their study, they obtained data with 5 classes from "kaggle.com". After applying data augmentation techniques to this dataset, they achieved an accuracy of 75.50% [13].

Wang et al., conducted a study on the classification of DR stages using InceptionV3. They performed their study with a dataset comprising 166 images and 5 classes. They achieved an accuracy of 63.23% using the InceptionV3 model [14].

Chen et al., applied InceptionV3 and data augmentation techniques for the detection of Diabetic Retinopathy (DR) using data obtained from "kaggle.com". After applying data augmentation, an accuracy of 80% achieved [15].

A literature review revealed that no studies covered all four categories of diabetic retinopathy (DR), glaucoma, cataracts, and normal eyes. Focusing on these specified classes, we employed InceptionV3 deep learning model, to extract features. Subsequently, classification was performed using four distinct machine learning models for pre- and post- addition of images to the dataset used in the study. Comparative analysis of the results was conducted.

The study is structured as follows: The second section encompasses the flowchart of the study, detailed information about the dataset utilized, explanations regarding the employed methodologies, and performance analysis techniques. The third section presents the experimental outcomes of the study. Finally, the fourth section elucidates the detailed results obtained from the study.

2. Material and Methods

The classification of diabetic retinopathy (DR), cataracts, glaucoma, and normal eyes were aimed in this study. To achieve this, the InceptionV3 model is retrained using transfer learning for images in the dataset before and after the dataset supplemented with additional images. The trained InceptionV3 model was utilized to extract 2048 image features. These features, obtained with the contribution of InceptionV3, were then employed in conjunction with machine learning models including logistic regression (LR), neural networks (NN), k-nearest neighbors (k-NN), and random forest algorithms for classification. During the study, cross-validation was employed to measure performance. Furthermore, precision and F1 score metrics were used alongside recall for analyzing and recording the performance of the models at the conclusion of the study. Throughout the study, ROC curves and AUC curves were employed to compare the learning performance of the LR, NN, k-NN, and Random Forest models. The flowchart of the study is provided in Figure 1.

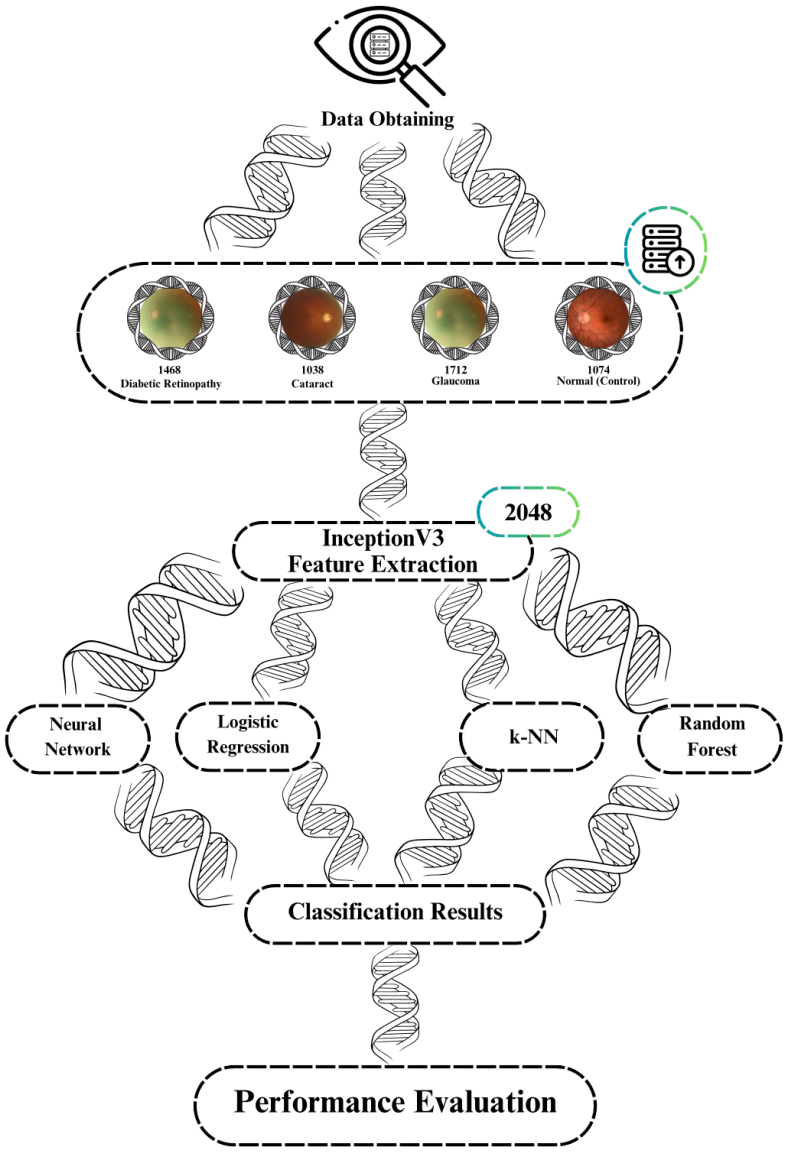

Figure 1. Classification Process Flow Diagram

2.1. Dataset Description

In this study, various eye diseases including diabetic retinopathy (DR), glaucoma, cataracts, and normal conditions were divided into four classes. Three different datasets were amalgamated for the study, all obtained from the "kaggle.com" platform [16-18]. Following the consolidation of datasets, the dataset comprises 1468 images for the DR class, 1712 images for the glaucoma class, 1038 images for the cataract class, and 1074 images representing the normal class. Table 1 provides the number of images for each class and their respective increments.

Table 1. Before and After Adding Additional Images to The Dataset: Data Counts

Before Additional Images		After Additional Images	
Classes	Number of total Images	Classes	Number of total Images
Normal Eye	1074	Normal Eye	1074
Glaucoma	1007	Glaucoma	1712
Cataracts	1038	Cataracts	1038
DR	1098	DR	1468

2.2. Machine Learning Models

The logistic regression (LR), neural network (NN), k-nearest neighbors (k-NN), and random forest (RF) models, frequently employed in the literature for the classification of eye diseases, have been applied. Detailed information regarding these models provided in the subsequent

subsections.

2.2.1. Logistic Regression (LR)

One of the most popular machine learning models, logistic regression (LR) is a widely used statistical model. Its primary aim is to focus on two different variables, where one is the dependent variable, and the other is the independent variable. LR aims to discern any relationship between these two variables. Unlike other models, LR does not require variables to follow a normal distribution [19, 20]. Since probabilities are estimated, this model has a limit between 0 and 1. This limitation arises from LR making predictions based on probabilities [21].

2.2.2. Neural Network (NN)

Neural networks (NN), an artificial intelligence model, algorithms that prioritize recognizing patterns in data similar to the human brain [22]. This algorithm relies on three main layers: input, hidden, and output. These layers consist of activation functions, weight functions, and interacting neurons [23]. The training of the model utilizes the backpropagation method [24].

2.2.3. k-Nearest Neighbors (k-NN)

The k-nearest neighbors (k-NN) model is a popular machine learning model. Instead of learning from training data, k-NN performs classification by memorizing it. It conducts the classification process by searching for the nearest neighbors of the data and categorizing them accordingly. During the execution of the algorithm, a value is computed for k, representing the number of neighbors. Each data point arriving for testing is calculated, and it is assigned to the appropriate class [25, 26].

2.2.4. Random Forest (RF)

Random Forest (RF) is a classification model consisting of multiple Decision Trees (DT). Each DT provides a classification process for the inputs to make a new classification. Subsequently, RF evaluates these classifications and selects the estimate with the highest accuracy [27]. RF has the capability to handle numerous variables in a dataset and is also a successful model in predicting missing data. Its major drawback lies in its lack of reproducibility. Additionally, interpreting the final model and subsequent results can be challenging due to its inclusion of multiple independent decision trees [28, 29].

2.3. Convolutional Neural Network (CNN)

CNN, which has many variations in the literature, is a deep learning architecture containing multiple layers. Activation maps in the layers are associated with input images. While the initial layers typically represent low-level features, subsequent layers often contain intermediate and high-level features. It has been observed multiple times that CNNs are used both as feature extractors and classifiers in classification structures [30-32].

2.4. InceptionV3

InceptionV3 is a comprehensive deep learning model comprising a total of 48 layers. The input layer of this model has image dimensions of 299 x 299. The architecture of InceptionV3 consists of both symmetric and asymmetric structures incorporating convolutional, pooling, dropout, and fully connected layers. Trained for the ImageNet competition in 2014, InceptionV3 is a sophisticated deep learning model developed for the purpose of object recognition [33, 34].

Within the additional images added to the dataset, a total of 5292 images were employed, yielding 2048 features extracted post-training via InceptionV3. Figure 2 [35] illustrates the architecture and processing pipeline of the InceptionV3 model utilized for feature extraction from images.

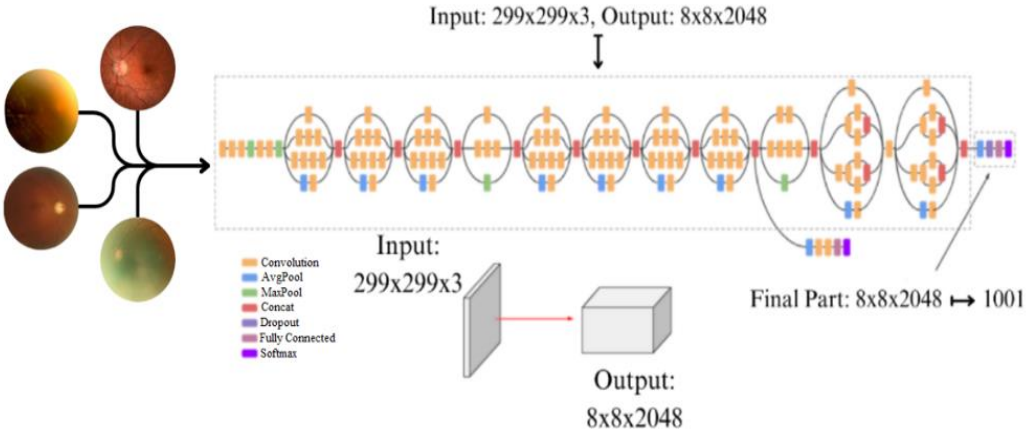

Figure 2. InceptionV3 Architecture

2.5. Confusion Matrix and Performance Metrics

The confusion matrix is employed for performance analysis of a generated or pre-existing model's classification predictions. This performance assessment is depicted by the confusion matrix, a matrix structure containing information about the classification predictions made by the model on test data post-training and related to the main classes [36]. For example, Table 2 below presents

a confusion matrix and its corresponding values. The confusion matrix consists of four different parameters: True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN). In the positive section, true positives indicate correct classifications, while true negatives indicate correct classifications for that class. Additionally, instances of positive classes classified as false negatives are referred to as false negatives, whereas instances of negative classes classified as false positives are referred to as false positives [37].

Table 2. Two-Classed Confusion Matrix

ACTUAL \ PREDICTED	0	1
1	TN	FP
0	FN	TP

In binary classification performance analyses, criteria such as Accuracy, F1-Score, Precision, Recall, and Specificity are employed, with calculation formulas provided in Table 3. The Receiver Operating Characteristic (ROC) curve is commonly employed to evaluate the performance of models post-training. When assessing the model's performance, the ROC curve approaches the top-left corner. Calculating the area under the ROC curve allows for the analysis of the model's predictive power. The area under the ROC curve is determined using the AUC, which ranges between 0 and 1 [38]. Table 3 illustrates the performance evaluation metrics for both binary and multiclass structures, along with the calculation formulas associated with these metrics [39].

Table 3. Performance Evaluation Metrics of Two-Class and Multi-Class Structures

Metrics	Two-Class Formula	Multi-Class Formula
Accuracy	$\frac{tp + tn}{tp + tn + fp + fn}$	$\frac{\sum_{i=1}^{1} \frac{tp_i + tn_i}{tp_i + fn_i + fp_i + tn_i}}{1}$
F-1 Score	$\frac{2tp}{2tp + fp + fn}$	$\frac{2 * \frac{\sum_{i=1}^{1} \frac{tp_i}{tp_i + fp_i}}{1} * \frac{\sum_{i=1}^{1} \frac{tp_i}{tp_i + fn_i}}{1}}{\frac{\sum_{i=1}^{1} \frac{tp_i}{tp_i + fp_i}}{1} + \frac{\sum_{i=1}^{1} \frac{tp_i}{tp_i + fn_i}}{1}}$
Precision	$\frac{tp}{tp + fp}$	$\frac{\sum_{i=1}^{1} \frac{tp_i}{tp_i + fp_i}}{1}$
Recall	$\frac{tp}{tp + fn}$	$\frac{\sum_{i=1}^{1} \frac{tp_i}{tp_i + fn_i}}{1}$

2.6. Cross Validation

Cross-validation with the argument K is a commonly used statistical method to evaluate the overall performance of models [40]. It divides the dataset into K equally sized folds, where one-fold is used as the test data and the rest as training data [41]. This splitting process is repeated for the number entered into the K argument. After repetition, the results are averaged [42], thus providing a measure of the average model performance [43]. Using the K argument to define the number 10, the dataset was divided into 10 equal parts, as illustrated in Figure 3, and the cross-validation method was applied.

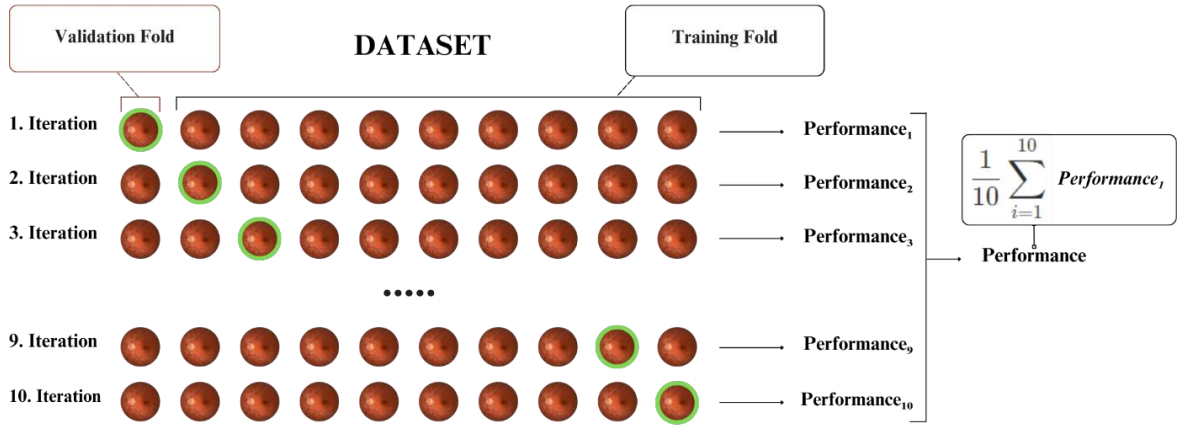

Figure 3. An Analysis of Cross Validation for K = 10

3. Experimental Results

Before supplemented the dataset with additional images, 4217 images were used, and post-training with the deep learning model InceptionV3, 2048 features were obtained. These features were then classified using machine learning models such as NN, LR, k-NN, and RF. Figure 4 illustrates the confusion matrix of the classification performance conducted with the NN model using the 2048 features obtained from InceptionV3, a deep learning model, prior to data image addition.

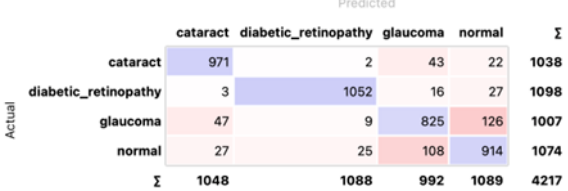

Figure 4. Confusion Matrix of NN Model Before Image Addition

The confusion matrix of the data predicted by the NN model on a class-by-class basis before image addition is depicted in Figure 4. Within this matrix, 27 images belonging to the Normal class were misclassified as cataracts. Additionally, 47 images that should have been classified as belonging to the Glaucoma class were mistakenly classified as cataracts. Furthermore, 3 images that should have belonged to the DR class were incorrectly predicted as cataracts.

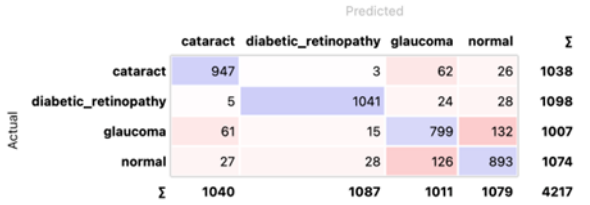

Figure 5. Confusion Matrix of LR Model Before Image Addition

The confusion matrix of the data predicted by the LR model on a class-by-class basis before image addition is depicted in Figure 5. Within this matrix, 27 images belonging to the Normal class were misclassified as cataracts. Moreover, 61 images that should have been classified as belonging to the Glaucoma class were mistakenly classified as cataracts. Additionally, 5 images that should have belonged to the DR class were incorrectly predicted as cataracts.

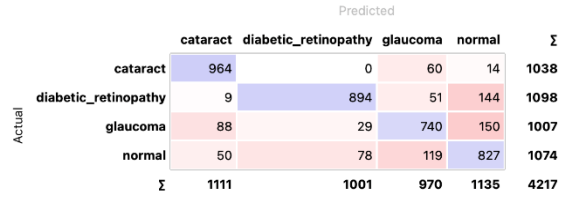

Figure 6. Confusion Matrix of k-NN Model Before Image Addition

The confusion matrix of the data predicted by the k-NN model on a class-by-class basis before image addition is depicted in Figure 6. Within this matrix, 50 images belonging to the Normal class were misclassified as cataracts. Additionally, 88 images that should have been classified as belonging to the Glaucoma class were mistakenly classified as cataracts. Furthermore, 9 images that should have belonged to the DR class were incorrectly predicted as cataracts.

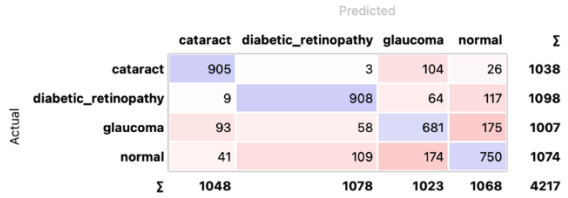

Figure 7. Confusion Matrix of RF Model Before Image Addition

The confusion matrix of the data predicted by the Random Forest model on a class-by-class basis before image addition is depicted in Figure 7. Within this matrix, 41 images belonging to the Normal class were misclassified as cataracts. Moreover, 93 images that should have been classified as belonging to the Glaucoma class were mistakenly classified as cataracts. Additionally, 9 images that should have belonged to the DR class were incorrectly predicted as cataracts.

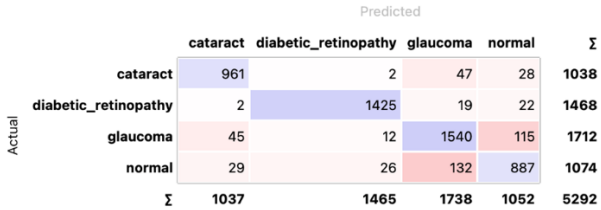

Figure 8. Confusion Matrix of NN Model After Image Addition

The confusion matrix of the data predicted by the NN model on a class-by-class basis after image addition is depicted in Figure 8. Within this matrix, 31 images belonging to the Normal class were misclassified as cataracts. Additionally, 54 images that should have been classified as belonging to the Glaucoma class were mistakenly classified as cataracts. Furthermore, 4 images that should have belonged to the DR class were incorrectly predicted as cataracts.

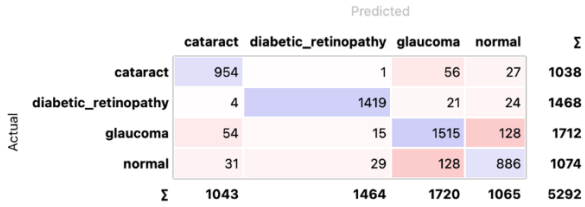

Figure 9. Confusion Matrix of LR Model After Image Addition

The confusion matrix of the data predicted by the LR model on a class-by-class basis after image addition is depicted in Figure 9. Within this matrix, 4 images that should have belonged to the DR class were incorrectly predicted as cataracts. Additionally, 54 images that should have been classified as belonging to the Glaucoma class were mistakenly classified as cataracts. Furthermore, 31 images belonging to the Normal class were misclassified as cataracts.

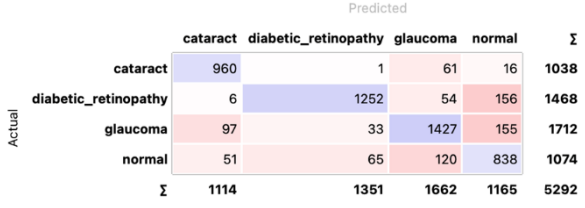

Figure 10. Confusion Matrix of k-NN Model After Image Addition

The confusion matrix of the data predicted by the k-NN model on a class-by-class basis after image addition is depicted in Figure 10. Within this matrix, 6 images that should have belonged to the DR class were incorrectly predicted as cataracts. Additionally, 97 images that should have been classified as belonging to the Glaucoma class were mistakenly classified as cataracts. Furthermore, 51 images belonging to the Normal class were misclassified as cataracts.

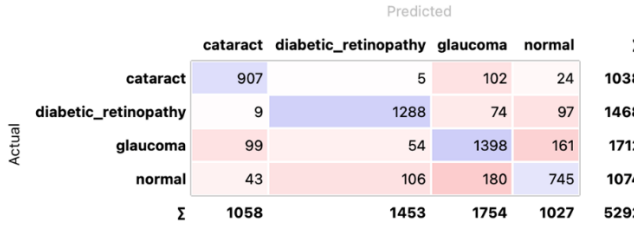

Figure 11. Confusion Matrix of RF Model After Image Addition

The confusion matrix of the data predicted by the Random Forest model on a class-by-class basis after image addition is depicted in Figure 11. Within this matrix, 9 images that should have belonged to the DR class were incorrectly predicted as cataracts. Additionally, 99 images that should have been classified as belonging to the Glaucoma class were mistakenly classified as cataracts. Furthermore, 43 images belonging to the Normal class were misclassified as cataracts.

Table 4. Performance Metrics Results for Machine Learning Models Before Image Addition

	Accuracy	F-1 Score	Precision	Recall	AUC
NN	%89.2	%89.2	%89.2	%89.2	%98.2
LR	%87.3	%87.3	%87.3	%87.3	%97.8
k-NN	%81.2	%81.2	%81.4	%81.2	%95.1
RF	%76.9	%76.9	%77	%76.9	%92.8

As seen in Table 4, before image addition, NN achieved a classification accuracy of 89.2%, LR achieved 87.3%, k-NN achieved 81.2%, and RF achieved 76.9%. While analyzing the classification accuracies, it is evident that the NN model achieved a higher classification accuracy compared to the other models.

Table 5. Performance Metrics Results for Machine Learning Models After Image Addition

	Accuracy	F-1 Score	Precision	Recall	AUC
NN	%90.9	%90.9	%90.9	%90.9	%98.7
LR	%90.2	%90.2	%90.2	%90.2	%98.5
k-NN	%84.6	%84.7	%85	%84.6	%96.5
RF	%82	%81.9	%81.9	%82	%95.4

As shown in Table 5, after image addition, NN achieved a classification accuracy of 90.9%, LR achieved 90.2%, k-NN achieved 84.6%, and RF achieved 82%. Upon examining the classification accuracies, it is observed that the NN model has a higher classification accuracy compared to the others.

Since the NN and LR machine learning models yielded the highest results in the study, the focus will be on these two machine learning models, NN and LR. ROC curves were used to analyze the classification performances and are shown in Figure 12. The NN model is represented in pink, the LR model in green, the k-NN model in orange, and the Random Forest model in purple.

Upon reviewing Figure 12, it can be discerned that in all ROC curves, the NN model exhibits higher sensitivity compared to the other models. Upon closer inspection, while LR and NN models have very similar ROC curves in the Normal, Cataract, and Diabetic Retinopathy classes, it can be observed that the NN model has significantly higher sensitivity in the Glaucoma class.

4. Conclusion

Within the research deep features were obtained using the CNN-based InceptionV3 model to detect various eye diseases, and four different machine learning models were employed for the classification of these features. The dataset consisted of pre-labelled 4217 images and comprised four classes: DR, Cataract, Glaucoma, and Normal, with the Normal class representing the control group.

In the initial stage, classification was conducted using the deep features obtained with the InceptionV3 model before resorting to image addition. According to the obtained results, the highest classification accuracy was achieved by the NN model at 89.2%, followed by the LR model at 87.3%, the k-NN model at 81.2%, and the Random Forest model at 76.9%. Subsequently, 705 additional images were added to the Glaucoma class and 370 to the DR class in the existing dataset. After the addition of images, the total number of images reached 5292. Following this, 2048 features per image were again extracted using the InceptionV3 model, and these features were classified using the machine learning models, NN, LR, k-NN, and Random Forest.

When comparing the performance metrics obtained before and after image addition to the dataset, it was observed that the supplementation of the dataset with additional images, the dataset improved the classification accuracy of the machine learning models. The NN model achieved the highest classification accuracy at 90.9% with a 1.7% increase, the LR model achieved a classification accuracy of 90.2% with a 2.9% increase, the k-NN model achieved a classification accuracy of 84.6% with a 3.4% increase. As for the Random Forest model, it achieved a classification accuracy of 82% with a 5.1% increase.

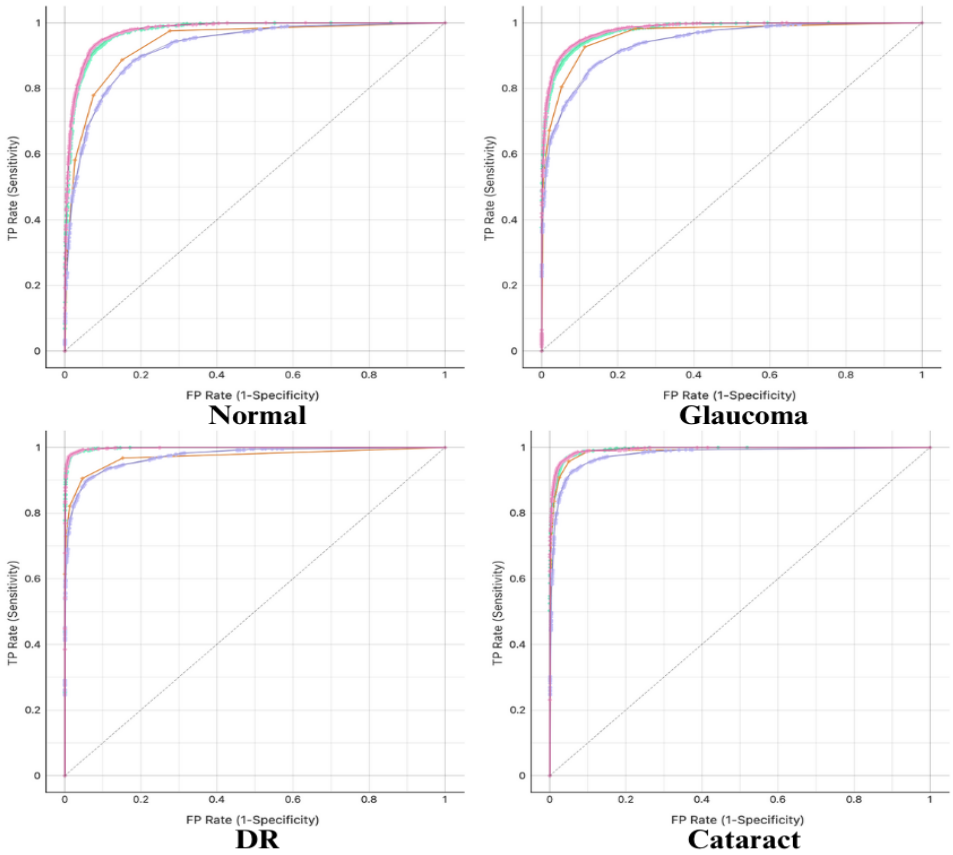

Figure 12. ROC Curves of NN, LR, k-NN and RF Machine Learning Models After Image Addition

For the overall performance evaluation of the study, recall, precision, and F-1 score methods were used. Additionally, the learning performance of the machine learning models was assessed by examining ROC curves and AUC values. For the image addition to the dataset, the NN model achieved F-1 score, precision, and recall values of 90.5%, while the LR model achieved 90.01% for all these metrics. As for the k-NN model, it attained an F-1 score of 84.7%, precision of 85%, and recall of 84.6%. Finally, the Random Forest model obtained an F-1 score, precision, and recall values of 81.9%.

In conclusion, it is anticipated that the methods implemented in the researches can be further developed and applied in the detection process of various eye diseases. These findings underscore the potential of artificial intelligence-supported techniques in the process of visual diagnosis. Furthermore, future research efforts could focus on increasing the dataset further and enhancing model developments.

Acknowledgments

This article was extracted from the project numbered 1919B012220496 carried out within the scope of TUBITAK 2209/A UNIVERSITY STUDENTS RESEARCH PROJECTS SUPPORT PROGRAM.

References

[1] Raza, A., et al., Classification of Eye Diseases and Detection of Cataract using Digital Fundus Imaging (DFI) and Inception-V4 Deep Learning Model, in 2021 International Conference on Frontiers of Information Technology (FIT). 2021. p. 137-142. DOI: 10.1109/FIT53504.2021.00034

[2] Umesh, L., M. Mrunalini, and S. Shinde, Review of image processing and machine learning techniques for eye disease detection and classification. International Research Journal of Engineering and Technology, 2016. 3(3): p. 547-551.

[3] Pan, Y., K. Zhao, and Z. Tan, Fundus image classification using Inception V3 and ResNet-50 for the early diagnostics of fundus diseases. Frontiers in Physiology, 2023. 14: p. 1126780. DOI: 10.3389/fphys.2023.1126780

[4] Bitto, A.K. and I. Mahmud, Multi categorical of common eye disease detect using convolutional neural network: a transfer learning approach. Bulletin of Electrical Engineering and Informatics, 2022. 11(4): p. 2378-2387. DOI: 10.11591/eei.v11i4.3834

[5] Hameed, S.R. and H.M. Ahmed, Eye diseases classification using back propagation with parabola learning rate. Al-Qadisiyah Journal Of Pure Science, 2021. 26(1): p. 1–9-1–9. DOI: 10.29350/qjps2021.26.1.1220

[6] Smaida, M., S. Yaroshchak, and Y. El Barg. DCGAN for Enhancing Eye Diseases Classification. in CMIS. 2021.

[7] Ahmad, H.M. and S.R. Hameed. Eye Diseases Classification Using Hierarchical MultiLabel Artificial Neural Network. IEEE. DOI: 10.1109/IT-ELA50150.2020.9253120

[8] Wang, W., et al., Learning Two-Stream CNN for Multi-Modal Age-Related Macular Degeneration Categorization. IEEE Journal of Biomedical and Health Informatics, 2022. 26(8): p. 4111-4122. DOI:10.1109/JBHI.2022.3171523

[9] Hirota, M., et al., Effect of color information on the diagnostic performance of glaucoma in deep learning using few fundus images. International Ophthalmology, 2020. 40(11): p. 3013-3022. DOI:10.1007/s10792-020-01485-3

[10] Sayres, R., et al., Using a deep learning algorithm and integrated gradients explanation to assist grading for diabetic retinopathy. Ophthalmology, 2019. 126(4): p. 552-564. DOI: 10.1016/j.ophtha.2018.11.016

[11] Serener, A. and S. Serte. Transfer learning for early and advanced glaucoma detection with convolutional neural networks. in 2019 Medical technologies congress (TIPTEKNO). 2019. IEEE. DOI: 10.1109/TIPTEKNO.2019.8894965

[12] Islam, M.T., et al. Source and camera independent ophthalmic disease recognition from fundus image using neural network. in 2019 IEEE International Conference on Signal Processing, Information, Communication & Systems (SPICSCON). 2019. IEEE. DOI: 10.1109/SPICSCON48833.2019.9065162

[13] Lam, C., et al., Automated detection of diabetic retinopathy using deep learning. AMIA summits on translational science proceedings, 2018. 2018: p. 147.

[14] Wang, X., et al. Diabetic retinopathy stage classification using convolutional neural networks. in 2018 IEEE International Conference on Information Reuse and Integration (IRI). 2018. IEEE. DOI: 10.1109/IRI.2018.00074

[15] Chen, H., et al. Detection of Diabetic Retinopathy using Deep Neural Network. in 2018 IEEE 23rd International Conference on Digital Signal Processing (DSP). 2018. DOI: 10.1109/ICDSP.2018.8631882

[16] Eye Diseases Classification Dataset. Available from: https://www.kaggle.com/datasets/gunavenkatdoddi/eye-diseases-classification.

[17] Glaucoma Detection Dataset. Available from: https://www.kaggle.com/datasets/sshikamaru/glaucoma-detection.

[18] Diabetic Retinopathy 224x224 Gaussian Filtered Dataset, Available from: https://www.kaggle.com/datasets/sovitrath/diabetic-retinopathy-224x224-gaussian-filtered.

[19] Cruyff, M.J., et al., A review of regression procedures for randomized response data, including univariate and multivariate logistic regression, the proportional odds model and item response model, and self-protective responses. Handbook of statistics, 2016. 34: p. 287-315. DOI: 10.1016/bs.host.2016.01.016

[20] Akhgarjand, C., et al., Does Ashwagandha supplementation have a beneficial effect on the management of anxiety and stress? A systematic review and meta-analysis of randomized controlled trials. Phytotherapy Research, 2022. 36(11): p. 4115-4124. DOI: 10.1002/ptr.7598

[21] Cinar, I. and M. Koklu, Classification of rice varieties using artificial intelligence methods. International Journal of Intelligent Systems and Applications in Engineering, 2019. 7(3): p. 188-194. DOI: 10.18201/ijisae.2019355381

[22] Wang, D., et al. Neural networks are more productive teachers than human raters: Active mixup for data-efficient knowledge distillation from a blackbox model. in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.

[23] Brownlee, J., What is the Difference Between a Batch and an Epoch in a Neural Network. Machine learning mastery, 2018. 20.

[24] Chicco, D., Siamese neural networks: An overview. Artificial neural networks, 2021: p. 73-94. DOI: 10.1007/978-1-0716-0826-5_3

[25] Cinar, I. and M. Koklu, Identification of Rice Varieties Using Machine Learning Algorithms. Journal of Agricultural Sciences, 2022: p. 9-9. DOI: 10.15832/ankutbd.862482

[26] Koklu, M. and K. Sabanci, Estimation of credit card customers payment status by using kNN and MLP. International Journal of Intelligent Systems and Applications in Engineering, 2016. 4(Special Issue-1): p. 249-251.

[27] Mao, W. and F. Wang, New advances in intelligence and security informatics. 2012: Academic Press.

[28] Dunham, K., Chapter 6-Phishing, SMishing, and Vishing. Mobile Malware Attacks and Defense, 2009: p. 125-196.

[29] Isik, M., Ozulku, B., Kursun, R., Taspinar, Y. S., Cinar, I., Yasin, E. T., & Koklu, M. (2024). Automated classification of hand-woven and machine-woven carpets based on morphological features using machine learning algorithms. The Journal of The Textile Institute, p. 1-10. DOI: 10.1080/00405000.2024.2309694

[30] Singh, D., et al., Classification and analysis of pistachio species with pre-trained deep learning models. Electronics, 2022. 11(7): p. 981. DOI: 10.3390/electronics11070981

[31] Koklu, M., et al., A CNN-SVM study based on selected deep features for grapevine leaves classification. Measurement, 2022. 188: p. 110425. DOI: 10.1016/j.measurement.2021.110425

[32] Koklu, M., I. Cinar, and Y.S. Taspinar, CNN-based bi-directional and directional long-short term memory network for determination of face mask. Biomedical signal processing and control, 2022. 71: p. 103216. DOI: 10.1016/j.bspc.2021.103216

[33] Demir, A., F. Yilmaz, and O. Kose. Early detection of skin cancer using deep learning architectures: resnet-101 and inception-v3. in 2019 medical technologies congress (TIPTEKNO). 2019. IEEE. DOI: 10.1109/TIPTEKNO47231.2019.8972045

[34] Sinha, R. and J. Clarke. When technology meets technology: Retrained 'Inception V3'classifier for NGS based pathogen detection. in 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2017. IEEE. DOI: 10.1109/BIBM.2017.8217942

[35] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826).

[36] Çataloluk, H., Gerçek tıbbi veriler üzerinde veri madenciliği yöntemlerini kullanarak hastalık teşhisi. 2012, Bilecik Üniversitesi, Fen Bilimleri Enstitüsü.

[37] Ruuska, S., et al., Evaluation of the confusion matrix method in the validation of an automated system for measuring feeding behaviour of cattle. Behavioural processes, 2018. 148: p. 56-62. DOI: 10.1016/j.beproc.2018.01.004

[38] Gencturk, B., Arsoy, S., Taspinar, Y. S., Cinar, I., Kursun, R., Yasin, E. T., and Koklu, M. (2024). Detection of hazelnut varieties and development of mobile application with CNN data fusion feature reduction-based models. European Food Research and Technology, 250(1), p. 97-110. DOI: 10.1007/s00217-023-04369-9

[39] Koklu, M. and I.A. Ozkan, Multiclass classification of dry beans using computer vision and machine learning techniques. Computers and Electronics in Agriculture, 2020. 174: p. 105507.

[40] Browne, M.W., Cross-validation methods. Journal of mathematical psychology, 2000. 44(1): p. 108-132. DOI: 10.1016/j.compag.2020.105507

[41] Bates, S., T. Hastie, and R. Tibshirani, Cross-validation: what does it estimate and how well does it do it? Journal of the American Statistical Association, 2023: p. 1-12. DOI: 10.1080/01621459.2023.2197686

[42] Xiong, Z., et al., Evaluating explorative prediction power of machine learning algorithms for materials discovery using k-fold forward cross-validation. Computational Materials Science, 2020. 171: p. 109203. DOI: 10.1016/j.commatsci.2019.109203

[43] Maleki, F., et al., Machine learning algorithm validation: from essentials to advanced applications and implications for regulatory certification and deployment. Neuroimaging Clinics, 2020. 30(4): p. 433-445. DOI: 10.1016/j.nic.2020.08.004