



OPEN Diagnostics of diabetic retinopathy based on fundus photos using machine learning methods with advanced feature engineering algorithms

Michał Gandor¹, Filip Pałka¹, Wojciech Książek¹, Paweł Pławiak^{1,2}✉, Aleksandra Siedlecka³, Tomasz Jeruzalski⁴, Magdalena Olender-Skorek³, Kamil Jonak^{5,6}, Robert Rejdak⁷, Katarzyna Nowomiejska⁷, Anna Matysik-Woźniak⁷, Miguel Teus^{8,9}, Carolina Arruabarrena Sanchez^{8,9} & Marek Sylwestrzak³

Diabetes is one of the main diseases posing a threat to healthcare systems. One of the complications of diabetes is diabetic retinopathy, which, if left untreated, can lead to serious consequences such as blindness. Early detection of this disease is critical to prevent disability and stop the process of vision loss. In our research, we aimed to develop and validate a machine learning model enabling early diagnosis of retinopathy disease. We were the first to conduct research using as many as eight public databases and one private database collected during the project implemented by the Ministry of Digital Affairs and the Ministry of Health of Poland. We analyzed 14,402 fundus photographs from patients, leveraging this large dataset to enhance the trustworthiness and validity of our findings. Such a large number of photos emphasizes the credibility and reliability of the results obtained. A significant innovation in our approach includes employing forty-six unique methods for feature selection and extraction, utilizing techniques such as CLAHE, B-CosFire, and Hough transform. We chose XgBoost and Random Forest algorithms for classification, with parameter tuning performed via the Optuna library. Our most successful model, employing the Random Forest algorithm combined with LBP and GLCM for feature extraction, reached a classification accuracy of 80.41%, F1-Score of 74.41%, and AUC of 0.80. The machine learning model we developed proved highly effective in the early detection of diabetic retinopathy. Further refinement is recommended to make this model a viable tool in clinical settings.

Keywords Diabetes, Diabetic retinopathy, Machine learning, Feature extraction, Feature selection, Artificial intelligence

Diabetes is one of the greatest and most important challenges faced by healthcare systems. It is one of the leading causes of death and disability worldwide. It affects people regardless of their country of residence, age or gender. It is a severe, chronic disease characterized by elevated blood glucose levels. According to estimates from the Global Burden of Diseases, Injuries and Risk Factors Study (GBD), diabetes is the eighth leading cause of death and disability in the world, with almost 460 million people in every country and age group living with the disease (as of 2019). This has been estimated to result in global healthcare spending of \$966 billion and is

¹Department of Computer Science, Faculty of Computer Science and Mathematics, Cracow University of Technology, Krakow, Poland. ²Institute of Theoretical and Applied Informatics, Polish Academy of Sciences, Gliwice, Poland.

³National Institute of Telecommunications, Szachowa 1, 04-894 Warszawa, Poland. ⁴Department of Management and Information Technology, Faculty of Economic Science, University of Warsaw, Krakowskie Przedmieście 26/28, 00-927 Warsaw, Poland. ⁵Department of Technical Informatics, Lublin University of Technology, 20-618 Lublin, Poland. ⁶Department of Clinical Neuropsychiatry, Medical University of Lublin, 20-059 Lublin, Poland. ⁷Chair and Department of General and Pediatric Ophthalmology, Medical University of Lublin, Chmielna 1, 20-079 Lublin, Poland. ⁸University of Alcalá, Madrid, Spain. ⁹Ophthalmology Department, University of Alcalá, Alcalá de Henares, Madrid, Spain. ✉email: pawel.plawiak@pk.edu.pl

projected to reach over \$1,054 billion by 2045. The results of the research that was conducted are alarming. In 2021, approximately 529 million people suffered from diabetes, which, out of an estimated population of 8 billion, shows that 6% of the population is sick. The dominant type of diabetes is type 2 diabetes, accounting for approximately 96% of all cases. According to research, approximately 50% of diabetes cases are caused by an increased BMI (body mass index). It can even be said with certainty that the leading causes of the disease are overweight, which turns into obesity, as well as a sedentary lifestyle and limited movement. It is expected that in 2050 the number of people with diabetes will be approximately 1.3 billion people. The development of diabetes can be significantly prevented, and in some situations, if detected early enough, it may be reversible. However, all evidence indicates that the incidence of diabetes is and will continue to increase worldwide and that this disease will represent one of the main challenges facing healthcare systems¹. Diabetes is also a major risk factor for ischemic heart disease and stroke, lower limb amputation, kidney failure, and diabetic retinopathy. Diabetic retinopathy is one of the most common complications of diabetes and is one of the leading causes of preventable blindness in the adult working population. The global incidence of blindness due to diabetic retinopathy increased by 14.9% to 18.5% between 1990 and 2020. With a rapidly aging global population, increasing lifespan of people with diabetes and lifestyle changes are expected to lead to an increased risk of diabetes, an increased burden of the disease and the need for eye care and treatment². Globally, the incidence of diabetic retinopathy among diabetic patients is estimated at 27.0%, which causes blindness in 0.4 million people worldwide - as confirmed by research conducted in 2015–2018³. It is believed that approximately 80% of people with type 2 diabetes develop retinopathy⁴.

With the increasing prevalence of diabetes and the high incidence of diabetic retinopathy among patients, combined with a limited number of ophthalmologists and rapid advancements in machine learning technologies, the use of artificial intelligence (AI) methods to enhance the diagnosis of this disease has become both achievable and essential. By 2050, the growing number of diabetes cases—and, consequently, diabetic retinopathy—underscores the critical role that AI, particularly machine learning, will play in addressing this healthcare challenge. Diabetic retinopathy is a severe condition that can lead to vision loss, profoundly impacting a patient's quality of life. Vision is a vital sense that enables individuals to perceive and interact with the world, and its loss is often associated with social isolation, difficulty securing employment, and mental health challenges such as depression. Currently, specialists' manual analysis of retinal images is prone to errors, especially in detecting early-stage disease, where timely intervention could prevent irreversible vision loss. Computer vision-based methods offer the potential to develop automated systems capable of identifying diabetic retinopathy in its early stages. Such systems not only improve patient outcomes by enabling earlier treatment but also alleviate the workload on healthcare professionals, who are often tasked with analyzing vast quantities of medical images. Furthermore, early detection reduces the financial burden associated with managing advanced stages of the disease, offering significant cost savings for healthcare systems⁵. For this purpose, classic machine learning methods are used, such as Random Forest or Support Vector Machine, along with advanced image processing methods, as well as deep neural networks - in particular convolutional networks, which have found a number of applications in medical diagnostics based on various types of medical data. Advanced machine learning methods have so far seen many applications in medicine, such as survival analysis of patients with hepatocellular carcinoma⁶, breast cancer prediction⁷, assessment of thyroid tumor malignancy⁸, classification of cardiac arrhythmias based on the ECG signal⁹, diagnosis of schizophrenia based on EEG¹⁰, or diagnosis of Covid-19 based on chest X-ray¹¹ or CT scan¹². Advanced machine learning models are also widely applied in other domains, such as biometrics¹³, which is particularly relevant today given its critical role in addressing cybersecurity challenges. Additionally, training physicians in the fundamentals of artificial intelligence and machine learning will be necessary to ensure they have at least a basic understanding of the tools they use and can trust the results.

The main goal of our research was to develop new machine learning models for early diagnosis of diabetic retinopathy based on fundus photographs. For this purpose, we used advanced image processing methods and classification algorithms. Additionally, our research utilized eight publicly available datasets along with our own data collected in Poland, comprising a total of 14,402 fundus images. This makes our study one of the largest of its kind reported in the literature. The substantial volume of data enhances the stability of our model and significantly improves its generalization capabilities. A key innovation of our study is the implementation of forty-six distinct scenarios for preprocess, feature selection, and extraction. Investigating such a broad range of options enabled us to develop highly effective classification algorithms. We anticipate that the models developed in this research will eventually find applications in clinical practice. Previous approaches presented in the literature have mainly focused on single methods of preprocessing, selection and feature extraction. The literature lacked a comprehensive experiment, which could indicate to researchers the most promising paths of fundus image processing for diabetic retinopathy diagnostics. Our research fills this gap by using a very broad experiment using a very large number of techniques such as GLCM, GLRM, PCA, LBP, as well as less known ones such as B-COSFIRE Filters. We tested not only these single methods, but also fusions of these approaches - designing advanced mechanisms for feature engineering. Initially, they could serve as digital assistants to support physicians in their diagnostic processes. In the future, these systems have the potential to evolve into fully autonomous diagnostic tools, and, considering the advancements in large language models (LLMs), may even interact directly with patients. Implementing such models will require robust hardware infrastructure, the development of software integrating the proposed algorithms, and a responsible approach to addressing cybersecurity concerns and ensuring the protection of personal data.

Related works

In recent years, there has been a growing interest among scientists in the use of machine learning methods in medical diagnostics. Many studies have been based on the use of fundus photography in the diagnosis of diabetic retinopathy. The growing interest may result from the increase in the incidence of diabetes and, consequently,

diabetic retinopathy¹⁴, as well as the increasing shortage of qualified medical personnel¹⁵. Seud et al.¹⁶ proposed the use of random forest in the diagnosis of diabetic retinopathy. Working on the basis of approximately 1,200 fundus photos, scientists extracted thirty-five features that served as input for the classification. The experiment was conducted using LOO validation. The proposed model achieved a classification accuracy of 74.1%. In¹⁷ Pratt et al. proposed the use of convolutional convolutional networks, which are the dominant model in image recognition tasks. They conducted their research on a publicly available set on Kaggle, and they achieved a high sensitivity of 95% and classification accuracy of 75% on a test set consisting of 5,000 samples. Acharya et al. in their article¹⁸ proposed using a support vector machine for disease diagnosis. Their research was based on photos of 331 patients. The obtained results are a sensitivity of 82% and a specificity of 86%. The work¹⁹ used various feature selection techniques such as LDA, PCA, and spatial invariant feature transform (SIFT) with the AlexNet deep neural network. The obtained classification accuracy of the model was 97.93%, 95.26%, and 94.40%, respectively. The research shows how key an element is to select features in designing effective machine learning models. Li et al.²⁰ used deep neural networks to diagnose diabetic retinopathy using the Inception-v3 network. They conducted their research on a dataset consisting of almost 20,000 images. Experiments were conducted using 10-fold cross-validation. The photos were resized to 299 by 299 pixels. Contrast-limited adaptive histogram equalization (CLAHE) was also used as a preprocessing method. The proposed model achieved an accuracy of 93.49% and a specificity of 93.45%. The Inception-ResNet-v2 network and methods related to transfer learning were used in the work²¹. The accuracy results were around 70–80% for the public Aptos and Messidor datasets. Nahiduzzama et al.²² used a parallel convolutional neural network in their research to perform feature extraction (120 features) and extreme machine learning in the classification task. The proposed models achieved an accuracy of 91.78% on the Kaggle DR 2015 dataset and 97.27% on the Aptos dataset. Many convolutional models were used in research by Kassani et al.²³ Xception, InceptionV3, MobileNet, ResNet50 and a new variant of Xception were used, which obtained the best results - accuracy of 83.09%, sensitivity of 88.24% and specificity of 87.00%. The models were trained on 2,657 images and tested on 343. The article²⁴ presents the use of the You Only Look Once (Yolo) V7 feature extraction algorithm together with the tailored quantum marine predator algorithm (QMPA) used in feature selection. The MobileNet V3 network was used as a classifier. CLAHE and Wiener filter techniques were used for preprocessing. The F1-score was approximately 93% on selected public datasets. The work²⁵ presents a combination of the concepts of deep features obtained from deep networks such as EfficientNet and DenseNet along with biologically inspired algorithms: Binary Bat Algorithm (BBA), Equilibrium Optimizer (EO), Gravity Search Algorithm (GSA), and Gray Wolf Optimizer (GWO) and classic machine learning methods such as SVM. The best of the designed models achieved a classification accuracy of 96.32%. The article²⁶ also uses several convolutional models to classify diabetic retinopathy. Resnet50, Inceptionv3, Xception, Dense121, and Dense169 were used. The authors decided to balance the dataset using data augmentation techniques. As part of this research, these networks were combined into an ensemble model that obtained an f1-score of 53.74% and an accuracy of 80.8%. Modi and Kumar²⁷ proposed a novel combination of deep forest and bat optimization algorithms. As part of the simulation, they compared the following models: KNN, ANN, SVM, VGG16, VGG19, InceptionV3 and Deep Forest. BA-DeepForest achieved a classification accuracy of 92.94%, a sensitivity of 95.67% and a specificity of 94.95%. To reduce the effect of model overfitting, the authors used 10-fold cross-validation. The work²⁸ proposes the use of several feature extraction algorithms: Maximum a Posteriori (MAP) algorithm, Principal Component Analysis (PCA), and Gray level co-occurrence matrix (GLCM) to build new classification models for diabetic retinopathy. The proposed model using a support vector machine achieved an accuracy of 77.3%, a sensitivity of 90.2% and a specificity of 74.1%. Machine learning algorithms have also been successfully applied to detect glaucoma. In the study²⁹, several deep learning models were proposed to identify this disease automatically. The best performance was achieved by a novel convolutional model developed by the authors. Its primary advantage was its relatively small number of layers, which significantly reduced training time. The model achieved a classification accuracy exceeding 80%, a result that can be considered highly satisfactory. In contrast, another study³⁰ proposed an alternative approach to early glaucoma detection, utilizing the Fast Fuzzy C-means clustering technique. This method enabled the design of a system with a remarkable prediction accuracy of 97.75%. These experiments were conducted using publicly available datasets containing fundus images. Additionally, the research presented in³¹ introduced an innovative hybrid approach by combining differential evolution with the K-Nearest Neighbors (KNN) algorithm for glaucoma detection. This hybrid model achieved a classification accuracy of 90%. Similarly, the study³² explored the application of biologically inspired algorithms, specifically the Grey Wolf Optimizer (GWO) and the Whale Optimization Algorithm (WOA). This approach facilitated the development of classification models with an impressive accuracy of 96.8%. These advanced models demonstrate significant potential for clinical application, offering reliable tools to assist ophthalmologists in the diagnosis and management of glaucoma.

One can notice the great interest of scientists in this topic. The analyzed works used both deep models and traditional classification techniques with advanced feature extraction methods. However, these models still require further research to support doctors and increase access to medical services in the future.

Materials and methods

Datasets

As part of our experiments, we used eight publicly available datasets and one new dataset obtained as part of a project aimed at screening patients with diabetes in Poland - for the purpose of early diagnosis. All of these datasets contain fundus images. Table 1 shows statistics for each dataset.

Employing a large number of datasets represents a significant shift from the traditional approach in this field, where researchers typically relied on just one or two datasets. This expanded data use allows for the development of models that are not only stronger but also more reliable. With more datasets, a broader information, base which enhances the model's ability to predict outcomes accurately. This approach significantly improves the

Name	Size
APTOS ³³	3662
Aria ³⁴	120
Drive ³⁵	40
E_optha ³⁶	442
Eye_diseases_classification_kaggle ³⁷	2172
HRF ³⁸	30
IDRiD ³⁹	516
ODIR ⁴⁰	4868
Topcon	4724
Total	14024

Table 1. Datasets used in the research.

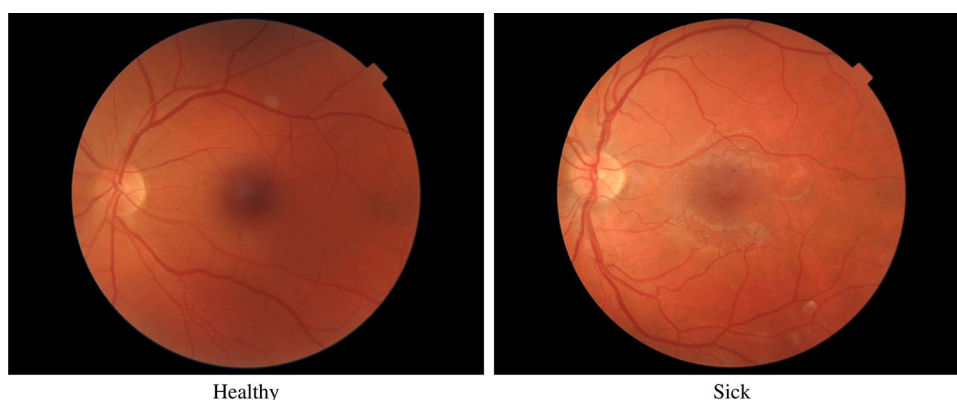


Fig. 1. Example of fundus photo for patients.

stability and accuracy of the models, offering a substantial improvement over previous methods which may have suffered from limited data diversity and potential biases. Modern machine learning models possess strong generalization capabilities. However, for these models to be suitable for clinical practice, they cannot be trained on a single dataset, acquired from a specific country or device. Such a limitation would make the model prone to overfitting, undermining its applicability in real-world clinical settings. Our research demonstrates that combining multiple datasets—despite originating from different sources—is feasible by appropriately converting and scaling the images. By leveraging a large, diverse dataset, the models we developed can more accurately reflect their real-world performance, ensuring they are capable of robust, reliable disease prediction in clinical environments.

Telediagnosics in ophthalmological examinations⁴¹ is a joint initiative of the Ministry of Health, the Minister of Digitization, the Medical University of Lublin and the Institute of Telecommunications - National Research Institute. One of the elements of the project is EyeBus, a specialized vehicle for telediagnosics equipped with medical devices enabling the acquisition of data by performing advanced eye tests in the field, as well as IT and telecommunications devices used to transmit the data obtained, in order to create an artificial intelligence algorithm and remote patient diagnoses. The project will accelerate diagnosis and provide access to tests for residents of areas with difficult access to advanced ophthalmological diagnostics. Pilot studies have been carried out in the Lubelskie Voivodeship, involving 5,000 participants. These studies took place from February to December 2023 across the whole region. Due to the particular characteristics of eye diseases, ophthalmologists evaluated patients ranging from 18 to 80 years old. Although this project was initially pilot in nature, plans are to expand it to cover the whole country in the future. This study was approved by the ethics committee of Medical University of Lublin (Commission resolution no.: KE-254/124/05/2022 of 26 May 2022). All methods were carried out in accordance with relevant guidelines and regulations. Informed consent was obtained from all subjects and/or their legal guardian(s).

Figure 1 shows an example photo for a sick and a healthy patient.

Data preprocessing and feature extraction

Appropriate data preprocessing is one of the essential tasks that must be performed to build effective machine learning models. A number of different approaches to preprocessing fundus images can be found in the literature. It is challenging to clearly determine which data transformation path is the best - so we decided to conduct a various experiments based on methods such as Clahe⁴², B-COSFIRE Filters⁴³ or Hough Transform⁴⁴. The CLAHE (Contrast Limited Adaptive Histogram Equalization) algorithm was employed to enhance the contrast

of fundus images. This technique enables improved visualization of retinal vessels and is among the most commonly used algorithms in the preprocessing of fundus images. Conversely, the B-CosFire filter algorithm is utilized for the extraction of retinal vessels—a challenging task due to the potential for vessels to be misidentified as noise, arising from imperfections in the processed images. In the final stage of this algorithm's application, binary thresholding is performed to generate a retinal vessel map⁴⁵.

For common classification algorithms such as Random Forest or XgBoost, we decided to perform four preprocessing scenarios:

- Variant 1 - called “pure” In this variant, the images were cropped, and resized, the background outside the eyeball circle was cut out, and the contrast was improved using the CLAHE method.
- Variant 2 - called “no_vessels” In this variant in addition to the previous one the veins were cut out based on the B-CosFire algorithm.
- Variant 3 - called “no_vessels_no_optic_disk” In this variant, the optical disc was additionally found and cut out using the Hough method.
- Variant 4 - called “no_optic_disk” In this variant in addition to the first variant, the optical disc was found and cut out using the Hough method.

For each of these four variants, global histogram equalization was also tested (variant name with “eq”). An average histogram of all images within each dataset was calculated, and all images were adjusted to the average histogram in each dataset respectively.

In the experiments conducted, a number of feature extraction methods were also used to select key features, reduce the size of the feature space and highlight important ones. The following algorithms were used:

- Principal Component Analysis (PCA)⁴⁶
- Local binary patterns (LBP)⁴⁷
- Gray-Level Co-Occurrence Matrix (GLCM)⁴⁸
- Gray-Level Run-Length Matrix (GLRLM)⁴⁹
- Histogram⁵⁰

GLCM and GLRLM feature extraction methods were conducted within smaller windows. Each image was divided into three windows for GLCM and twelve for GLRLM.

During this research, we carried out forty-six experiments involving diverse configurations of preprocessing, feature selection, and extraction. The goal was to develop classification models that achieved the highest possible accuracy. These varied experiments allowed exploring multiple approaches to refine our methods and enhance the performance of our models.

Table 2 illustrates the appearance of the photos after applying different preprocessing variants. It is evident that the methods employed vary significantly from each other, highlighting their value for inclusion in the research.

Experiments

The experiments utilized traditional classification techniques, including XgBoost and RandomForest. Random Forests are among the learning methods used in classification and regression problems, which function by creating many decision trees in the training phase, the result of which is the average results of all trees. They belong to the group of bagging-based classifiers⁵¹. The XgBoost (Extreme Gradient Boosting) algorithm was also used, achieving excellent results in many issues. It is a gradient boosting model belonging to the group of ensemble classifiers⁵². Both the XgBoost and Random Forest algorithms belong to the group of very advanced and effective machine learning models that are useful in many fields, including the classification of medical data. The random forest achieved an efficiency of as much as 91.87% in the problem of predicting the recurrence of pancreatic cancer⁵³, 80.43% classification accuracy in diagnosing glaucoma⁵⁴ or early diagnosis of depression⁵⁵. The XgBoost algorithm has demonstrated high efficacy in breast cancer prediction - in a group of 500 patients, this efficacy was as high as 97%⁵⁶. Equally promising results for this algorithm were obtained in the problem of stroke prediction - this time, XgBoost combined with PCA achieved a classification accuracy of 95%⁵⁷. The OPTUNA library was used to optimize the parameters of these two classifiers⁵⁸ and the metric used for optimization was f1-score. The data was split into training and test sets in a 0.7 to 0.3 ratio keeping the original distribution between classes. The training dataset was used during parameter tuning with Optuna with 10-split cross-validation. The final hyperparameters set received after tuning was used to train the final model using the same training dataset, where tests were performed using the test set not used before. The results discussed in this paper were specifically obtained from the test set, which was not involved in the optimization or training stages.

In order to build the most effective machine learning algorithm for diagnosing diabetic retinopathy, a number of experiments were carried out related to various preprocessing methods, feature selection and extraction, and various classification models. Classification accuracy and the f1-score measures were used as the basic metrics for model evaluation.

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

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN} \quad (2)$$

The following measures were also used helpfully:



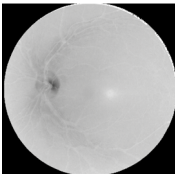
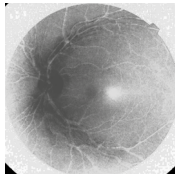
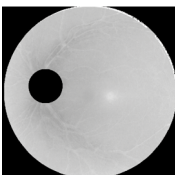
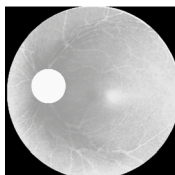
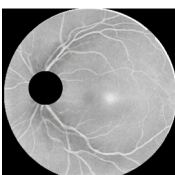
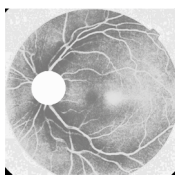
Preprocessing methods	
 (a) pure	 (b) pure_eq
 (c) no_vessels	 (d) no_vessels_eq
 (e) no_vessels_no_optic_disk	 (f) no_vessels_no_optic_disk_eq
 (g) no_optic_disk	 (h) no_optic_disk_eq

Table 2. Comparison of different preprocessing methods.

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$Sensitivity = Recall = \frac{TP}{TP + FN} \tag{5}$$

$$Specificity = \frac{TN}{FP + TN} \tag{6}$$

where:

- TP: is the number of True Positives
- TN: is the number of True Negatives
- FP: is the number of False Positives
- FN: is the number of False Negatives

Additionally, ROC curves were drawn for the best models and the AUC value was calculated
The source code of the designed models was written in Python. The following libraries were used:

- Scikit-learn⁵⁹
- XgBoost⁶⁰
- Optuna⁶¹
- Scikit-image⁶²

- PyFeats⁶³
- Scipy⁶⁴
- OpenCV⁶⁵

The calculations were conducted on a machine having following the specifications:

- Processor: Intel(R) Xeon(R) CPU E5-2670 v3 @ 2.30 Ghz 2.30 Ghz (two processors)
- RAM: 512 GB
- Operating system: Windows Server 2019 (64-bits)

Figure 2 shows a simplified diagram of our experiment.

The experiment can be divided into main stages such as data acquisition and their initial preprocessing, feature engineering, optimization and training of machine learning models and their evaluation.

Results

During the research, forty-six feature engineering scenarios were implemented. The XgBoost and Random Forest algorithms were employed as the primary classifiers. Hyperparameters of both classifiers were optimized using Optuna to find the best fitting values. Optimization ranges are presented in Table 3.

Table 4 shows the complete results achieved using Random Forest. The best result was achieved for the “pure” scenario using a combination of two feature extraction methods: Local binary patterns (LBP) and Gray-Level Co-Occurrence Matrix (GLCM). This model achieved a classification accuracy of 80.41% and an f1-score of 0.744. Table 5 shows the hyperparameters used in the final model.

Figure 3 shows the confusion matrix for this model and Figure 4 shows the ROC curve.

It can be seen in the confusion matrix that the model performs well in recognizing cases from both classes. The AUC area under the ROC curve is equal to 0.80.

Similar experiments were conducted using the XgBoost algorithm, which is recognized as one of the most effective and widely used classifiers in current machine learning research. The results for this model are displayed in Table 6.

The best results were achieved by the model in the “no_vessels_eq” preprocessing scenario using the Gray-Level Run-Length Matrix feature extraction method. The designed model had a classification accuracy of 0.816 and an f1-score of 0.729. The high performance achieved with the XGBoost algorithm resulted from several factors. First, a large dataset was collected, comprising both public data and data from the implemented project. Second, we employed forty-six preprocessing scenarios to identify the optimal configuration. Lastly, the use of the Optuna framework was crucial in selecting the appropriate parameters for the XGBoost algorithm. Given the complexity of XGBoost and the large number of parameters it requires, configuring it without advanced tools like Optuna would have been challenging. Complete hyperparameter values can be found in Table 7.

Figures 5 and 6 show the confusion matrix and ROC curve of this model.

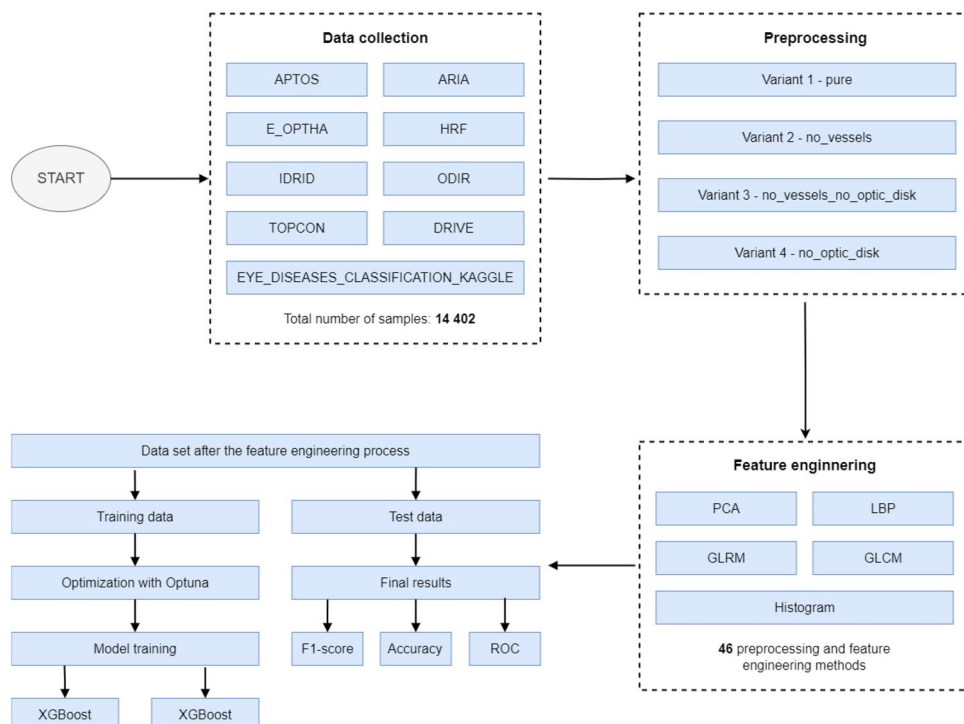


Fig. 2. Experimental scheme.

Classifier	Parameter	Values
XgBoost	tree_method	'exact', 'auto'
	booster	'gbtree', 'gblinear', "dart"
	lambda	1e-8 - 10
	alpha	1e-8 - 10
	subsample	0.2 - 10
	colsample_bytree	0.2 - 10
	max_depth	1 - 15
	min_child_weight	1 - 10
	eta	1e-8 - 1
	gamma	1e-8 - 1
	grow_policy	'depthwise', 'lossguide'
	sample_type	'uniform', 'weighted'
	rate_drop	1e-8 - 1
	skip_drop	1e-8 - 1
Random Forest	max_depth	1-50
	max_features	2-50
	min_samples_split	1-50
	n_estimators	10-1000
	bootstrap	True, False
	criterion	'gini', 'entropy'

Table 3. Hyperparameters ranges used during optimization.

Analyzing these results, it can be observed that those are similar to the results of the Random Forest algorithm. Similarly, the model correctly handles the classification of cases from both classes. The area under the ROC curve is 0.78.

Discussion

This article explored various strategies for constructing classification models to diagnose diabetic retinopathy using fundus images. Our research included well-known and commonly used classifiers such as XgBoost and Random Forest combined with sophisticated preprocessing methods such as CLAHE and B-CosFire filter and advanced feature selection and extraction techniques like PCA, LBP, GLCM, and GLRM. Our goal was to check the effectiveness of these methods in the early diagnosis of diabetic retinopathy. We conducted advanced experiments involving forty-six different preprocessing and feature engineering variants to develop highly effective diagnostic models. Many tested scenarios were designed to identify the most promising techniques for predicting diabetic retinopathy from fundus images and the ones that should be disregarded. For instance, the “pure and histogram” scenario using the Random Forest algorithm resulted in a classification model with an accuracy of only 55%, while the “pure + lbp_glcm” scenario yielded an accuracy of over 80%. This highlights the significant potential of certain algorithms, particularly those based on texture analysis, such as GLCM and GLRM, in addressing the problem at hand. Table 8 presents the top five algorithms, ranked by their F1-score.

The Random Forest models generally outperformed those using the XGBoost algorithm, although the performance differences were relatively small. The best configuration of the XGBoost algorithm achieved a classification accuracy of 0.816 and an F1-score of 0.729. In comparison, the Random Forest algorithm achieved a similar classification accuracy of 0.804, but slightly outperformed XGBoost with a higher F1-score of 0.744. The minor differences in the results highlight the similarity in the underlying principles of these models, as both are ensemble methods with a single decision tree serving as their foundational element. Several stages of the project contributed to these strong results, including acquiring a large dataset, initial preprocessing, feature selection and extraction, and optimizing model parameters using the Optuna framework.

Our research represents one of the first experiments to utilize such a diverse array of datasets, leading to reduced model overfitting and increased reliability. While some scientific studies report higher model performance, these results are often achieved by training on a single dataset, which tends to produce overfitted models with limited generalization capabilities. In contrast, our approach surpasses most existing efforts by employing a significantly broader range of preprocessing and feature extraction methods—forty-six in total. Unfortunately, machine learning models trained on a single dataset from a single population, often even on sets obtained from a single device for performing fundus examination, are susceptible to overfitting. Such models in the literature often achieve accuracy above 90%, which unfortunately does not translate into real performance. Often, models tested on other fundus sets (still in the problem of predicting diabetic retinopathy) are characterized by much lower efficiency. Such studies on many datasets have often not been shown before - due to the lack of availability of such data. In this work, we have shown the availability of a number of different datasets containing data from patients with fundus images in the problem of detecting diabetic retinopathy - so today conducting such extensive research is possible and we can look again at certain approaches to the classification of this disease and extend the research with new data - which will constitute an additional validation of these algorithms.

Index	f1	image_type	preprocessing_type	Recall	Precision	Accuracy
0	0.722	pure	lbp_histogram	0.701	0.744	0.798
1	0.744	pure	lbp_glcml	0.761	0.728	0.804
2	0.508	pure	histogram	0.613	0.433	0.555
3	0.726	pure	glcml	0.745	0.708	0.790
4	0.735	pure	glrlml	0.748	0.723	0.798
5	0.743	pure	lbp_glrlml	0.769	0.718	0.800
6	0.629	pure	pca	0.630	0.629	0.722
7	0.687	pure	lbp_pca	0.689	0.686	0.765
8	0.687	pure_eq	lbp_histogram	0.642	0.738	0.781
9	0.726	pure_eq	lbp_glcml	0.696	0.758	0.803
10	0.529	pure_eq	histogram	0.677	0.434	0.549
11	0.716	pure_eq	glcml	0.703	0.730	0.792
12	0.724	pure_eq	glrlml	0.709	0.739	0.797
13	0.729	pure_eq	lbp_glrlml	0.723	0.735	0.799
14	0.565	pure_eq	pca	0.556	0.574	0.679
15	0.670	pure_eq	lbp_pca	0.690	0.651	0.746
16	0.703	no_vessels	lbp_histogram	0.727	0.680	0.770
17	0.723	no_vessels	lbp_glcml	0.723	0.724	0.793
18	0.647	no_vessels	histogram	0.670	0.626	0.726
19	0.704	no_vessels	glcml	0.695	0.714	0.782
20	0.723	no_vessels	glrlml	0.739	0.708	0.788
21	0.734	no_vessels	lbp_glrlml	0.734	0.734	0.801
22	0.674	no_vessels	pca	0.664	0.685	0.760
23	0.665	no_vessels	lbp_pca	0.647	0.683	0.755
24	0.706	no_vessels_eq	lbp_histogram	0.743	0.673	0.768
25	0.731	no_vessels_eq	lbp_glcml	0.762	0.702	0.790
26	0.649	no_vessels_eq	histogram	0.685	0.617	0.723
27	0.711	no_vessels_eq	glcml	0.704	0.718	0.785
28	0.734	no_vessels_eq	glrlml	0.707	0.762	0.808
29	0.731	no_vessels_eq	lbp_glrlml	0.733	0.730	0.798
30	0.664	no_vessels_eq	pca	0.656	0.673	0.752
31	0.664	no_vessels_eq	lbp_pca	0.610	0.727	0.768
32	0.693	no_vessels_no_optic_disk	lbp_histogram	0.703	0.683	0.767
33	0.723	no_vessels_no_optic_disk	lbp_glcml	0.706	0.740	0.797
34	0.641	no_vessels_no_optic_disk	histogram	0.673	0.612	0.718
35	0.709	no_vessels_no_optic_disk	glcml	0.695	0.724	0.787
36	0.728	no_vessels_no_optic_disk	glrlml	0.759	0.700	0.788
37	0.737	no_vessels_no_optic_disk	lbp_glrlml	0.756	0.718	0.797
38	0.654	no_vessels_no_optic_disk	pca	0.635	0.675	0.749
39	0.644	no_vessels_no_optic_disk	lbp_pca	0.588	0.712	0.757
40	0.669	no_vessels_no_optic_disk_eq	lbp_histogram	0.648	0.691	0.760
41	0.705	no_vessels_no_optic_disk_eq	lbp_glcml	0.685	0.727	0.786
42	0.608	no_vessels_no_optic_disk_eq	histogram	0.592	0.624	0.714
43	0.701	no_vessels_no_optic_disk_eq	glcml	0.673	0.732	0.785
44	0.714	no_vessels_no_optic_disk_eq	glrlml	0.693	0.736	0.792
45	0.719	no_vessels_no_optic_disk_eq	lbp_glrlml	0.681	0.762	0.801
46	0.624	no_vessels_no_optic_disk_eq	pca	0.638	0.611	0.725
47	0.656	no_vessels_no_optic_disk_eq	lbp_pca	0.613	0.705	0.759

Table 4. Results obtained for Random Forest.

Additionally, we used eight publicly available datasets alongside one private dataset, each containing fundus images that were used to predict whether individuals were healthy or exhibited signs of diabetic retinopathy. As part of the experiment, these datasets were combined and then split into a 70% training set and a 30% test set. For the training set, we applied 10-fold cross-validation, which ensured robust evaluation and helped mitigate overfitting. This comprehensive approach not only reduced overfitting but also provided a realistic assessment of the classification capabilities of the designed algorithms.

Parameter	Value
max_depth	50
max_features	26
min_samples_split	22
min_samples_leaf	16
n_estimators	862
bootstrap	True
criterion	entropy

Table 5. Hyperparameters used in the final model of Random Forest.

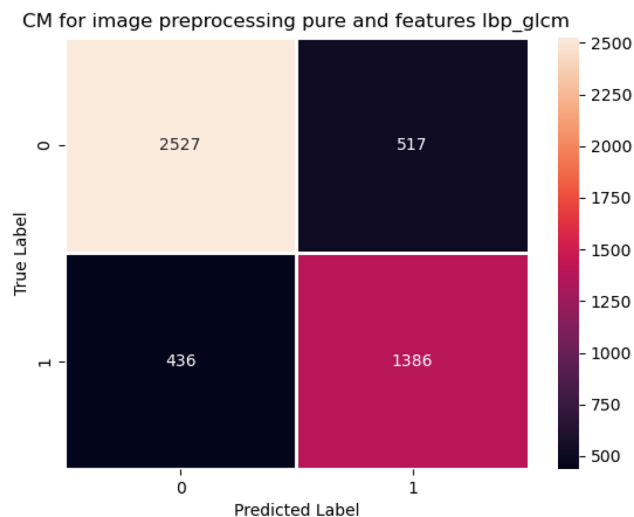


Fig. 3. Confusion matrix for pure and lbp_glc with Random Forest.

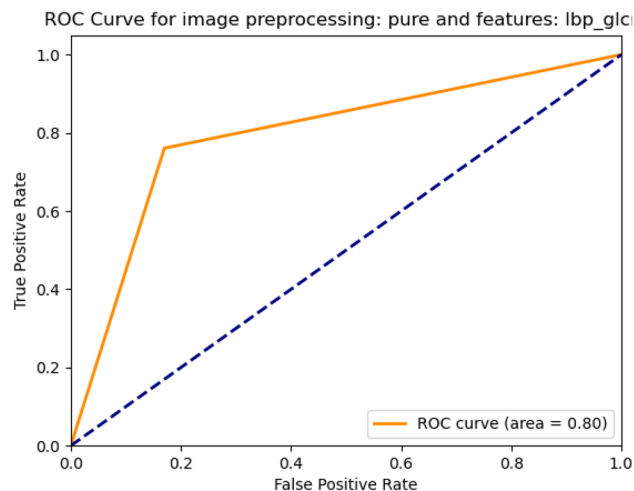


Fig. 4. ROC for pure and lbp_glc with Random Forest.

It should be emphasized that the literature includes studies employing techniques such as GLCM, GLRM, or PCA for the automatic diagnosis of diabetic retinopathy. For instance, in the work⁶⁶, the GLCM algorithm was used to extract features from fundus images. However, this research was conducted using a dataset containing only 280 images. In our study, the dataset consisted of 14,402 images. Similar studies are found in⁶⁷, where feature extraction methods such as LBP, LTP, DSIFT, and HOG were utilized. The authors worked with a dataset comprising 1,384 fundus images. In another work⁶⁸, models were built using features derived from algorithms such as histograms, GLCM, GLRM, and wavelet features, based on a dataset of 2,500 images. The study⁶⁹

Index	f1	image_type	preprocessing_type	Recall	Precision	Accuracy
0	0.718	pure	lbp_histogram	0.654	0.797	0.808
1	0.727	pure	lbp_glcml	0.650	0.826	0.818
2	0.390	pure	histogram	0.372	0.411	0.566
3	0.714	pure	glcml	0.662	0.775	0.801
4	0.722	pure	glrlml	0.658	0.801	0.811
5	0.714	pure	lbp_glrlml	0.631	0.821	0.811
6	0.623	pure	pca	0.556	0.709	0.748
7	0.663	pure	lbp_pca	0.593	0.753	0.775
8	0.690	pure_eq	lbp_histogram	0.599	0.813	0.798
9	0.722	pure_eq	lbp_glcml	0.645	0.821	0.814
10	0.421	pure_eq	histogram	0.422	0.420	0.566
11	0.693	pure_eq	glcml	0.612	0.799	0.797
12	0.698	pure_eq	glrlml	0.632	0.780	0.795
13	0.715	pure_eq	lbp_glrlml	0.627	0.831	0.813
14	0.555	pure_eq	pca	0.478	0.663	0.714
15	0.640	pure_eq	lbp_pca	0.560	0.745	0.764
16	0.690	no_vessels	lbp_histogram	0.623	0.773	0.790
17	0.720	no_vessels	lbp_glcml	0.658	0.795	0.808
18	0.613	no_vessels	histogram	0.550	0.692	0.740
19	0.684	no_vessels	glcml	0.607	0.783	0.790
20	0.708	no_vessels	glrlml	0.650	0.777	0.799
21	0.725	no_vessels	lbp_glrlml	0.664	0.797	0.811
22	0.650	no_vessels	pca	0.570	0.755	0.770
23	0.644	no_vessels	lbp_pca	0.552	0.771	0.771
24	0.697	no_vessels_eq	lbp_histogram	0.634	0.775	0.794
25	0.712	no_vessels_eq	lbp_glcml	0.651	0.785	0.803
26	0.622	no_vessels_eq	histogram	0.555	0.706	0.747
27	0.681	no_vessels_eq	glcml	0.613	0.767	0.785
28	0.729	no_vessels_eq	glrlml	0.660	0.813	0.816
29	0.721	no_vessels_eq	lbp_glrlml	0.644	0.819	0.813
30	0.640	no_vessels_eq	pca	0.559	0.749	0.765
31	0.641	no_vessels_eq	lbp_pca	0.550	0.767	0.769
32	0.678	no_vessels_no_optic_disk	lbp_histogram	0.616	0.755	0.781
33	0.716	no_vessels_no_optic_disk	lbp_glcml	0.650	0.798	0.807
34	0.619	no_vessels_no_optic_disk	histogram	0.557	0.696	0.743
35	0.696	no_vessels_no_optic_disk	glcml	0.617	0.798	0.798
36	0.714	no_vessels_no_optic_disk	glrlml	0.653	0.788	0.804
37	0.703	no_vessels_no_optic_disk	lbp_glrlml	0.622	0.809	0.804
38	0.625	no_vessels_no_optic_disk	pca	0.542	0.738	0.756
39	0.628	no_vessels_no_optic_disk	lbp_pca	0.531	0.769	0.765
40	0.672	no_vessels_no_optic_disk_eq	lbp_histogram	0.596	0.771	0.782
41	0.701	no_vessels_no_optic_disk_eq	lbp_glcml	0.625	0.797	0.800
42	0.570	no_vessels_no_optic_disk_eq	histogram	0.487	0.689	0.725
43	0.694	no_vessels_no_optic_disk_eq	glcml	0.634	0.767	0.791
44	0.713	no_vessels_no_optic_disk_eq	glrlml	0.632	0.816	0.809
45	0.706	no_vessels_no_optic_disk_eq	lbp_glrlml	0.627	0.807	0.804
46	0.600	no_vessels_no_optic_disk_eq	pca	0.516	0.716	0.754
47	0.633	no_vessels_no_optic_disk_eq	lbp_pca	0.543	0.759	0.764

Table 6. Results obtained for XgBoost.

combined deep learning with the PCA algorithm, conducting experiments on a dataset of 1,445 fundus images. Furthermore, previous scientific studies rarely explore the fusion of these methods. Instead, most works focus on comparative analyses⁷⁰, examining features obtained using the GLCM and GLRM methods.

Table 9 presents a comparison with selected works in the literature. One can observe a number of classification methods used in the literature, such as basic SVM type prediction algorithms, ensemble learning and convolutional networks. Moreover, most of the experiments were conducted on a single dataset - which

Parameter	value
tree_method	auto
booster	gbtree
lambda	1.092
alpha	4.31
subsample	0.72
colsample_bytree	0.84
max_depth	14
min_child_weight	3
eta	0.13
gamma	0.021
grow_policy	depthwise

Table 7. Hyperparameters used in the final model of XgBoost.

CM for image preprocessing no_vessels_eq and features glrlm

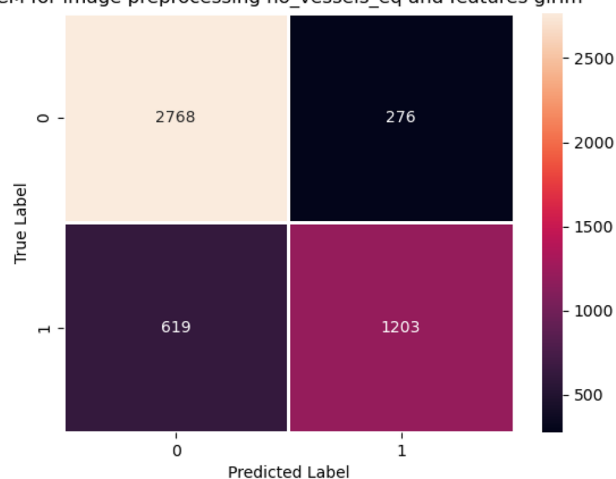


Fig. 5. Confusion matrix for pure and lbp_glcmm with XgBoost.

ROC Curve for image preprocessing: no_vessels_eq and features: glrlm

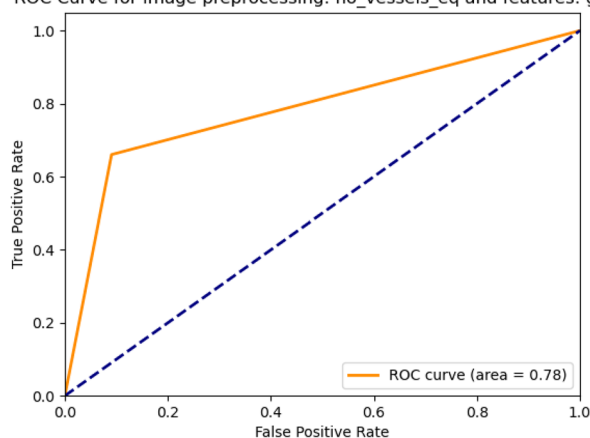


Fig. 6. ROC for pure and lbp_glcmm with XgBoost.

is what distinguishes our research - as many as 9 datasets used. The primary objective of our research was to conduct these experiments on a large dataset sourced from various origins to assess the actual effectiveness of these techniques objectively. This is crucial because small datasets can lead to overstated performance due to overfitting, rather than reflecting true model accuracy. Modern machine learning models must demonstrate

Classifier	f1-score	image_type	Preprocessing_type	Recall	Precision	Accuracy
Random Forest	0.744	pure	lbp_glcm	0.761	0.728	0.804
Random Forest	0.743	pure	lbp_glrlm	0.769	0.718	0.800
Random Forest	0.737	no_vessels_no_optic_disk	lbp_glrlm	0.756	0.718	0.797
Random Forest	0.735	pure	glrlm	0.748	0.723	0.798
Random Forest	0.734	no_vessels	lbp_glrlm	0.734	0.734	0.801

Table 8. Summary of the best results achieved.

Reference	Method	Number of datasets	Results
17	Convolutional neural networks	1	Acc: 75%
18	SVM	1	Acc: 85.9%
21	Inception-ResNet-v2	2	Acc: 72.33% and 82.18%
26	Ensemble learning + CNN	1	F1-score: 80.8%
23	Modified Xception Architecture	1	Acc: 83.09
This study	Random Forest + advanced feature engineering	9	Acc: 80.4

Table 9. Comparison with state-of-the-art algorithms.

high generalization capabilities. It is unacceptable for a model to perform well on one set of fundus images while failing completely on another set addressing the same problem, such as diabetic retinopathy. Moreover, our research introduces several novel methods, such as combining LBP with GLRM or GLCM, and innovative preprocessing approaches, including the removal of the eyeball or retinal vessels. These contributions underscore the necessity and value of our research. We hope the results and proposed models presented here will guide and inspire future research in this field.

Key innovations of our research include:

- An experiment containing many different preprocessing and feature engineering techniques
- Integration of seven publicly available datasets
- The use of effective classifiers such as XgBoost and Random Forest

However, our study also has limitations:

- Lack of use of biologically inspired methods in parameter optimization and feature selection
- Limited number of classification methods
- No experiment related to deep features

Future research on this field will focus on advancing optimization methods, especially the use of those inspired by biology, such as genetic algorithms or PSO. Biologically inspired metaheuristics have been effectively applied to a wide range of advanced optimization problems, including the selection of machine learning parameters and feature selection. In this context, techniques such as the genetic algorithm⁷¹, bee algorithm⁷², and whale algorithm⁷³ have been utilized. These approaches leverage the principles of wrapper-based feature selection, allowing the development of classification models that often outperform those constructed using traditional filter-based methods. Plans also include expanding the range of classifiers tested, including logistic regression and support vector machines, and incorporating deep convolutional methods for enhanced feature extraction and classification tasks. In particular, the authors plan to focus on pre-trained convolutional models, such as VGG16, VGG19, ResNet, ResNet-152, GoogleNet, and various variants of the EfficientNet architecture. In particular, the authors plan to focus on pre-trained modern models, such as Vision Transformers (ViT)⁷⁴, Swin Transformer⁷⁵, ConvNeXt⁷⁶ but also various variants of the EfficientNet architecture such as EfficientNetV2⁷⁷. We also plan further experiments related to XgBoost and Random Forest classifiers - in order to develop these methods through advanced parameter selection, as well as designing ensemble models using weighted voting or the concept of model contamination - where XgBoost and Random Forest algorithms will be used as key components of the ensemble model. The primary objective of further work on these models will be to enhance their accuracy, with a particular focus on minimizing type I errors. This approach aims to ensure that the models are reliable and suitable for deployment in clinical practice. Further research will also focus on analyzing the explainability of models based on extracted features.

Conclusions

Early diagnosis of diabetic retinopathy is necessary to prevent the progression of the disease and prevent the development of blindness. Artificial intelligence algorithms can support these activities through early diagnosis of the disease. Both simple classification methods and deep methods allow the introduction of machine learning algorithms in clinical practice. However, it is necessary to conduct further research to build even more effective models. Additionally, further research should consider the explainability of models to make them more

understandable to doctors and gain their trust. According to the authors, in the coming years, models that support the diagnosis of diabetic retinopathy will become an integral part of the healthcare system, leading to faster diagnoses, shorter waiting times for doctors, and an improved quality of life for patients.

Data availability

Research data for eight publicly available datasets are presented in Section 3.1. Datasets. Data obtained as part of the Telediagnosics in ophthalmological examinations project cannot be made available for privacy reasons. They are available upon reasonable request from the corresponding author (P.P.).

Received: 27 May 2024; Accepted: 11 June 2025

Published online: 03 October 2025

References

- Ong, K. L. et al. Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: A systematic analysis for the Global Burden of Disease Study 2021. *Lancet* **402**, 203–234 (2023).
- Teo, Z. L. et al. Global prevalence of diabetic retinopathy and projection of burden through 2045: Systematic review and meta-analysis. *Ophthalmology* **128**, 1580–1591 (2021).
- Thomas, R., Halim, S., Gurudas, S., Sivaprasad, S. & Owens, D. Idf diabetes atlas: A review of studies utilising retinal photography on the global prevalence of diabetes related retinopathy between 2015 and 2018. *Diabetes Res. Clin. Pract.* **157**, 107840 (2019).
- Zegeye, A. F., Temachu, Y. Z. & Mekonnen, C. K. Prevalence and factors associated with diabetes retinopathy among type 2 diabetic patients at Northwest Amhara Comprehensive Specialized Hospitals, Northwest Ethiopia 2021. *BMC Ophthalmol.* **23**, 9 (2023).
- Senapati, A., Tripathy, H. K., Sharma, V. & Gandomi, A. H. Artificial intelligence for diabetic retinopathy detection: A systematic review. *Inform. Med. Unlocked* **45**, 101445. <https://doi.org/10.1016/j.imu.2024.101445> (2024).
- Ksiażek, W., Gandor, M. & Plawiak, P. Comparison of various approaches to combine logistic regression with genetic algorithms in survival prediction of hepatocellular carcinoma. *Comput. Biol. Medicine* **134**, (2021).
- Khanna, M., Singh, L. K., Shrivastava, K. & Singh, R. An enhanced and efficient approach for feature selection for chronic human disease prediction: A breast cancer study. *Heliyon* **10**, e26799. <https://doi.org/10.1016/j.heliyon.2024.e26799> (2024).
- Ksiażek, W. Explainable thyroid cancer diagnosis through two-level machine learning optimization with an improved naked mole-rat algorithm. *Cancers* **16**, 4128. <https://doi.org/10.3390/cancers16244128> (2024).
- Plawiak, P. & Acharya, U. R. Novel deep genetic ensemble of classifiers for arrhythmia detection using ecg signals. *Neural Comput. Appl.* **32**, 11137–11161. <https://doi.org/10.1007/s00521-018-03980-2> (2020).
- Ruiz de Miras, J., Ibáñez-Molina, A., Soriano, M. & Iglesias-Parro, S. Schizophrenia classification using machine learning on resting state eeg signal. *Biomed. Signal Process. Control.* **79**, (2023).
- Brunese, L., Martinelli, F., Mercaldo, F. & Santone, A. Machine learning for coronavirus covid-19 detection from chest x-rays. *Procedia Comput. Sci.* **176**, 2212–2221 (2020).
- Zhao, W., Jiang, W. & Qiu, X. Deep learning for COVID-19 detection based on CT images. *Sci. Rep.* <https://doi.org/10.1038/s41598-021-93832-2> (2021).
- Singh, L. K., Khanna, M. & Garg, H. Multimodal biometric based on fusion of ridge features with minutiae features and face features. *Int. J. Inf. Syst. Model. Des.* **11**, 37–57. <https://doi.org/10.4018/ijismd.2020010103> (2020).
- Debele, G. R. et al. Incidence of Diabetic Retinopathy and Its Predictors Among Newly Diagnosed Type 1 and Type 2 Diabetic Patients: A Retrospective Follow-up Study at Tertiary Health-care Setting of Ethiopia. *Diabetes Metab Syndr Obes* **14**, 1305–1313 (2021).
- Oseka, M., Jamroz-Witkowska, A. & Mulak, M. Selected problems and challenges of ophthalmic care system in Poland. *OphthaTherapy* **8**, 100–106. <https://doi.org/10.24292/01.ot.300321.2> (2021).
- Seoud, L., Chelbi, J. & Cheriet, F. Automatic grading of diabetic retinopathy on a public database (2015).
- Pratt, H., Coenen, F., Broadbent, D. M., Harding, S. P. & Zheng, Y. Convolutional neural networks for diabetic retinopathy. *Procedia Comput. Sci.* **90**, 200–205 (2016).
- Acharya, U. R., Lim, C. M., Ng, E. Y. K., Chee, C. & Tamura, T. Computer-based detection of diabetes retinopathy stages using digital fundus images. *Proc. Inst. Mech. Eng. Part H J. Eng. Med.* **223**, 545–553 (2009).
- Mansour, R. F. Deep-learning-based automatic computer-aided diagnosis system for diabetic retinopathy. *Biomed. Eng. Lett.* **8**, 41–57. <https://doi.org/10.1007/s13534-017-0047-y> (2017).
- Li, F. et al. Automatic detection of diabetic retinopathy in retinal fundus photographs based on deep learning algorithm. *Transl. Vis. Sci. Technol.* **8**, 4. <https://doi.org/10.1167/tvst.8.6.4> (2019).
- Gangwar, A. K. spsampsps Ravi, V. *Diabetic Retinopathy Detection Using Transfer Learning and Deep Learning*, 679–689 (Springer Singapore, 2020).
- Nahiduzzaman, M. et al. Diabetic retinopathy identification using parallel convolutional neural network based feature extractor and elm classifier. *Expert Syst. Appl.* **217**, 119557. <https://doi.org/10.1016/j.eswa.2023.119557> (2023).
- Kassani, S. H. et al. Diabetic retinopathy classification using a modified xception architecture. In *2019 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, <https://doi.org/10.1109/isspit47144.2019.9001846> (IEEE, 2019).
- Wahab Sait, A. R. A lightweight diabetic retinopathy detection model using a deep-learning technique. *Diagnostics* **13**, 3120. <https://doi.org/10.3390/diagnostics13193120> (2023).
- Canayaz, M. Classification of diabetic retinopathy with feature selection over deep features using nature-inspired wrapper methods. *Appl. Soft Comput.* **128**, 109462. <https://doi.org/10.1016/j.asoc.2022.109462> (2022).
- Qummar, S. et al. A deep learning ensemble approach for diabetic retinopathy detection. *IEEE Access* **7**, 150530–150539. <https://doi.org/10.1109/access.2019.2947484> (2019).
- Modi, P. & Kumar, Y. Smart detection and diagnosis of diabetic retinopathy using bat based feature selection algorithm and deep forest technique. *Comput. Ind. Eng.* **182**, 109364. <https://doi.org/10.1016/j.cie.2023.109364> (2023).
- Hardas, M., Mathur, S., Bhaskar, A. & Kalla, M. Retinal fundus image classification for diabetic retinopathy using svm predictions. *Phys. Eng. Sci. Med.* **45**, 781–791. <https://doi.org/10.1007/s13246-022-01143-1> (2022).
- Singh, L. K., Garg, H. spsampsps Pooja. *Automated Glaucoma Type Identification Using Machine Learning or Deep Learning Techniques*, 241–263 (Springer Singapore, 2019).
- Singh, L. K., Pooja & Garg, H. Detection of glaucoma in retinal fundus images using fast fuzzy c means clustering approach. In *2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, 397–403, <https://doi.org/10.1109/iccis48478.2019.8974539> (IEEE, 2019).
- Singh, L. K., Pooja, & Garg, H. Detection of glaucoma in retinal images based on multiobjective approach. *Int. J. Appl. Evol. Comput.* **11**, 15–27. <https://doi.org/10.4018/ijae.2020040102> (2020).
- Singh, L. K., Khanna, M., Thawkar, S. & Singh, R. A novel hybridized feature selection strategy for the effective prediction of glaucoma in retinal fundus images. *Multimedia Tools Appl.* **83**, 46087–46159. <https://doi.org/10.1007/s11042-023-17081-3> (2023).

33. Aptos dataset. <https://www.kaggle.com/competitions/aptos2019-blindness-detection/data> [Accessed: (17.12.2023)].
34. Aria dataset. <https://paperswithcode.com/dataset/aria/> [Accessed: (17.12.2023)].
35. Drive dataset. <https://www.kaggle.com/datasets/andrewmvd/drive-digital-retinal-images-for-vessel-extraction> [Accessed: (17.12.2023)].
36. Eoptha dataset. <https://www.adcis.net/en/third-party/e-optha/> [Accessed: (17.12.2023)].
37. Eye diseases classification kaggle dataset. <https://www.kaggle.com/datasets/gunavenkatdoddi/eye-diseases-classification> [Accessed: (17.12.2023)].
38. Hrf dataset. <https://www5.cs.fau.de/research/data/fundus-images/> [Accessed: (17.12.2023)].
39. Idrd dataset. <https://ieee-dataport.org/open-access/indian-diabetic-retinopathy-image-dataset-idrid> [Accessed: (17.12.2023)].
40. Odir dataset. <https://www.kaggle.com/datasets/andrewmvd/ocular-disease-recognition-odir5k> [Accessed: (17.12.2023)].
41. Telediagnosics in ophthalmological examinations (2023).
42. Hayati, M. et al. Impact of CLAHE-based image enhancement for diabetic retinopathy classification through deep learning. *Procedia Comput. Sci.* **216**, 57–66. <https://doi.org/10.1016/j.procs.2022.12.111> (2023).
43. Azzopardi, G. & Azzopardi, N. Trainable cosfire filters for keypoint detection and pattern recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**, 490–503. <https://doi.org/10.1109/tpami.2012.106> (2013).
44. Zhu, X. & Rangayyan, R. M. Detection of the optic disc in images of the retina using the hough transform. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, <https://doi.org/10.1109/iembs.2008.4649971> (IEEE, 2008).
45. Li, W. et al. Retinal vessel segmentation based on B-cosfire filters in fundus images. *Front. Public Health* <https://doi.org/10.3389/fpubh.2022.914973> (2022).
46. Kumar, V., Biswas, S., Rajput, D. S., Patel, H. & Tiwari, B. Pca-based incremental extreme learning machine (pca-ielm) for covid-19 patient diagnosis using chest x-ray images. *Comput. Intell. Neurosci.* **2022**, 1–17. <https://doi.org/10.1155/2022/9107430> (2022).
47. Kaplan, K., Kaya, Y., Kuncan, M. & Ertunç, H. M. Brain tumor classification using modified local binary patterns (lbp) feature extraction methods. *Med. Hypotheses* **139**, 109696. <https://doi.org/10.1016/j.mehy.2020.109696> (2020).
48. Nyasulu, C. et al. A comparative study of machine learning-based classification of tomato fungal diseases: Application of GLCM texture features. *Heliyon* **9**, e21697. <https://doi.org/10.1016/j.heliyon.2023.e21697> (2023).
49. Murugan, A., Nair, S. A. H., Preethi, A. A. P. & Kumar, K. P. S. Diagnosis of skin cancer using machine learning techniques. *Microprocess. Microsyst.* **81**, 103727. <https://doi.org/10.1016/j.micpro.2020.103727> (2021).
50. Najjar, F. H. et al. Histogram features extraction for edge detection approach. In *2022 5th International Conference on Engineering Technology and its Applications (IICETA)*, 373–378. <https://doi.org/10.1109/IICETA54559.2022.9888697> (2022).
51. Breiman, L. *Machine Learning* **45**, 5–32. <https://doi.org/10.1023/a:1010933404324> (2001).
52. Chen, T. & Guestrin, C. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, <https://doi.org/10.1145/2939672.2939785> (ACM, 2016).
53. Nopour, R. Prediction of 12-month recurrence of pancreatic cancer using machine learning and prognostic factors. *BMC Med. Inform. Decis. Mak.* <https://doi.org/10.1186/s12911-024-02766-y> (2024).
54. K, A. et al. Effect of multi filters in glaucoma detection using random forest classifier. *Meas. Sens.* **25**, 100566. <https://doi.org/10.1016/j.measen.2022.100566> (2023).
55. Bader, M., Abdelwanis, M., Maalouf, M. & Jelinek, H. F. Detecting depression severity using weighted random forest and oxidative stress biomarkers. *Sci. Rep.* <https://doi.org/10.1038/s41598-024-67251-y> (2024).
56. Islam, T. et al. Predictive modeling for breast cancer classification in the context of Bangladeshi patients by use of machine learning approach with explainable ai. *Sci. Rep.* <https://doi.org/10.1038/s41598-024-57740-5> (2024).
57. Mochurad, L., Babii, V., Boliubash, Y. & Mochurad, Y. Improving stroke risk prediction by integrating xgboost, optimized principal component analysis, and explainable artificial intelligence. *BMC Med. Inform. Decis. Mak.* <https://doi.org/10.1186/s12911-025-02894-z> (2025).
58. Akiba, T., Sano, S., Yanase, T., Ohta, T. & Koyama, M. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2019).
59. Pedregosa, F. et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
60. Chen, T. & Guestrin, C. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, 785–794. <https://doi.org/10.1145/2939672.2939785> (ACM, New York, NY, USA, 2016).
61. Akiba, T., Sano, S., Yanase, T., Ohta, T. & Koyama, M. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2019).
62. van der Walt, S. et al. Scikit-image: Image processing in python. *PeerJ*, e453 (2014).
63. Giakoumoglou, N. Pyfeats: Open-source software for image feature extraction. <https://github.com/giakou4/pyfeats> (2021).
64. Virtanen, P. et al. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272. <https://doi.org/10.1038/s41592-019-0686-2> (2020).
65. Bradski, G. The OpenCV Library. *Dr. Dobbs Journal of Software Tools* (2000).
66. Mujeeb Rahman, K. K., Nasor, M. & Imran, A. Automatic screening of diabetic retinopathy using fundus images and machine learning algorithms. *Diagnostics* **12**, 2262. <https://doi.org/10.3390/diagnostics12092262> (2022).
67. Bibi, I., Mir, J. & Raja, G. Automated detection of diabetic retinopathy in fundus images using fused features. *Phys. Eng. Sci. Med.* **43**, 1253–1264. <https://doi.org/10.1007/s13246-020-00929-5> (2020).
68. Ali, A. et al. Machine learning based automated segmentation and hybrid feature analysis for diabetic retinopathy classification using fundus image. *Entropy* **22**, 567. <https://doi.org/10.3390/e22050567> (2020).
69. Usman, T. M., Saheed, Y. K., Ignace, D. & Nsang, A. Diabetic retinopathy detection using principal component analysis multi-label feature extraction and classification. *Int. J. Cogn. Comput. Eng.* **4**, 78–88. <https://doi.org/10.1016/j.ijcce.2023.02.002> (2023).
70. Makmur, N. M., Kwan, F., Rana, A. D. & Kurniadi, F. I. Comparing local binary pattern and gray level co-occurrence matrix for feature extraction in diabetic retinopathy classification. *Procedia Comput. Sci.* **227**, 355–363. <https://doi.org/10.1016/j.procs.2023.10.534> (2023).
71. Żelasko, D., Ksiażek, W. & Pławiak, P. Transmission quality classification with use of fusion of neural network and genetic algorithm in pay and require multi-agent managed network. *Sensors* **21**, 4090. <https://doi.org/10.3390/s21124090> (2021).
72. Anuradha, K., Krishna, M. V. & Mallik, B. Bio inspired boolean artificial bee colony based feature selection algorithm for sentiment classification. *Meas. Sens.* **32**, 101034. <https://doi.org/10.1016/j.measen.2024.101034> (2024).
73. Nadimi-Shahraki, M. H., Zamani, H. & Mirjalili, S. Enhanced whale optimization algorithm for medical feature selection: A COVID-19 case study. *Comput. Biol. Med.* **148**, 105858. <https://doi.org/10.1016/j.combiomed.2022.105858> (2022).
74. Dosovitskiy, A. et al. An image is worth 16x16 words: Transformers for image recognition at scale. <https://doi.org/10.48550/ARXIV.2010.11929> (2020).
75. Liu, Z. et al. Swin transformer: Hierarchical vision transformer using shifted windows <https://doi.org/10.48550/ARXIV.2103.14030> (2021).
76. Liu, Z. et al. A convnet for the 2020s, <https://doi.org/10.48550/ARXIV.2201.03545> (2022).
77. Tan, M. & Le, Q. V. Efficientnetv2: Smaller models and faster training. <https://doi.org/10.48550/ARXIV.2104.00298> (2021).

Author contributions

Conceptualization: M.G. and F.P.; methodology: F.P. and W.K.; software: M.G., F.P., A.S.; validation: W.K., F.P.; formal analysis: W.K.; investigation: W.K.; resources: F.P., A.S.; data curation: M.G., A.S.; writing—original draft preparation: W.K.; writing—review and editing: F.P., P.P., K.J., R.R., K.N., A.W., M.T., C.A., M.S.; visualization: F.P.; supervision: P.P., T.J.; project administration: M.O.S.; funding acquisition: T.J., M.O.S. All authors have read and agreed to the published version of the manuscript.

Funding

The telediagnosics project in ophthalmological examinations is a joint initiative of the Ministry of Health, the Minister of Digitization, the Medical University of Lublin and the Institute of Telecommunications - National Research Institute. Source of financing: Targeted grant from the Minister of Digitization.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to P.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025