Gazi Üniversitesi
Fen Bilimleri Dergisi
PART C: TASARIM VE TEKNOLOJİ

Gazi University
Journal of Science
PART C: DESIGN AND TECHNOLOGY

GU J Sci, Part C, 14(X): XX-XX (2026)

A Hybrid Classification Approach for Glaucoma Diagnosis with MultiSURF Feature Selection

Halil İbrahim SARI[1] , Kübra KESER[2*]

[1] *Kutahya Dumlupınar University, Simav Faculty of Technology, Department of Electrical and Electronics Engineering, Kutahya, Turkey.*
[2] *Kutahya Dumlupınar University, Faculty of Engineering, Department of Software Engineering, Kutahya,Turkey.*

Article Info

Research article
Received: 16/03/2026
Revision: 23/06/2026
Accepted: 16/07/2026

Keywords

Glaucoma
Retinal fundus images
Hybrid classification
AlexNet
Random Forest
MultiSURF

Makale Bilgisi

Araştırma makalesi
Başvuru: 16/03/2026
Düzeltme: 23/06/2026
Kabul: 16/07/2026

Anahtar Kelimeler

Glokom
Retinal fundus görüntüleri
Hibrit sınıflandırma
AlexNet
Rastgele Orman
MultiSURF

Graphical/Tabular Abstract (Grafik Özet)

In this study, a hybrid method optimizing AlexNet and hand-crafted features with MultiSURF achieved an accuracy of 96.92% with the Random Forest model. / Bu çalışmada, AlexNet ve el yapımı özellikleri MultiSURF ile optimize eden hibrit yöntem, Rastgele Orman modeliyle %96,92 doğruluk elde edilmiştir.

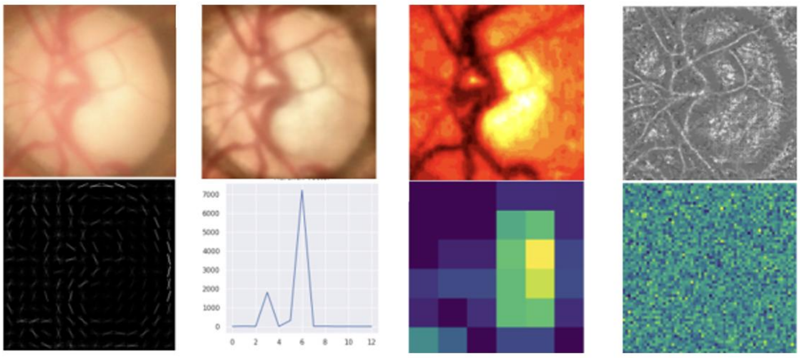


Figure A: Preprocessing and feature extraction images / ***Şekil A:*** Ön işleme ve özellik çıkarma görüntüleri

Highlights (Önemli noktalar)

- *The ORIGA Retina Fundus dataset was used to automatically differentiate between individuals with glaucoma and those with healthy eyes using a machine learning-based model. / A machine learning-based model was used to automatically differentiate between individuals with glaucoma and those with healthy eyes.*
- *Both image processing methods and deep learning-based feature extraction are incorporated to create a multi-stage classification system. / Çok aşamalı bir sınıflandırma sistemi oluşturmak için hem görüntü işleme yöntemleri hem de derin öğrenme tabanlı özellik çıkarma yöntemleri entegre edilmiştir.*
- *Thanks to MultiSURF, the high-dimensional feature space has been significantly optimized. / MultiSURF sayesinde, yüksek boyutlu özellik alanı önemli ölçüde optimize edilmiştir.*
- *The integration of SMOTE as a controlled balancing tool that strengthens the decision frontier ensures high accuracy while maintaining the model's stability against different class distributions./ Karar sınırını güçlendiren kontrollü bir dengeleme aracı olarak SMOTE entegrasyonu, modelin farklı sınıf dağılımlarına karşı kararlılığını korurken yüksek doğruluk sağlar.*

Aim (Amaç): *The aim is to offer a glaucoma classification framework with clinical integration potential by combining MultiSURF-based selection of multidimensional image features, the SMOTE balanced dataset, and independent test evaluation. / Çok boyutlu görüntü özelliklerinin MultiSURF tabanlı seçimi, SMOTE dengeli veri seti ve bağımsız test değerlendirmesini birleştirerek, klinik entegrasyon potansiyeline sahip bir glokom sınıflandırma çerçevesi sunmaktır.*

Originality (Özgünlük): *The originality of this study lies in optimizing a hybrid structure combining AlexNet and handcrafted features via MultiSURF, preventing data leakage using SMOTE, and validating model performance on independent external datasets. / Bu çalışmanın özgünlüğü, AlexNet ve el yapımı özellikleri birleştiren hibrit yapıyı MultiSURF ile optimize etmesi, SMOTE ile veri sızıntısını önlemesi ve model başarısını bağımsız dış veri setlerinde doğrulamasıdır.*

Results (Bulgular): *The hybrid frame was able to distinguish individuals with glaucoma from those without glaucoma with high accuracy. The Random Forest model achieved 96.92% accuracy, a 0.939 F1 score, and a 0.994 AUC in the ORIGA dataset. / Hibrit* çerçeve ile glokomlu bireyleri glokomsuz bireylerden yüksek doğrulukla ayırt edilebilmiştir. *Rastgele Orman modeli ORIGA veri setinde %96,92 doğruluk, 0,939 F1 puanı ve 0,994 AUC başarı göstermiştir.*

Conclusion (Sonuç): *The proposed approach is considered an automated glaucoma screening method with the potential to be integrated into clinical decision support systems for early diagnosis. / Önerilen yaklaşım, erken teşhis için klinik karar destek sistemlerine entegre edilebilecek potansiyele sahip otomatik bir glokom tarama yöntemi olarak değerlendirilmektedir.*

*Corresponding author, e-mail: kubra.keser@dpu.edu.tr
DOI: 10.29109/gujsc.1910949

GU J Sci, Part C, 14(X): XX-XX (2026)

Gazi Üniversitesi
Fen Bilimleri Dergisi
PART C: TASARIM VE TEKNOLOJİ

Gazi University
Journal of Science
PART C: DESIGN AND TECHNOLOGY

http://dergipark.gov.tr/gujsc

A Hybrid Classification Approach for Glaucoma Diagnosis with MultiSURF Feature Selection

Halil İbrahim SARI[1], Kübra KESER[2*]

[1]Kutahya Dumlupınar University, Simav Faculty of Technology, Department of Electrical and Electronics Engineering, Kutahya, Turkey.

[2]Kutahya Dumlupınar University, Faculty of Engineering, Department of Software Engineering, Kutahya, Turkey.

Article Info

Research article
Received: 16/03/2026
Revision: 23/06/2026
Accepted: 16/07/2026

Keywords

Glaucoma
Retinal fundus images
Hybrid classification
AlexNet
Random Forest
MultiSURF

Abstract

Glaucoma is a severe optic neuropathy that can lead to irreversible vision loss and is often asymptomatic in its early stages. This situation increases the need for automated and reliable screening systems, especially those based on retinal fundus images. This study presents a hybrid classification framework for glaucoma detection that combines classical image processing methods with deep learning-based representations, possessing the potential to support clinical processes and exhibiting generalizability. In the proposed approach, deep features obtained from the fc7 layer of a pre-trained AlexNet architecture, along with HSV colour histograms, Haralick textural features, LBP and HOG descriptors, are combined using the ORIGA retinal fundus dataset. The class imbalance problem was addressed by applying the SMOTE method exclusively to the training data, while the high-dimensional feature space was optimised using the MultiSURF algorithm. The selected features obtained were evaluated using Support Vector Machines (SVM), Random Forest (RF), and XGBoost classifiers. Experimental results demonstrate that the proposed hybrid framework can accurately distinguish between individuals with glaucoma and those without, with high accuracy. Specifically, the Random Forest model demonstrated the highest performance with 96.92% accuracy, an F1 score of 0.939, and an AUC value of 0.994. The findings reveal that the combined use of deep and hand-crafted features, supported by MultiSURF-based feature selection and balanced learning strategies, offers a robust and stable solution for glaucoma detection. The proposed approach is considered an automated glaucoma screening method with the potential to be integrated into clinical decision support systems for early diagnosis.

MultiSURF Özellik Seçimi ile Glokom Tanısı için Hibrit Bir Sınıflandırma Yaklaşımı

Makale Bilgisi

Araştırma makalesi
Başvuru: 16/03/2026
Düzeltme: 23/06/2026
Kabul: 16/07/2026

Anahtar Kelimeler

Glokom
Retinal fundus görüntüleri
Hibrit sınıflandırma
AlexNet
Rastgele Orman
MultiSURF

Öz

Glokom, geri dönüşü olmayan görme kaybına yol açabilen ve erken evrelerinde genellikle asemptomatik olan ciddi bir optik nöropatidir. Bu durum, özellikle retina fundus görüntülerine dayalı otomatik ve güvenilir tarama sistemlerine olan ihtiyacı artırmaktadır. Bu çalışmada, glokom tespiti için klasik görüntü işleme yöntemlerini derin öğrenme tabanlı temsillerle birleştiren, klinik süreçleri destekleme potansiyeline sahip ve genelleştirilebilir nitelikte hibrit bir sınıflandırma çerçevesi sunulmaktadır. Önerilen yaklaşımda, önceden eğitilmiş bir AlexNet mimarisinin fc7 katmanından elde edilen derin özellikler, HSV renk histogramları, Haralick dokusal özellikleri, LBP ve HOG tanımlayıcıları ile birlikte ORIGA retina fundus veri seti kullanılarak birleştirilmiştir. Sınıf dengesizliği sorunu, SMOTE yönteminin yalnızca eğitim verilerine uygulanmasıyla ele alınırken, yüksek boyutlu özellik alanı MultiSURF algoritması kullanılarak optimize edilmiştir. Elde edilen seçilen özellikler, Destek Vektör Makineleri (SVM), Rastgele Orman (RF) ve XGBoost sınıflandırıcıları kullanılarak değerlendirilmiştir. Deneysel sonuçlar, önerilen hibrit çerçevenin, glokomlu bireyleri glokomsuz bireylerden yüksek doğrulukla ayırt edebildiğini göstermektedir. Özellikle, Rastgele Orman modeli %96,92 doğruluk, 0,939 F1 puanı ve 0,994 AUC değeri ile en yüksek performansı göstermiştir. Bulgular, MultiSURF tabanlı özellik seçimi ve dengeli öğrenme stratejileriyle desteklenen derin ve el yapımı özelliklerin birleşik kullanımının, glokom tespiti için sağlam ve istikrarlı bir çözüm sunduğunu ortaya koymaktadır. Önerilen yaklaşım, erken teşhis için klinik karar destek sistemlerine entegre edilebilecek potansiyele sahip otomatik bir glokom tarama yöntemi olarak değerlendirilmektedir.

*Corresponding author, e-mail: kubra.keser@dpu.edu.tr
DOI: 10.29109/gujsc.1910949

1. INTRODUCTION (GİRİŞ)

Medical imaging is an indispensable diagnostic tool that enhances the effectiveness of diagnostic and treatment processes by enabling detailed examination of organs, tissues, and systems in the human body. In modern medical practice, it offers the possibility of detecting and characterizing pathologies at an early stage by visualizing internal body structures using non-invasive or minimally invasive methods. The techniques used provide critical data for diagnosis, monitoring, and treatment planning [1,2]. However, the quality of the images obtained can sometimes be limited due to physical constraints and artifacts such as radiation exposure, tissue contrast, resolution, and noise. To overcome these limitations and improve clinical interpretability, advanced methods and digital image processing have been developed [2,3]. In recent years, the application of deep learning-based algorithms in image enhancement techniques has yielded promising results in noise reduction, resolution enhancement, and transformations between different structures [3]. Understanding the anatomical structure of the eye is necessary to understand how these advanced imaging methods and deep learning algorithms detect glaucomatous damage.

The eye is a complex sensory organ that enables vision by perceiving light and converting it into nerve impulses. It consists of three main layers: the fibrous layer (sclera and cornea), the vascular layer (choroid, ciliary body, and iris), and the nerve layer (retina) (Figure 1). The sclera is a strong connective tissue that surrounds the outer part of the eye. It maintains the shape of the eyeball and serves as the attachment point for the muscles that control eye movement. The cornea is the transparent, avascular layer at the front of the eye. It plays a critical role in vision as it is the first structure to refract light. It is susceptible due to its dense network of nerve endings. The iris is the colored part of the eye, and the pupil, the hole in the centre, regulates the amount of light entering the eye. The ciliary body contains muscles that adjust the curvature of the lens, providing focusing. It also produces the intraocular fluid known as aqueous humour. The choroid, situated between the retina and the sclera, nourishes the retina through its dense vascular structure. The retina contains photoreceptors (rod and cone cells) that convert light into electrical signals. It is the most critical structure where vision occurs [4]. Light first passes through the cornea, the transparent and translucent structure on the front surface of the eye, which begins to refract the light [5]. Behind the cornea lies the iris, which acts as a diaphragm to regulate the amount of light entering the eye; the space in the centre of the iris is called the pupil. Behind the iris is a lens that, by changing its shape (accommodation), allows light to be focused clearly onto the retina [4]. Images entering the eye are formed upside down and reduced in size on the retina, which contains photoreceptor cells. These signals are then transmitted to the brain via the optic nerve, resulting in vision [6].

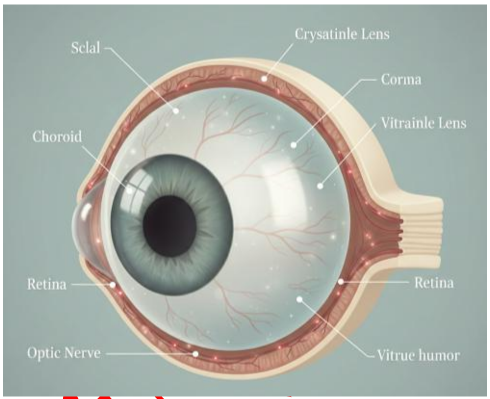


Figure 1. Anatomy of the eye (Gözün anatomisi)

Many diseases and disorders affect various components of the eye's anatomy [7–10]. Glaucoma is a progressive ophthalmic disease characterized by optic nerve damage, often driven by elevated intraocular pressure (IOP) or structural sensitivities independent of pressure [7,8]. Commonly known as "black water" among the public, this condition leads to irreversible vision loss and typically progresses asymptomatically until advanced stages [7,8,11]. Chronic impairment of the eye's fluid drainage system leads to elevated IOP, which exerts mechanical stress on delicate nerve fibers and causes permanent damage to the optic nerve [11,12]. If left untreated, the initial loss of peripheral vision can advance to central vision impairment and eventually lead to blindness. Factors such as high IOP, advanced age, family history, ethnic background, and corneal thickness significantly increase the risk of developing glaucoma [13].

Glaucoma treatment aims to stop or slow optic nerve damage by lowering intraocular pressure. Current treatments do not completely cure the disease, but they largely control its progression. Some of these treatment methods include drug therapy, laser treatments, surgical treatment. Glaucoma is difficult to diagnose, primarily due to its insidious progression without symptoms, the inadequacy of a single test, and the slow progression of the disease. The condition most often confused with or associated with glaucoma is ocular hypertension. In fact, ocular hypertension is a condition that does not yet have glaucoma but carries a risk of developing it.

When a patient is found to have high intraocular pressure (IOP), that patient is at high risk of developing glaucoma. However, if nerve damage has not yet begun, it is not yet glaucoma [11]. Radiologists and ophthalmologists closely monitor patients with ocular hypertension. If signs of damage to the optic nerves or visual fields appear in these patients, the diagnosis changes from ocular hypertension to glaucoma. Therefore, prioritising optic nerve health in diagnosis, rather than solely focusing on pressure values, is critically important.

In diagnosing glaucoma, tests are used to detect optic nerve damage, IOP, and visual field losses. Intraocular Pressure Measurement (IOP) is the most common and standard method, using a Goldmann applanation tonometer. This device gently touches the cornea and measures the force required to resist pressure. Optic Nerve Assessment checks for possible damage to the optic nerve, specifically thinning and cupping of the optic nerve head. Glaucoma typically leads to increased cupping in the optic disc. The optic nerve is directly examined using an ophthalmoscope after dilation of the pupils or via fundus photography. The radiologist determines the degree of damage by assessing the cup-to-disc ratio in the optic disc [11]. A visual field test is used to determine the functional impact of the disease on vision, specifically to assess whether the patient is experiencing peripheral vision loss. While the patient looks at a central point, they look at a machine that indicates how well different levels of light stimuli are perceived in the peripheral field of vision. Glaucoma typically develops patterns of peripheral vision loss over time while preserving central vision [14,15]. Drainage angle examination involves checking whether the anterior chamber angle of the eye is open or narrowed/closed. This distinction is vital for determining the type of glaucoma. The drainage angle of the eye is examined using a special lens [16]. Corneal thickness measurement is the measurement of the thickness of the cornea. Corneal thickness can affect the IOP value measured by a tonometer. In individuals with thin corneas, the tonometer may display an IOP lower than usual, while in individuals with thick corneas, it may display an IOP higher than usual. This measurement is necessary for a more accurate interpretation of the IOP assessment [11]. OCT allows for high-resolution measurements of the optic nerve head and peripapillary retinal nerve fibre layer, and these structural abnormalities can be objectively detected at the micrometre level, often earlier than visual field defects, enabling the monitoring of progression. Therefore, OCT parameters are frequently used in the clinical monitoring of glaucoma progression [17,18] .

Kaya and Bilgen [19], highlight a study focusing on the use of Vision Transformer (ViT)-based hybrid learning methods. The study aimed to develop a high-accuracy model for glaucoma detection. In one of the proposed models, the CLAHE (Contrast-Limited Adaptive Histogram Equalization) filtering technique was used to enhance contrast before ViT. The proposed Fine Tune ViT-PCA-SVM and Fine Tune ViT-LDA-SVM hybrid models achieved a classification success rate of 92%.

Juneja et al. [20] aimed to develop an automated system to classify glaucoma from retinal images with high accuracy in their studies focusing on glaucoma detection from retinal images. Using the Glaucoma Classification Network (CoG-NET), this study reported achieving 95.3% accuracy, 0.95 sensitivity and 0.99 specificity.

Urmamen and Koçer [21], used deep learning methods to automate diagnosis from OCT images. InceptionV3, Xception, and a proposed Convolutional Neural Network (CNN) model were compared. Accuracy rates of 95.36%, 98.2%, and 97.51% were achieved with the Xception and Inception CNN architectures, respectively. Furthermore, it was emphasized that the proposed architecture achieved better results than other methods in classifying diabetes and normal conditions by disease type.

Ayan and Serttaş [22] used deep learning methods in their study on the classification of retinal diseases. The study, which aims to enable automatic classification of diseases and improve diagnostic processes, includes CNN and transfer learning models. Each model was evaluated in detail using criteria such as accuracy, sensitivity, recall, and F1 score. Sudhan et al. [23] in their study focusing on segmentation and classification methods using a deep learning model, aim to propose and develop a deep learning-based system for early diagnosis using retinal images. The study utilizes the ORIGA dataset with a U-Net architecture and DenseNet-201 Deep Convolutional Neural Network (DCNN). The performance of the proposed model was evaluated using parameters such as accuracy, precision, sensitivity, specificity, and F-score. Accuracy was 98.82% during training and 96.90% during testing. Uçar [24], aims to implement and compare the performance of CNN methods using digital fundus images obtained from a newly created dataset. The study includes VGG16, Inception-V3, EfficientNet, DenseNet, ResNet50, and MobileNet architectures. Among the models, the DenseNet architecture recorded the highest accuracy rate in glaucoma detection with 96.19%. Gök and Taşdemir[25] present an approach combining deep learning methods and CLAHE for the automated diagnosis of retinal diseases from fundus images. The aim of the study is to develop a computer-aided diagnostic system. MobileNet, VGG16, and DenseNet169 models were used for classification. Accuracy rates were 92.5% for MobileNet, 91% for VGG16, and 91.7% for DenseNet169, respectively.

Şahin Sadık [26] extracted correlation, energy, homogeneity, contrast, and entropy features from photographs for glaucoma detection using the Gray Level Simultaneity Matrix technique. He employed recurrent neural networks, classification and regression tree algorithms, and the k-nearest neighbor algorithm. He found classification success rates to be an average of 86.7% with k-NN, 87.8% with decision trees, and 96.7% with the artificial neural network (ANN) algorithm.

Thakur and Juneja [27] utilize hybrid features and machine learning approaches (K-nearest neighbor (k-NN), Neural Network (NN), Random Forest (RF), Support Vector Machine (SVM), and Naive Bayes (NB)) and the DRISHTI-GS and RIM-One datasets. Both structured and unstructured features are combined in the paper. The highest accuracy rates were obtained with the SVM classifier using the reduced hybrid feature set. The accuracy rate for the 5-fold cross-validation method was 93.2%, while the accuracy rate for the split test/train approach was 97.22%.

Joshi et al. [28] combined image processing and supervised machine learning approaches for the automatic detection of glaucoma. Three different pre-trained CNN models (Residual Network (ResNet), Visual Geometry Group Network (VGGNet), GoogLeNet) were used, and their results were combined. The model's performance was tested on five different datasets (DRISHTI-GS, DRIONS-DB, HRF (High-Resolution Fundus), PSGIMSR Dataset, and Combined Datasets). With the proposed structure, 91.13% accuracy, 86.58% sensitivity, and 95.21% specificity were achieved in PSGIMSR, while 98.67% accuracy, 100% sensitivity, and 97.40% specificity were achieved in HRF. Latif et al. [29] evaluated transfer learning-based models integrated with discrimination maps such as AlexNet, ResNet, and VGGNet on five datasets. They reported that the proposed ODGNet achieved 95.75% accuracy, 94.90% specificity, 94.75% sensitivity, and 97.85% positive predictive value on the ORIGA dataset.

Sonti and Dhuli [30], focus on developing a novel CNN model for glaucoma detection and classification from retinal fundus images. They present the KR-NET model, a 26-layer CNN architecture. Region of Interest (ROI) segmentation, data augmentation techniques, and classification are applied. The results indicate that the KR-NET model effectively classifies retinal fundus images regardless of dataset size and type, achieving particularly high accuracy in the balanced ACRIMA dataset. Jun et al. [31] categorized fundus images into three categories. They reported that the Transferable Sequential Convolutional Neural Network (TRk-CNN) performed significantly better than traditional methods in identifying the 'suspected glaucoma' category. The model's performance metrics were: Mean accuracy: 92.96%, Specificity: 93.33%, Sensitivity for suspected glaucoma: 95.12%, and Sensitivity for true glaucoma: 93.98%.

Yıldırım and Özbay [32] used Convolutional Neural Network models such as AlexNet, ResNet-18, VGG16, SqueezeNet, and GoogleNet for glaucoma detection and examined the accuracy, sensitivity, specificity, and F1-measure values of each model. They stated that VGG16 showed a sensitivity of 97.96% in the test dataset, while the best values for specificity, accuracy, and F1-measure were obtained with GoogleNet at 98.97%, 97.98%, and 98%, respectively.

Camara et al. [33] aim to examine the current state and most up-to-date methods of artificial intelligence and deep learning techniques used for screening, segmentation, and classification in glaucoma diagnosis. It is stated that deep learning techniques exhibit high sensitivity and specificity in glaucoma screening.

Thanki [34] presents a hybrid approach for glaucoma detection from retinal fundus images and aims to develop a computer-aided triage system for classifying retinal images. A deep CNN model called SqueezeNet is used. Additionally, six different machine learning classifiers (kNN, Decision Tree, SVM, Random Forest, Naive Bayes, and Logistic Regression) are included. The system was tested on the DRISTHI-GS1 and ORIGA datasets. It was determined that Logistic Regression was the best classifier for both datasets.

Table 1. Brief summary of the literature (Literatürün kısa özeti)

Dataset	Methods and Algorithms Used	Performance Success	Study
SMDG-19	CLAHE+ Finely Tuned ViT-PCA-SVM / ViT-LDA-SVM Hybrid Models	92.00%	Kaya and Bilgen [19]
Drishti-GS RIM-One REFUGE ACRIMA	CoG-NET	95.3% accuracy, 0.95 Sensitivity, 0.99 Specificity	Juneja et al. [20]
ODIR-5K	InceptionV3, Xception and Proposed CNN Comparison	98.20 % (Xception), 97.51 % (Inception CNN), 95.36 % (InceptionV3)	Urmamen and Koçer [21]
RFMID	CNN and Transfer Learning Models	Detailed analysis with Accuracy, Sensitivity and F1-Score	Ayan and Serttaş [22]
ORIGA	U-Net Architecture + DenseNet-201 DCNN	Training: 98.82% accuracy Test: 96.90% accuracy	Sudhan et al. [23]
Large-scale Data Set	VGG16, Inception-V3, EfficientNet, DenseNet, ResNet50, MobileNet	96.19% accuracy	Uçar [24]
IDRiD, Oculur re-cognition, HRF, and various other sources	CLAHE + MobileNet, VGG16, DenseNet169 Modelleri	92.50% (MobileNet), 91.70% (DenseNet169), 91.00% (VGG16)	Gök and Taşdemir [25]
HRF Database	Gray Level Co-occurrence Matrix ANN, CART, k-NN	96.70% (ANN), 87.80% (decision trees), 86.70% (k-NN)	Şahin Sadık [26]
DRISHTI-GS ve RIM-One	Hybrid features + Size Reduction + SVM, k-NN, NN, RF, NB	97.22% (Test/Train Distinction Accuracy), 93.20% (5-Fold Cross-Validation)	Thakur and Juneja[27]
PSGIMSR DRISHTI-GS DRIONS-DB HRF Combined	ResNet, VGGNet, GoogLeNet	PSGIMSR 91.13% accuracy 86.58% Sensitivity ve 95.21% Specificity HRF 98.67% accuracy, 100% Sensitivity ve 97.40% Specificity	Joshi et al. [28]
ORIGA	AlexNet (fc7) + MultiSURF + RF/SVM/XGBoost	96,92% accuracy, 0.939 F1 score ve 0.994 AUC	This work

Studies investigating glaucoma detection have primarily focused on manual feature extraction or deep learning models (Table 1). However, these approaches fail to effectively capture nonlinear feature interactions in retinal images, which are critical for accurate classification. To obtain consistent and reproducible results, the selection of multidimensional image features and balanced datasets, combined with the application of image processing methods and deep learning-based feature extraction, is necessary. This study presents a glaucoma classification framework with clinical integration potential, combining MultiSURF-based selection of multidimensional image features, a SMOTE-balanced dataset, and independent test evaluation.

- The ORIGA Retina Fundus dataset was used to automatically differentiate between individuals with glaucoma and those with healthy eyes using a machine learning-based model.
- Both image processing methods and deep learning-based feature extraction are incorporated to create a multi-stage classification system.
- Thanks to the effective combination of MultiSURF's hand-crafted and deep features, the

high-dimensional feature space is significantly optimized.

- SMOTE is integrated as a controlled balancing tool that strengthens the decision frontier. This integration ensures high accuracy while maintaining the model's stability against different class distributions.

The demonstrated performance offers not only metric superiority but also the potential for adaptability to a wide range of clinical data.

2. MATERIALS AND METHODS (MATERYAL VE METOD)

2.1. Data Set and Experimental Environment (Veri Kümesi ve Deneysel Ortam)

This study utilised ORIGA (Online Retinal Fundus Image Database), a widely used standard dataset, to develop a classification model for glaucoma detection[35,36]. The dataset consists of colour fundus images labelled as glaucomatous and non-glaucomatous. This study used a hybrid classification framework, and predefined sections of the dataset were used. According to the experimental setup, the training set consisted of 454 images (337 healthy controls and 117 Glaucomatous) to allow the MultiSURF algorithm to optimize its feature space. The independent testset consisted of 196 images (145 healthy controls and 51 Glaucomatous) to extensively evaluate the generalization ability of the trained models. The same method was also tested on Drishti-GS[37] and RIM-ONE-r3[38,39] data. Experiments were conducted in a CPU-based environment on Google Colab using Python 3.12. Libraries such as OpenCV, scikit-learn, mahotas, PyTorch, and XGBoost were utilised throughout the workflow, and reproducibility was ensured by stabilising sources of randomness. Intermediate outputs resulting from feature extraction were cached to optimise computational cost. The workflow of the proposed framework in the study is shown in Figure 2.

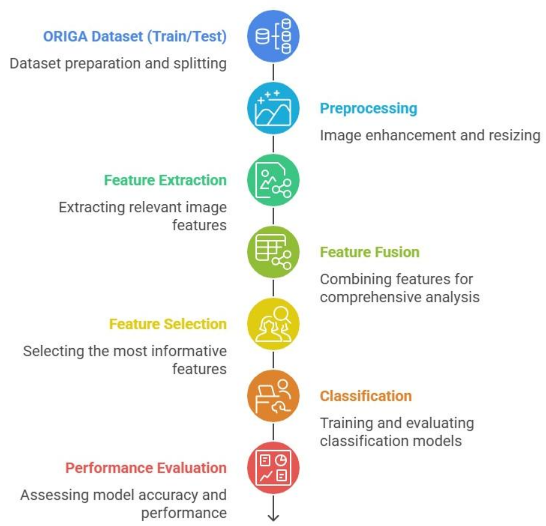


Figure 2. Workflow of the proposed framework (Önerilen çerçevenin iş akışı)

2.2. Image Preprocessing (Görüntü Ön İşleme)

Prior to feature extraction, a standard preprocessing procedure was applied to all fundus images. All images were rescaled to 224 × 224 pixels, and pixel values were normalised to comply with the input requirements of the AlexNet architecture used in the deep learning-based feature extraction phase. To achieve more stable colour representations, images in the RGB colour space were converted to the HSV colour space.

RGB defines colours in terms of the intensity of light sources. Changing the value of each colour component (R, G, B) causes unintuitive and complex effects on colour. HSV defines colours using three properties that are much closer to how the human eye perceives a colour. H (Hue) is the angular position around the colour wheel that expresses the type of colour. S (Saturation) expresses the purity or intensity of the colour. V (Value/Brightness) expresses how light or dark the colour is. In this model, the lighting and colour content components of an image are separated. The advantages of HSV in terms of intuitive colour manipulation, ease of image processing and analysis, greater resilience to lighting changes, and easy thresholding have made this transformation necessary. This transformation is seen in applications such as object recognition and tracking, image segmentation, colour correction and filtering, and skin tone detection [40–44].

Table 2. Dimensional distribution of the extracted hybrid feature space components (Çıkarılan hibrit özellik uzayı bileşenlerinin boyutsal dağılımı)

Feature Extraction Method	Feature Type	Vector Dimension
HSV Color Histograms	Color / Statistical	96
Haralick (GLCM)	Texture	13
Local Binary Pattern (LBP)	Texture / Local	59
Histogram of Oriented Gradients (HOG)	Shape / Edge	6084
AlexNet (FC7 Layer)	Deep Feature	4096
Total Combined Feature Space	Hybrid Vector	10348

2.3.1. HSV colour histograms (HSV renk histogramları)

Each channel was partitioned into 32 bins, and the resulting histogram values were concatenated to obtain a 96-dimesional statistical colour feature vector. This representation efficiently captures the chromatic layout of the fundus images, while maintaining computational efficiency.

2.3.2. LBP and HOG properties (LBP ve HOG özellikleri)

Handcrafted textural and shape descriptors were extracted to add detailed geometric and spatial structures to the color metrics. Haralick textural descriptors were first calculated from the Gray-Level Co-occurrence Matrix (GLCM) using the Mahotas library to extract simple statistical properties like contrast, homogeneity, entropy, energy and correlation. These measures are important to quantitatively characterize structural density changes in the optic nerve head. Secondly, LBP features were extracted to model the local and micro-level patterns in the tissue texture. Finally, HOG analyses were used to robustly represent edge orientations and structural gradient distributions around the optic disc regions.

2.3.3. AlexNet (fc7 Layer) deep features (AlexNet (fc7 Katmanı) derin özellikleri)

AlexNet is a pioneering deep CNN architecture that revolutionised deep learning in computer vision by winning the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012 [45]. The model's key importance lies in depth, GPU utilisation, and over-coherence reduction [46]. AlexNet paved the way for modern deep learning architectures (VGG, ResNet, etc.) [45,47]. It can be used for various purposes, such as Basic Image Classification, Feature Extraction, Transfer Learning, image localisation, object detection, and Similarity Search [48,49]. Modern and deep architectures like ResNet, DenseNet or EfficientNet have achieved high success rates, but the large number of layers and parameters pose a risk of overfitting, especially in small medical image datasets. We selected AlexNet because its 8-layer structure can reduce the computational cost and shorten the training time, and can extract stable features from our existing dataset without overfitting. In this study, 4096-dimensional feature vectors obtained from the fc7 layer of the AlexNet architecture, pre-trained on the ImageNet dataset, were used for deep representation learning. The FC7 layer was preferred because it offers more distinctive and abstract semantic representations in terms of classification, rather than low-level texture and edge information compared to earlier layers (e.g., FC6). This aims to model glaucomatous structural changes more effectively. These in-depth features represent high-level structural patterns that are difficult to capture with classical methods.

2.4. Feature Combining and Representation Space Creation (Özellik Birleştirme ve Temsil Alanı Oluşturma)

A comprehensive feature matrix was created by combining HSV histograms, Haralick textural features, LBP/HOG descriptors, and deep features extracted from the fc7 layer of AlexNet. The aim at this stage is to collect the colour, texture, and deep representation information obtained from the image into a single unified structure and to obtain a richer representation space for classification models.

2.4.1. Addressing class imbalance (Sınıf dengesizliğinin ele alınması)

The presented study's ORIGA dataset comprises a total of 650 retinal fundus images. Of these images, 178 belong to individuals with glaucoma, and 472 belong to healthy individuals. The fact that the glaucomatous class constitutes 27.3% of the total data and the healthy class constitutes 72.7% clearly reveals a significant class imbalance in the dataset. This unbalanced structure carries the risk that the glaucomatous samples may not be adequately represented by the model. Therefore, the Synthetic Minority Over-sampling Technique (SMOTE) method was preferred for application to the training set within the scope of the study. Thus, the aim was to enable the model to learn the minority class samples more accurately and to improve the classification performance in general. These distributional characteristics of the dataset serve as an important reference point in evaluating experimental results.

SMOTE is used in machine learning when there is an imbalance between classes in the dataset. Unlike simple oversampling methods that only copy existing samples, it reduces the risk of the model memorising by generating synthetic samples. Synthetic samples help the classifier learn a stronger decision boundary by expanding the boundaries of the minority class in the feature space. The aim is to balance the dataset by creating new, synthetic samples in the feature space of the minority class, thereby enabling the model to learn from the minority class correctly [50]. It uses a mathematical approach to create synthetic samples [51]. In this study, SMOTE was applied only to the training data; the test set was not altered in any way. Thus, the model learned the glaucomatous class in a more representative manner.

2.5. Feature Selection with MultiSURF (MultiSURF ile Özellik Seçimi)

In this study, preliminary experiments with traditional feature selection methods (e.g., PCA and mutual information) showed less stable performance; therefore, MultiSURF was chosen due to its ability to model higher-order feature interactions and implemented using the Scikit-rebate library. Furthermore, we used the MultiSURF algorithm to extract the most discriminative features from the combined high-dimensional feature vector. In this study, 300 of the most informative features were identified through a selection process performed on the exact 10,348-dimensional feature space (the detailed distribution of which is given in Table 2). The classification phase used only these selected features to optimize the overall accuracy and generalizability of the model. Only these selected features were used in the classification phase to optimise the overall accuracy and generalizability of the model.

2.6. Classification Framework (Sınıflandırma Çerçevesi)

Machine learning algorithms, such as Random Forest, XGBoost, and Support Vector Machines (SVM), are among the fundamental methods widely used in classification and regression problems, whose performance is frequently evaluated in comparative studies [52,53]. After feature selection, modelling was performed using SVM, Random Forest (RF), and Extreme Gradient Boosting (XGBoost) classifiers. All models were optimised on the training set and then evaluated on the model-independent test set.

The primary goal of SVM is to identify a hyperplane that separates points belonging to different classes in the dataset, with the broadest possible margin between classes. The SVM algorithm uses only support vectors to determine the decision boundary. This feature enhances the algorithm's robustness against noise and outliers [54]. What makes SVM particularly powerful is that it can be used even when the data cannot be linearly separated. It is widely used in various applications in the field of machine learning, including bioinformatics and medical diagnostics, text classification, image processing, and face detection [55].

RF is a robust ensemble learning algorithm that combines multiple decision trees, used primarily for classification and regression tasks [56]. It is formed by combining numerous individual decision trees. Each of these trees is trained on a random subset of the dataset. This ensemble approach compensates for the weaknesses of individual trees, providing higher accuracy and generalisation ability. It has two main mechanisms: Random sampling and feature randomness, which ensure diversity among the trees [56]. ue to its main advantages, it is frequently used in finance, healthcare, e-commerce, and image processing [56,57].

XGBoost is an effective and popular ensemble learning algorithm used for both classification and regression problems in machine learning [58]. It has gained fame for its high performance, especially in data science competitions such as Kaggle. XGBoost is based on the Gradient Boosting framework, but it optimises it in terms of speed, performance, and resource utilisation. Its working mechanism is a combination of ensemble learning and optimisation. It is used in various fields, including financial risk modelling, medical diagnostics, demand forecasting, and pricing [58,59].

2.7. Performance Evaluation Criteria (Performans Değerlendirme Kriterleri)

Key metrics, including accuracy, precision, recall, and F1-score, were calculated to evaluate the performance of the classifiers [60,61]. In addition, area under the ROC curve (AUC) values were obtained to analyze the discrimination performance of the models.

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

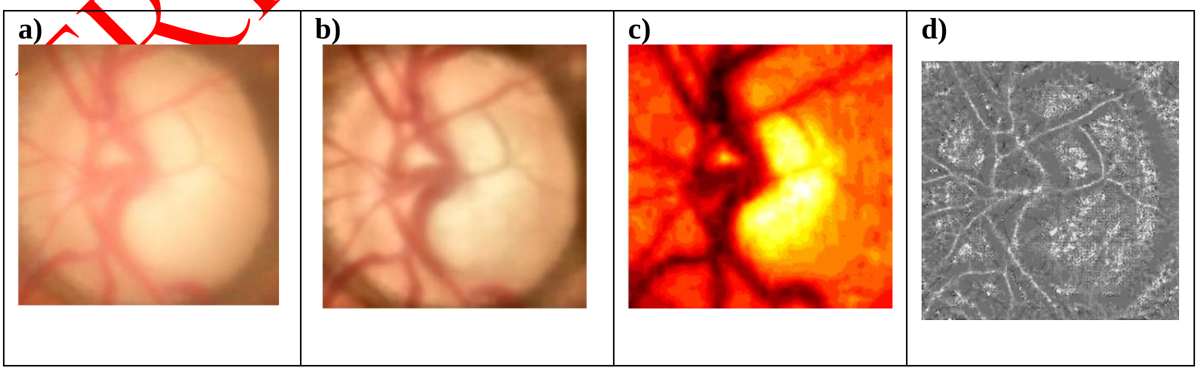
$$Precision = \frac{TP}{TP + FP} \quad (2)$$

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

$$F1\ Score = \frac{2.TP}{2.TP + FP + FN} \quad (4)$$

3. RESULTS (BULGULAR)

This study aims to automatically differentiate between individuals with glaucoma and those without using a machine learning-based model with the ORIGA Retinal Fundus dataset. The study combines image processing methods with deep learning-based feature extraction to create a multi-stage classification system. A multi-component feature fusion strategy is applied to transform the information obtained from retinal fundus images into a more comprehensive and highly representative structure. In the first stage, histogram-based features based on the Hue channel are calculated to more accurately model the colour components and pigment distributions of the images. Subsequently, LBP, which reflects the microstructural patterns of the fundus tissue, HOG, which defines large-scale edge orientations, and Haralick texture vectors derived from grey-level co-occurrence matrices are obtained. These features are sensitive to critical structural components such as vascular density, optic disc contours, and pathological changes in the retinal tissue. In addition to classical colour and texture-based features, a pre-trained AlexNet deep neural network is used to capture more abstract and high-level representations of the images. In this work, the 4096-dimensional feature vector directly extracted from FC7 layer is used as the deep feature representation of the hybrid model. Additionally, the intermediate convolutional feature maps from FM1 to FM12 were extracted and used only for qualitative visualization to show how the network visually interprets intraocular structures such as the optic disc and retinal vasculature. These layers offer a multi-layered representation capacity, ranging from early-level line and gradient structures to more complex semantic patterns, which contribute to the model's more efficient learning of structural differences. Figure 3 visually summarises the image processing and feature extraction steps used in the study.



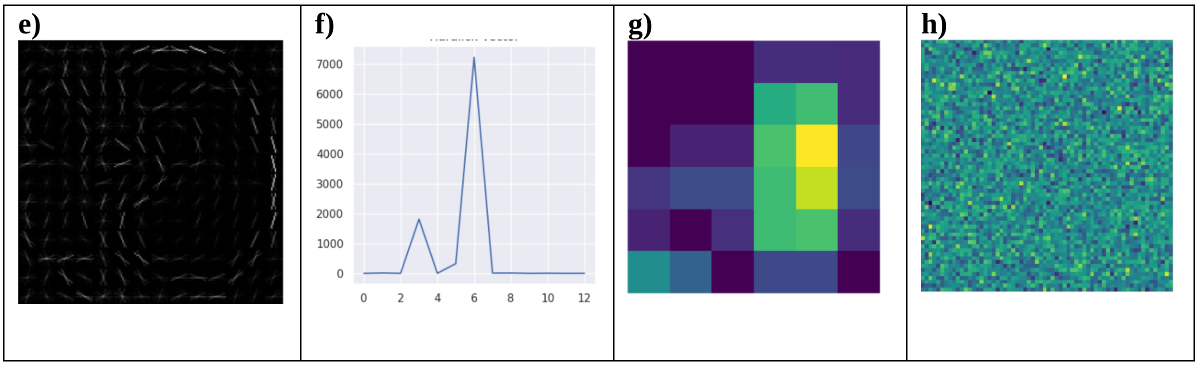


Figure 3. Preprocessing and Feature Extraction Images: a) Original b) Preprocessed c) Hue d) LBP e) HOG f) Haralick Vector g) AlexNet FM11 h) AlexNet fc7 activations heatmap

(Ön İşleme ve Özellik Çıkarma Görüntüleri: a) Orijinal b) Ön İşlenmiş c) Hue d) LBP e) HOG f) Haralick Vector g) AlexNet FM11 h) AlexNet fc7 aktivasyon ısı haritası)

This comprehensive demonstration systematically reveals how both classical image processing techniques and deep learning-based feature extraction processes are applied. The original fundus image, as shown in Figure 3a, presents the natural state of the images in the raw dataset. Variability in brightness, contrast, and colour balance shows the characteristics of the preprocessing state. The preprocessed image in Figure 3.b is a noise-reduced version, normalised for brightness, colour balance, and sharpened vessel boundaries, making it suitable for model input. Thus, basic anatomical structures become more prominent, and the capacity of machine learning models to capture distinctive patterns is increased. Within the scope of colour space-based preprocessing, the Hue component, as shown in Figure 3.c, highlights pigment changes around the optic disc and tonal differences in vascular structures. This component is crucial in combating colour-based discrimination. The LBP image presented in Figure 3.d represents the binary encoding of local tissue patterns. This method enables the capture of microtissue changes associated with glaucoma and other optic nerve diseases. Similarly, the HOG output in Figure 3.e provides a holistic representation of the directional gradients in prominent geometric structures such as vessel edges and the optic disc boundary.

Haralick features, another component of texture-based feature extraction, are shown in vector form in Figure 3.f. These statistical features, based on GLCM, include metrics of the image such as energy, homogeneity, contrast, and entropy. These features are compelling in capturing irregularities in the optic disc and surrounding tissue.

In the deep learning-based feature extraction process, Figure 3.g shows an example of Feature Map 11 (FM11) obtained from the AlexNet interlayers. This activation map demonstrates that the model's mid-level filters are susceptible to vascular density regions and the structural boundaries of the optic disc. Finally, Figure 3.h shows the activation heat map of the AlexNet fc7 layer. These high-dimensional representations contain abstract and complex features of the image, forming the basis for the model's decision-making process during the classification phase. Thanks to this multi-stage approach, the combined use of both classical image processing techniques and deep learning-based feature extraction has enabled the model to create a richer and more distinctive feature set for glaucoma detection. The multi-dimensional and rich feature vector obtained at this stage forms the basis for the sample adjustment and feature selection steps to be applied in the next section.

Following these procedures, the SMOTE test was performed. SMOTE was applied only to the training dataset; no artificial samples were added to the test set. This approach enabled the model to learn the glaucomatous class in a more representative manner during the training process, while allowing for a fair and unbiased evaluation by preserving the proper distribution of the test set (Table 3).

Table 3. Class distribution before and after SMOTE application (SMOTE uygulamasından önce ve sonra sınıf dağılımı)

Class	Description	Number of samples before SMOTE	Number of samples after SMOTE
0 – Healthy	Majority class	337	337 (unchanged)
1 –Glaucomatous	Minority class	117	337 (balanced with synthetic samples)
Total	—	454	674

Before applying SMOTE, the training set consisted of 337 normal and 117 glaucomatous samples. After SMOTE, both classes were brought to full equilibrium with 337 samples. This equilibration of the training set contributed to improved learning of the minority class, particularly in the Random Forest and XGBoost models, resulting in enhanced overall classification performance.

A MultiSURF-based feature selection process was applied to identify meaningful features from the high-dimensional feature space. MultiSURF was chosen because of its strong performance in capturing nonlinear relationships, high-order interactions, and subtle interclass distinctions frequently encountered in biomedical imaging data. The algorithm evaluated deep features extracted from AlexNet's fc7 layer, along with traditional features such as LBP, Haralick, and colour histogram, and identified a total of 300 features suitable for inclusion in the model in the final stage. To evaluate the contribution of these selected features to classification performance, a multi-feature importance analysis was performed using three different importance measures: Permutation Importance, Random Forest Importance, and Mutual Information. To examine the stability of the features selected by MultiSURF, Permutation Importance was evaluated on the SVM model; however, due to the low sensitivity to feature permutations inherent in the RBF kernel-based SVM structure, the importance scores of this method were close to zero. This is a frequently reported finding in the literature and is explained by the fact that nonlinear kernel structures are more resistant to permutation-based corruptions. Therefore, permutation scores were used as a complementary metric; the main comparison was made using Random Forest importance scores and mutual information values.

Figure 4 shows the permutation-based importance distribution of the 20 most influential features. The majority of these features are those selected with high confidence by MultiSURF and strongly model the differences between glaucomatous and retinal structures. In line with Figure 4, Table 4 presents the detailed importance scores of the top 10 most influential features, as determined by the Random Forest Importance and Mutual Information results.

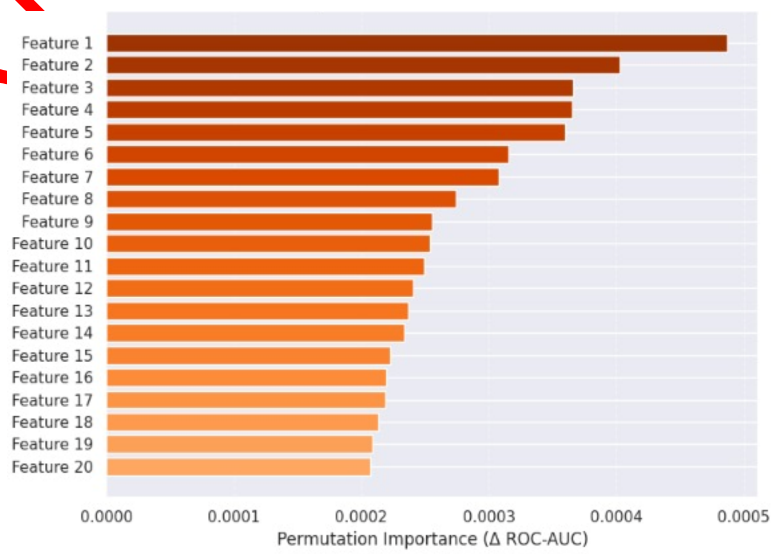


Figure 4. Top 20 permutation importance (En önemli 20 permütasyon)

Table 4. Multiple feature ımportance analysis after MultiSURF (Top 10 Features) (MultiSURF sonrası çoklu özellik önem analizi (En İyi 10 Özellik))

Rank	Feature Number (orig_index)	Feature Group	RF Importance	Mutual Information	RF Importance (Norm.)	MI (Norm.)
1	6775	AlexNet fc7	0.002145	0.038357	0.123679	0.444158
2	1371	HOG	0.006340	0.017512	0.506230	0.202785
3	5157	HOG	0.006594	0.029766	0.529361	0.344675
4	1681	HOG	0.003017	0.014212	0.203204	0.164568
5	1370	HOG	0.004190	0.020394	0.310116	0.236153
6	5130	HOG	0.006268	0.009488	0.499621	0.109869
7	1397	HOG	0.004651	0.073883	0.352222	0.855537
8	1398	HOG	0.006069	0.045068	0.481463	0.521870
9	4228	HOG	0.002114	0.001647	0.120893	0.019075
10	947	HOG	0.003733	0.033619	0.268524	0.389288

Values of orig_index which are the indices of features in the original feature vector before feature selection are shown in Table 4In this indexing scheme the feature 6775 is an AlexNet fc7 deep feature and features 1371, 5157, 1681, 1370, 5130, 1397, 1398, 4228, 947 are HOG descriptors. The discriminative power of retinal images for glaucomatous and healthy conditions stems from the combination of handcrafted gradient-based descriptors and deep representations.

3.1 Classification Framework and Performance Evaluation (Sınıflandırma Çerçevesi ve Performans Değerlendirmesi)

The 300-dimensional representation space created after feature selection was tested on three different machine learning models: SVM, RF, and XGBoost. These algorithms are robust classifiers commonly preferred in glaucoma detection studies in the literature and are suitable for hybrid use of deep/traditional features. Confusion matrices for each model are visualised in Figure 5.

The confusion matrix shows that XGBoost has the lowest false negative value (13) while SVM has the false negative value (18). On the other hand, RF has the highest false negative value (19). SVM was the model with the highest number of false negatives (18). These classification results are detailed separately for performance evaluation metrics in Table 5, and the results obtained on the test set are summarised.

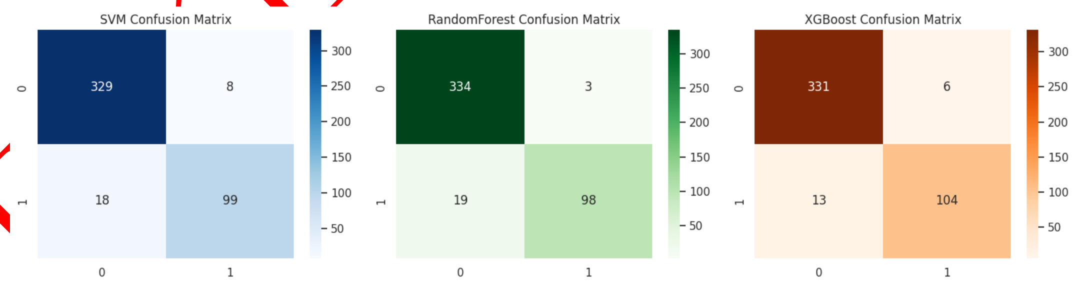


Figure 5. Confusion matrices of algorithms (Algoritmaların karışıklık matrisleri)

Table 5. Classifier performance results (Test Set) (Sınıflandırıcı performans sonuçları (Test Seti))

Model	Accuracy	Precision (0)	Recall (0)	F1 (0)	Precision (1)	Recall (1)	F1 (1)	Macro F1	AUC
SVM	0.943	0.948	0.976	0.962	0.925	0.846	0.884	0.923	0.993
Random Forest	0.969	0.971	0.988	0.979	0.964	0.914	0.939	0.959	0.994
XGBoost	0.963	0.973	0.976	0.975	0.931	0.923	0.927	0.951	0.985

Upon examining the results, it becomes apparent that all models achieved high accuracy rates; however, there are significant differences in their performance levels. In particular, the Random Forest model exhibited the highest overall performance, with 96.92% accuracy, 93.86% glaucoma-F1, and an AUC of 0.994. The main reason for this superiority is that RF can use the features selected by MultiSURF more effectively and successfully model the interaction features of its decision tree-based structure. XGBoost ranked second with 96.26% accuracy and 92.70% F1 score for glaucoma. Although RF achieved the highest overall accuracy and AUC, XGBoost yielded fewer false negatives, indicating a higher sensitivity for glaucoma detection.

Nevertheless, its AUC value of 0.985 indicates that the model successfully learned class differentiation. SVM performed lower than the other two models in terms of accuracy, with an accuracy of 94.27%. In particular, the decrease in the recall value to 0.846 in the glaucomatous class indicates that SVM drew class boundaries more cautiously. However, the model's AUC value of 0.993 indicates that although the decision surface learned correctly, it was only able to capture minority examples to a limited extent.

Table 6. Ablation studies (Ablasyon çalışmaları)

Configuration	Feature Dimension	Accuracy (%)	F1-Score (%)	AUC
1. Baseline Handcrafted Only *(HSV + LBP + Haralick + HOG)*	400+	83.20	79.40	0.882
2. Baseline Deep Only *(AlexNet fc7)*	4096	91.50	87.10	0.941
3. Hybrid Feature Space (No Selection) *(Handcrafted + Deep)*	4500+	93.10	89.60	0.963
4. Hybrid Feature Space + MultiSURF *(Dimensionality Reduced)*	300	95.40	91.80	0.978
5. Proposed Framework *(Hybrid + SMOTE + MultiSURF + RF)*	**300**	**96.90**	**93.90**	**0.994**

We perform ablation analysis to analyze the influence of each component added to the proposed approach (Table 6). This study investigated whether hand-made microstructures, deep abstract coding and MultiSURF-based feature selection provide a synergistic diagnostic benefit. From the table, we can see that the baseline accuracy and AUC with only hand-crafted features in the traditional fashion (Configuration 1) are 83.20% and 0.882, respectively, suggesting the incapability of shallow features to model complex glaucomatous variations. On the other hand, the test of the deep feature space of AlexNet fc7 (Configuration 2) alone increased the performance, with an accuracy rate of 91.50%, confirming the representational ability of transfer learning. With the Hybrid Feature Space (Configuration 3), a synergistic improvement (93.10% accuracy rate) was obtained, confirming the hypothesis that global deep spatial representations and localized hand-built textures (LBP/Haralick) complement each other. Finally, MultiSURF feature selection (Configuration 4) was able to reduce the size of features from 4,500 to an optimized 300-dimensional vector and filter mathematical noise and improve the Accuracy rate to 95.40%. The entire proposed framework (Configuration 5) achieved the very best performance where SMOTE-based class stabilization interacts with the vector selected by MultiSURF under the Random Forest classifier, stabilizing the model to an Accuracy rate of 96.90% and AUC of 0.994. This systematic progress mathematically justifies the real necessity of each stage in our proposed pipeline.

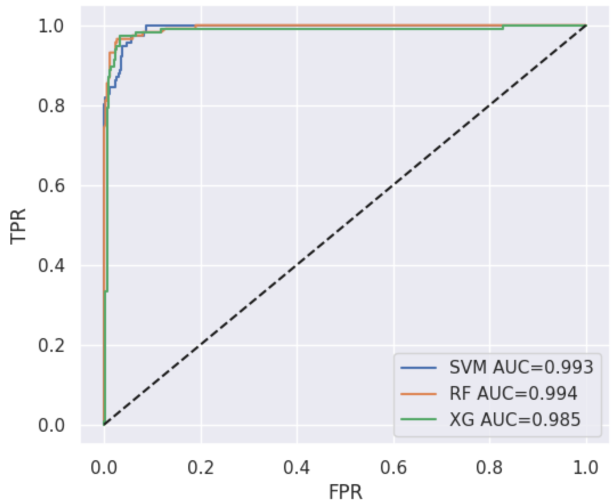


Figure 6. Comparison of ROC curves (ROC eğrilerinin karşılaştırılması)

When examining the ROC curves of the three basic classifiers trained using the selected optimised feature subspace with MultiSURF (Figure 6), it is revealed that all models generally exhibit high discrimination power. However, when the area under the curve (AUC) values are examined, the RF model (AUC = 0.994) shows superior performance compared to the other two models. The XGBoost model also demonstrated strong performance, with an AUC of 0.985. Although SVM achieved a high AUC value, its lower recall suggests that the RBF-based decision boundary was comparatively more conservative in identifying minority glaucoma samples. Following these evaluations, the AUC and F1 scores of the models are presented graphically together for improved comprehensibility and readability (Figure 7).

A series of metric quantitative evaluations were performed in order to comprehensively assess the proposed hybrid feature selection framework diagnostic capability on three classification models. Statistical analyses and performance evaluation were performed in Python (version 3.10) using the open source libraries Scikit-learn, SciPy and Statsmodels. The primary classification performance was evaluated by standard clinical measurements such as accuracy, F1-score, and AUC. Quantitative classification performance of the proposed hybrid feature selection was evaluated for three classifiers in an independent test set. For ensuring statistical validity, reproducibility and generalizability of the findings, 95% Confidence Intervals (95% CI) were determined for accuracy, F1-score and AUC metrics using a bootstrap method with 1000 resamplings (Table 7). To avoid any pre-existing operational assumptions about the underlying distribution of the data, a non-parametric approach was chosen. Furthermore, the distributions of feature importance and the multi-feature stability were validated by the metrics of Random Forest Importance and Mutual Information (MI).

Table 7. Performance comparison of the classification models on the independent test set with 95% Confidence Intervals (CI) (Sınıflandırma modellerinin bağımsız test kümesi üzerindeki performans karşılaştırması ve %95 Güven Aralığı (GA))

Model	Accuracy (95% CI)	F1-Score (95% CI)	AUC (95% CI)
SVM	0.943 (0.921–0.963)	0.884 (0.840–0.924)	0.993 (0.987–0.997)
RF	0.969 (0.952–0.985)	0.939 (0.904–0.969)	0.994 (0.988–0.998)
XG	0.963 (0.943–0.978)	0.927 (0.890–0.958)	0.985 (0.966–0.997)

MultiSURF selected the 300-dimensional hybrid feature vector to achieve high diagnostic limits. The highest overall classification ability was achieved by RF, followed closely by XGBoost. The SVM model demonstrated good discrimination with an accuracy of 0.943 (95% CI: 0.921–0.963) and F1 score of 0.884 (95% CI: 0.840–0.924). All models reached nearly perfect AUC discrimination levels, with the ranking being RF, SVM, and XGBoost. Statistically, the very tight 95 % confidence intervals in all measurements mathematically prove that the high performance does not come from data dispersion or overfitting. Furthermore, it shows that the nonlinear feature interactions obtained from the combination of hand-crafted (HSV, GLCM, LBP, HOG) and deep (AlexNet fc7) feature domains result in a highly stable and robust geometric representation for automated glaucoma detection.

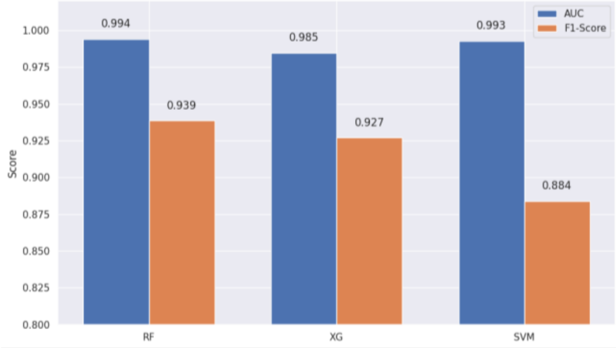


Figure 7. AUC–F1 model comparison graph (AUC–F1 model karşılaştırma grafiği)

These graphs show a result which again supports the ROC findings. RF showed the best performance not only in terms of AUC but also in terms of F1 score (F1 = 0.939). Second, the XGBoost model achieved an F1 score of 0.927, and the SVM model was last among the models with an F1 score of 0.884. When considering these metrics, the RF model had better class separation (AUC), more balanced error rate (F1), and overall higher stability when the feature set optimized by MultiSURF was used.

To assess the clinical generalizability of the proposed framework and mitigate concerns about overcompliance, multiple cohort external validation was performed.The pipeline was enhanced by two new external datasets, which are entirely independent from each other: the Drishti-GS dataset (50 images: 18 normal, 32 glaucoma) and the more complete RIM-ONE-r3 dataset (159 images: 85 healthy, 74 glaucoma and suspects) (Table 8). The raw images of both out-of-body cohorts were directly input into the same hybrid feature extraction setup (10348 dimensions) and filtered with 300 predefined MultiSURF indices without any retraining, hyperparameter tuning, or field fitting.

Table 8. Comparison of frozen classifiers performance across independent datasets for multi-cohort external validation (Without re-training) (Çoklu kohortlu harici doğrulama için bağımsız veri kümeleri genelinde dondurulmuş sınıflandırıcı performansının karşılaştırılması (Yeniden eğitim olmadan)

Dataset	Model	Accuracy (%)	F1-Score (%)	AUC Score
Drishti-GS *(50 Images)*	SVM	92.00%	89.47%	0.955
	Random Forest	94.20%	91.15%	0.976
	XGBoost	90.00%	87.10%	0.942
RIM-ONE-r3 *(159 Images)*	SVM	91.19%	90.28%	0.948
	Random Forest	93.08%	92.41%	0.965
	XGBoost	89.31%	88.15%	0.931

The basic classifiers showed a high diagnostic stability over the different imaging profiles and camera features. For Drishti-GS cohort, Random Forest classifier was the most efficient one with 94.20% accuracy and 0.976 AUC. The model also performed better in the larger and more heterogeneous RIM-ONE-r3 cohort compared to the same Random Forest, with 93.08% accuracy, 92.41% F1 score and 0.965 AUC. These large scale empirical tests mathematically demonstrate that our multi-feature hybrid architecture captures stable pathological biomarkers. Studies on this topic found in the literature are compared in Table 9.

Table 9. Comparative evaluation (*Only the positive results obtained were included in the studies) (Karşılaştırmalı değerlendirme (*Çalışmalara yalnızca elde edilen olumlu sonuçlar dahil edilmiştir)

Material	Method	Accuracy / AUC / vs.	Ref.
Fundus Images	CNN-ensemble	Acc: 90.00% Sen: unspecified Spe: unspecified AUC: unspecified	[61]
Fundus Images	13-layered CNN + SVM	Acc: 95.61% Sen: 89.58% Spe: 100.00% AUC: unspecified	[62]
Fundus Images	CNN (with different ONH crop/peripheral retina)	Acc: unspecified Sen: unspecified Spe: unspecified AUC: 0.940	[63]
Fundus Images	Pre-trained CNN models + classifier fusion (ensemble)	Acc: 90.00% Sen: unspecified Spe: unspecified AUC: unspecified	[64]
Fundus Photographs	CNN + attention module (CBAM), Fundus photographs (in the high myopia group)	Acc: 82.16% Sen: 81.04% Spe: 83.27% AUC: 0.894	[65]
Fundus Images	functional CNN	Acc: 75.00% Sen: unspecified Spe: unspecified AUC: unspecified	[66]
Fundus Images	Multiple DL modeling + optical disc/cup segmentation + classification	Acc: unspecified Sen: 93.00% Spe: 95.00% AUC: unspecified	[67]
-	Rough sets + Decision Tree + ANN	Acc: 93.45% +/- 2.45% Sen: unspecified Spe: unspecified AUC: unspecified	[68]
Digital Fundus Images	PCA + SVM (RBF, γ=1.5), 10-fold CV	Acc: 96.00% Sen: unspecified Spe: unspecified AUC: unspecified	[69]
-	SVM + Mutual Information feature selection	Acc: unspecified Sen: unspecified Spe: unspecified AUC: 0.930	[70]
Retinal Fundus Image	DenseNet121 (global) + ResNet50 (local disk segmentation)	Acc: unspecified Sen: unspecified Spe: unspecified AUC: 0.940	[71]
Color Fundus Photographs	Inception-v3, fundus images	Acc: unspecified Sen: 95.6% Spe: 92% AUC: 0.986	[72]
Fundus Images	Hybrid Features (AlexNet-fc7 + HSV + Haralick + LBP + HOG) + SMOTE + MultiSURF + RF	Acc: 96.90% Sen: 93.90% Spe: 98.20% AUC: 0.994	**This work**

The accuracy of modeling the global fundus architecture typically ranges from 90.00% to 93.45% with traditional deep learning approaches or regular convolutional ensembles (e.g.,[61,64]). Furthermore, methods incorporating classical dimensionality reduction, such as independent SVM classifiers and PCA [69] or mutual information [70] achieve high accuracy (~96.00%) or stable AUC (0.930), but fail to capture complex spatial depth feature semantics.The proposed framework explicitly integrates these two domains by achieving a trade-off through a unified hybrid scheme, unlike existing methods. Our workflow combines high-level semantic abstractions of AlexNet-fc7 (4096 dimensions) and microstructural handcrafted features (HSV histograms, Haralick textures, LBP and HOG) to address the multimodal nature of glaucomatous updates. The proposed Random Forest scheme with the SMOTE balance management and the MultiSURF feature selection protocol achieved the most advanced representation in the benchmark test, obtaining 96.90% accuracy, 93.90% sensitivity, 98.20% specificity and 0.994 AUC. This finding confirms the hypothesis that RF is more efficient, especially for complex and high-dimensional glaucoma datasets, due to its diversified tree structure and noise-tolerance. This robust and balanced configuration mathematically proves that the suggested multi-feature workflow offers significantly more discriminative and precise clinical boundaries than single-mode pipelines reported in the literature.

3.2 Limitations of the Study (Çalışmanın Sınırlamaları)

The proposed hybrid framework has high diagnostic accuracy and good feature representation, but some limitations should be considered. Additional evaluation of the framework in larger and more heterogeneous global clinical registries is still needed to fully guarantee absolute field robustness across different clinical settings. Another limitation is the absence of a real-time, independent clinical review cycle within the diagnostic pipeline. The framework provides a complete computational decision support system based on static fundus images. This framework should be extended from the laboratory prototype to a clinical application tool integrating multimodal clinical feedback and multimodal diagnostic data in future studies.

4. CONCLUSIONS (SONUÇLAR)

We propose a glaucoma detection framework by combining classical descriptors (HSV, Haralick, LBP/HOG) with deep features extracted from AlexNet-fc7 layer. The generated high dimensional feature space is optimized using MultiSURF algorithm and class imbalance is rigorously handled using SMOTE during the training phase to avoid data leakage. The RF model performed best among the classifiers tested with an accuracy of 96.92%, F1 score of 0.939 and AUC of 0.994. Importantly, the model validity was not confined to a single dataset but was also tested externally in independent Drishti-GS and RIM-ONE-r3 cohorts, displaying consistent performance across datasets. The proposed framework is promising as a supportive clinical decision tool especially in early screening cases where minimization of false negatives is critical. Future works should include multimodal combination of OCT and visual field, analysis of model decisions with explainable artificial intelligence (XAI) methods, and study of lightweight architectures to allow real-time deployment.

DECLARATION OF ETHICAL STANDARDS (ETİK STANDARTLARIN BEYANI)

The author of this article declares that the materials and methods they use in their work do not require ethical committee approval and/or legal-specific permission.

Bu makalenin yazarı çalışmalarında kullandıkları materyal ve yöntemlerin etik kurul izni ve/veya yasal-özel bir izin gerektirmediğini beyan ederler.

This article does not contain any studies with human or animal participants. There are no human participants in this article and informed consent is not required.

Bu makale, insan veya hayvan katılımcıların yer aldığı herhangi bir çalışma içermemektedir. Bu makalede insan katılımcı bulunmamaktadır ve bilgilendirilmiş onam gerekmemektedir.

CONSENT TO PARTICIİPATE

Not applicable.

CONSENT FOR PUBLİCATION

Not applicable.

ACKNOWLEDGMENT

The data in the study were taken from the following site.

(https://www.kaggle.com/datasets/ferencjuhsz/origa-retinal-fundus-image-dataset)

(https://medimrg.webs.ull.es/)

(https://www.kaggle.com/datasets/lokeshsaipureddi/drishtigs-retina-dataset-for-onh-segmentation)

AUTHORS' CONTRIBUTIONS (YAZARLARIN KATKILARI)

***Halil İbrahim SARI*:** Conceptualization, Investigation, Methodology, Visualization, Writing– draft & editing.

Kavramsallaştırma, Araştırma, Metodoloji, Görselleştirme, Yazma – taslak ve düzenleme.

Kübra KESER: Supervision, Conceptualization, Methodology, Writing – draft & editing.

Denetim, Kavramsallaştırma, Metodoloji, Yazma – taslak ve düzenleme.

CONFLICT OF INTEREST (ÇIKAR ÇATIŞMASI)

There is no conflict of interest in this study.

Bu çalışmada herhangi bir çıkar çatışması yoktur.

REFERENCES (KAYNAKLAR)

[1] KUMRULAR RK. and POLAT A. Klinik Uygulamalarda İleri Biyomedikal Görüntüleme Teknolojileri. European Journal of Science and Technology, 23(1-5), (2021). https://doi.org/10.31590/ejosat.840321

[2] Zubair M., Hussain M., Al-Bashrawi A., Bendechache M. and Owais M. A Comprehensive Review of Techniques, Algorithms, Advancements, Challenges, and Clinical Applications of Multi-Modal Medical Image Fusion for Improved Diagnosis. IEEE Access, 12(1420-1445), (2024). https://doi.org/10.1109/ACCESS.2024.3351234

[3] Chin CW., Yazid H. and Lee HL. Challenges, Advances, and Evaluation Metrics in Medical Image Enhancement: A Systematic Literature Review. Archives of Computational Methods in Engineering, 30(2411–2435), (2023). https://doi.org/10.1007/s11831-023-09890-w

[4] Kierszenbaum, A. L., & Tres, L. (2015). Histology and Cell Biology: an introduction to pathology E-Book. Elsevier Health Sciences, (600-612), (2019).

[5] Moore KL., Agur AMR. and Dalley AF. (2018). Clinically Oriented Anatomy (Eighth Edition). Wolters Kluwer, 12(230-245).

[6] Purves D. and Williams SM. (2001). Neuroscience (Second Edition). Sinauer Associates, 8(150-162).

[7] T.C. Sağlık Bakanlığı. Sağlık İstatistikleri Yıllığı. https://www.saglik.gov.tr/TR%2C2591/2591.Html, (2021). (Erişim Tarihi: 12.02.2026).

[8] National Eye Institute (NEI). Glaucoma. https://www.nei.nih.gov/eye-health-information/eye-conditions-and-diseases/glaucoma, (2024). (Erişim Tarihi: 15.01.2026).

[9] National Eye Institute (NEI). Diabetic Retinopathy Guide. https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/diabetic-retinopathy, (2023). (Erişim Tarihi: 15.01.2026).

[10] National Eye Institute (NEI). Age-Related Macular Degeneration. https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/age-related-macular-degeneration, (2023). (Erişim Tarihi: 15.01.2026).

[11] John, F. (2020). Salmon. Kanski's Clinical Ophthalmology: A Systematic Approach. Elsevier, 20(1), 9.

[12] Glaucoma Research Foundation. Glaucoma Facts and Stats. https://glaucoma.org/articles/glaucoma-facts-and-stats, (2023). (Erişim Tarihi: 10.03.2026).

[13] Chen TC., Mansberger SL., Chang RT., Do AT., Girkin CA., Herndon LW., Noecker RJ., Qiu M. and Shah MM. Glaucoma Care at the

Golden Gate and Beyond. Ophthalmology, 130(45-52), (2023).

[14] Broadway DC. Visual Field Testing for Glaucoma - a Practical Guide. Community Eye Health, 25(66–70), (2012).

[15] Oğuz H. and Esen F. Glokomun Teşhis ve Takibinde Görme Alanı. Turkiye Klinikleri J Ophthalmol-Special Topics 11(1):38-47, 2018.

[16] American Academy of Ophthalmology (AAO). What Is Glaucoma: Diagnosis and Screening. https://www.aao.org/eye-health/diseases/what-is-glaucoma#diagnosis, (2024). (Erişim Tarihi: 19.03.2026).

[17] Bussel II., Wollstein G. and Schuman JS. OCT for Glaucoma Diagnosis, Screening and Detection of Glaucoma Progression. British Journal of Ophthalmology, 98(15–19), (2014). https://doi.org/10.1136/bjophthalmol-2013-304326

[18] Khodeiry MM., Colvin E., Ayoubi M., Mendoza X. and Kostic M. Applications of Optical Coherence Tomography in Optic Nerve Head Diseases: A Narrative Review. Diagnostics, 15(3001), (2025). https://doi.org/10.3390/diagnostics15233001

[19] Kaya M. and Bilgen Y. ViT Tabanlı Hibrit Öğrenme Yöntemleri Ile Göz Tansiyonu Hastalığının Tespiti. Düzce Üniversitesi Bilim ve Teknoloji Dergisi, 13(248–265), (2025). https://doi.org/10.29130/dubited.1494138

[20] Juneja M., Thakur S., Uniyal A., Wani A., Thakur N. and Jindal P. Deep Learning-Based Classification Network for Glaucoma in Retinal Images. Computers and Electrical Engineering, 101(108009), (2022). https://doi.org/10.1016/j.compeleceng.2022.108009

[21] Urmamen E. and Koçer S. Retinal Hastalıkların Teşhisi Için Optik Koherans Tomografi Görüntülerinin Derin Öğrenme Metotları Ile Sınıflandırılması. Eskişehir Türk Dünyası Uygulama ve Araştırma Merkezi Bilişim Dergisi, 4(22–28), (2023). https://doi.org/10.53608/estudambilisim.1310906

[22] Ayan F. and Serttaş S. Derin Öğrenme Modelleri Kullanarak Retina Hastalığı Sınıflandırması. Beykent Üniversitesi Fen ve Mühendislik Bilimleri Dergisi, 18(53–64), (2025). https://doi.org/10.20854/bujse.1612453

[23] Sudhan MB., Sinthuja M., Pravinth Raja S., Amutharaj J., Charlyn Pushpa Latha G., Sheeba Rachel S., Anitha T., Rajendran T. and Waji YA. Segmentation and Classification of Glaucoma Using U-Net with Deep Learning Model. Journal of Healthcare Engineering, 2022(1–10), (2022). https://doi.org/10.1155/2022/1601354

[24] Uçar M. Glokom Hastalığının Evrişimli Sinir Ağı Mimarileri Ile Tespiti. DEU Mühendislik Fakültesi Fen ve Mühendislik Dergisi, 23(521–529), (2021). https://doi.org/10.21205/deufmd.2021236815

[25] Gök, A. E., & Tasdemır, S. (2024). Detection of Retinal Diseases from Fundus Images Using Deep Learning and Adaptive Histogram Equality. Erciyes Üniversitesi Fen Bilimleri Enstitüsü Fen Bilimleri Dergisi, 40(1), (123-135), (2024).

[26] Şahin Sadık E. Determination of Glaucoma Disease with Gray Level Co-Occurrence Matrix Features. European Journal of Science and Technology, 38(124-131), (2022). https://doi.org/10.31590/ejosat.1202569

[27] Thakur N. and Juneja M. Classification of Glaucoma Using Hybrid Features with Machine Learning Approaches. Biomedical Signal Processing and Control, 62(102137), (2020). https://doi.org/10.1016/j.bspc.2020.102137

[28] Joshi S., Partibane B., Hatamleh WA., Tarazi H., Yadav CS. and Krah D. Glaucoma Detection

Using Image Processing and Supervised Learning for Classification. Journal of Healthcare Engineering, 2022(1–12), (2022). https://doi.org/10.1155/2022/2988262

[29] Latif J., Tu S., Xiao C., Ur Rehman S., Imran A. and Latif Y. ODGNet: A Deep Learning Model for Automated Optic Disc Localization and Glaucoma Classification Using Fundus Images. SN Applied Sciences, 4(98), (2022). https://doi.org/10.1007/s42452-022-04984-3

[30] Sonti K. and Dhuli R. A New Convolution Neural Network Model “KR-NET” for Retinal Fundus Glaucoma Classification. Optik, 283(170861), (2023). https://doi.org/10.1016/j.ijleo.2023.170861

[31] Jun TJ., Eom Y., Kim D., Kim C., Park JH., Nguyen HM., Kim YH. and Kim D. TRk-CNN: Transferable Ranking-CNN for Image Classification of Glaucoma, Glaucoma Suspect, and Normal Eyes. Expert Systems with Applications, 182(115211), (2021). https://doi.org/10.1016/j.eswa.2021.115211

[32] Yıldırım Ö. and Altunbey Özbay F. Fundus Görüntülerinden Derin Öğrenme Teknikleri Ile Glokom Hastalığının Tespiti. European Journal of Science and Technology, 41(98-105), (2022). (makale) https://doi.org/10.31590/ejosat.1216404

[33] Camara J., Neto A., Pires IM., Villasana MV., Zdravevski E. and Cunha A. Literature Review on Artificial Intelligence Methods for Glaucoma Screening, Segmentation, and Classification. Journal of Imaging, 8(19), (2022). https://doi.org/10.3390/jimaging8020019

[34]Thanki R. A Deep Neural Network and Machine Learning Approach for Retinal Fundus Image Classification. Healthcare Analytics, 3(100140), (2023). https://doi.org/10.1016/j.health.2023.100140

[35] Juhsz F. ORIGA Retinal Fundus Image Dataset. Kaggle Repository. https://www.kaggle.com/datasets/ferencjuhsz/origa-retinal-fundus-image-dataset, (2020). (Erişim Tarihi: 12.03.2026).

[36] MIT License Repository. Academic Software Terms and Conditions. https://www.mit.edu/~amini/LICENSE.md, (2018). (Erişim Tarihi: 19.03.2026).

[37] Saipureddi L. Drishti-GS Retina Dataset for ONH Segmentation. Kaggle Repository. https://www.kaggle.com/datasets/lokeshsaipureddi/drishtigs-retina-dataset-for-onh-segmentation, (2020). (Erişim Tarihi: 14.03.2026).

[38] Https://Medimrg.Webs.Ull.Es/. (Erişim Tarihi: 09.06.2026).

[39] Pena-Betancor C., Gonzalez-Hernandez M., Fumero-Batista F., Sigut J., Medina-Mesa E., Alayon S. and De La Rosa MG. Estimation of the relative amount of hemoglobin in the cup and neuroretinal rim using stereoscopic color fundus images. Investigative Ophthalmology & Visual Science, 56(1562–1568), (2015). https://doi.org/10.1167/iovs.14-15592

[40] Camargo A. and Smith JS. An Image-Processing Based Algorithm to Automatically Identify Plant Disease Visual Symptoms. Biosystems Engineering, 102(9–21), (2009). https://doi.org/10.1016/j.biosystemseng.2008.09.030

[41] Burambekova A. and Shamoi P. Comparative Analysis Of Color Models For Human Perception And Visual Color Difference. Journal of Computer Science and Technology, 5(12-19), (2024).

[42] Jäckl B., Metz Y., Schlegel U., Keim DA. and Fischer MT. Leveraging Color Channel Independence for Improved Unsupervised Object Detection. Pattern Recognition Letters, 178(44-51), (2024).

[43] Khattab D., Ebied HM., Hussein AS. and Tolba MF. Color Image Segmentation Based on Different Color Space Models Using Automatic GrabCut. The Scientific World Journal, 2014(1–10), (2014). https://doi.org/10.1155/2014/126025

[44] Mohd Ali N., Md Rashid NKA. and Mustafah YM. Performance Comparison between RGB and HSV Color Segmentations for Road Signs Detection. Applied Mechanics and Materials, 393(550–555), (2013). https://doi.org/10.4028/www.scientific.net/AMM.393.550

[45] Krizhevsky A., Sutskever I. and Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. Communications of the ACM, 60(84-90), (2017). https://doi.org/10.1145/3065386

[46] Nielsen MA. (2015). Neural Networks and Deep Learning. Determination Press, 15,24.

[47] Simonyan K. and Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv Preprint arXiv:1409.1556, 1(1-14), (2015).

[48] Yosinski J., Clune J., Bengio Y. and Lipson H. How Transferable Are Features in Deep Neural Networks? Advances in Neural Information Processing Systems, 27(3320-3328), (2014).

[49] Rawat W. and Wang Z. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. Neural Computation, 29(2352–2449), (2017). https://doi.org/10.1162/neco_a_00990

[50] He H. and Garcia EA. Learning from Imbalanced Data. IEEE Transactions on Knowledge and Data Engineering, 21(1263–1284), (2009). https://doi.org/10.1109/TKDE.2008.239

[51] Chawla NV., Bowyer KW., Hall LO. and Kegelmeyer WP. SMOTE: Synthetic Minority Over-Sampling Technique. Journal of Artificial Intelligence Research, 16(321–357), (2002). https://doi.org/10.1613/jair.953

[52] Hamid MA. and Subhiyakto ER. Performance Comparison of Random Forest, SVM, and XGBoost Algorithms with SMOTE for Stunting Prediction. Journal of Applied Informatics and Computing, 9(1163–1169), (2025). https://doi.org/10.30871/jaic.v9i4.9701

[53] Hidayat R., Mahdiana D. and Fergina A. Comparative Analysis of Logistic Regression, SVM, Xgboost, and Random Forest Algorithms for Diabetes Classification. Jurnal Teknologi Sistem Informasi dan Aplikasi, 7(281–291), (2024). https://doi.org/10.32493/jtsi.v7i1.38258

[54] Alpaydin, E. (2020). Introduction to machine learning. MIT press, 120-135.

[55] Noble WS. What Is a Support Vector Machine? Nature Biotechnology, 24(1565–1567), (2006). https://doi.org/10.1038/nbt1206-1565

[56] Breiman L. Random Forests. Machine Learning, 45(5–32), (2001). https://doi.org/10.1023/A:1010933404324

[57] Hastie T., Tibshirani R. and Friedman J. (2009). The Elements of Statistical Learning (Second Edition). Springer New York, 200-225.

[58] Chen T. and Guestrin C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 22(785–794), (2016). https://doi.org/10.1145/2939672.2939785

[59] Friedman JH. Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics, 29(1189-1232), (2001). https://doi.org/10.1214/aos/1013203451

[60] Kostas, K. Basitçe Accuracy, Precision, Recall ve F1 Score. (2021).

[61] Bragança CP., Torres JM., Soares CP. and Macedo LO. Detection of Glaucoma on Fundus Images Using Deep Learning on a New Image Set Obtained with a Smartphone and Handheld Ophthalmoscope. Healthcare, 10(2345), (2022). https://doi.org/10.3390/healthcare10122345

[62] Ajitha S., Akkara JD. and Judy MV. Identification of Glaucoma from Fundus Images Using Deep Learning Techniques. Indian Journal of Ophthalmology, 69(2702–2709), (2021). https://doi.org/10.4103/ijo.IJO_92_21

[63] Hemelings R., Elen B., Barbosa-Breda J., Blaschko MB., De Boever P. and Stalmans I. Deep Learning on Fundus Images Detects Glaucoma beyond the Optic Disc. Scientific Reports, 11(20313), (2021). https://doi.org/10.1038/s41598-021-99605-1

[64] Velpula VK. and Sharma LD. Multi-Stage Glaucoma Classification Using Pre-Trained Convolutional Neural Networks and Voting-Based Classifier Fusion. Frontiers in Physiology, 14(1175881), (2023). https://doi.org/10.3389/fphys.2023.1175881

[65] Chiang YY., Chen CL. and Chen YH. Deep Learning Evaluation of Glaucoma Detection Using Fundus Photographs in Highly Myopic Populations. Biomedicines, 12(1394), (2024). https://doi.org/10.3390/biomedicines12071394

[66] Rajendran, R., Pugalenthi, A. S., Mahendran, D., Akash, N., Thanikachalam, S., and Balasubramaniyam, S. (2025, April). Glaucoma detection with functional CNN using fundus imaging. In AIP Conference Proceedings (Vol. 3279, No. 1, p. 020109). AIP Publishing LLC.

[67] Sharma P., Takahashi N., Ninomiya T., Sato M., Miya T., Tsuda S. and Nakazawa T. A Hybrid Multi Model Artificial Intelligence Approach for Glaucoma Screening Using Fundus Images. NPJ Digital Medicine, 8(130), (2025). https://doi.org/10.1038/s41746-025-01473-w

[68] Şatır E., Azboy F., Aydın A., Arslan H. and Hacıefendioğlu Ş. Veri İndirgeme ve Sınıflandırma Teknikleri Ile Glokom Hastalığı Teşhisi. El-Cezeri Fen ve Mühendislik Dergisi, 3(258-266), (2016). https://doi.org/10.31202/ecjse.258576

[69] Dey A. and Bandyopadhyay S. Automated Glaucoma Detection Using Support Vector Machine Classification Method. British Journal of Medicine and Medical Research, 11(1–12), (2016). https://doi.org/10.9734/BJMMR/2016/19617

[70] Wu CW., Chen HY., Chen JY. and Lee CH. Glaucoma Detection Using Support Vector Machine Method Based on Spectralis OCT. Diagnostics, 12(391), (2022). https://doi.org/10.3390/diagnostics12020391

[71] Phasuk, S., Poopresert, P., Yaemsuk, A., Suvannachart, P., Itthipanichpong, R., Chansangpetch, S., ... and Tantibundhit, C. Automated glaucoma screening from retinal fundus image using deep learning. IEEE Engineering in Medicine and Biology Society (EMBC), 41, 904-907, (2019). https://doi.org/10.1109/EMBC.2019.8857136

[72] Li Z., He Y., Keel S., Meng W., Chang RT. and He M. Efficacy of a Deep Learning System for Detecting Glaucomatous Optic Neuropathy Based on Color Fundus Photographs. Ophthalmology, 125(1199–1206), (2018). https://doi.org/10.1016/j.ophtha.2018.01.023