

Transfer Learning and Multi-Feature Fusion-Based Deep Learning Model for Idiopathic Macular Hole Diagnosis and Grading from Optical Coherence Tomography Images

Ye-Ting Lin¹, Xu Xiong¹, Ying-Ping Zheng², Qiong Zhou¹

¹Department of Ophthalmology, The First Affiliated Hospital, Jiangxi Medical College, Nanchang University, Nanchang, Jiangxi, People's Republic of China; ²Department of Product Design, Jiangxi Normal University, Nanchang, Jiangxi, People's Republic of China

Correspondence: Qiong Zhou, The First Affiliated Hospital of Nanchang University, 17 Yongwai Zheng Street, Donghu District, Nanchang, Jiangxi, People's Republic of China, Email qiongzd06@163.com

Background: Idiopathic macular hole is an ophthalmic disease that seriously affects vision, and its early diagnosis and treatment have important clinical significance to reduce the occurrence of blindness. At present, OCT is the gold standard for diagnosing this disease, but its application is limited due to the need for professional ophthalmologist to diagnose it. The introduction of artificial intelligence will break this situation and make its diagnosis efficient, and how to build an effective predictive model is the key to the problem, and more clinical trials are still needed to verify it.

Objective: This study aims to evaluate the role of deep learning systems in Idiopathic Macular Hole diagnosis, grading, and prediction.

Methods: A single-center, retrospective study used binocular OCT images from IMH patients at the First Affiliated Hospital of Nanchang University (November 2019 - January 2023). A deep learning algorithm, including traditional omics, Resnet101, and fusion models incorporating multi-feature fusion and transfer learning, was developed. Model performance was evaluated using accuracy and AUC. Logistic regression analyzed clinical factors, and a nomogram predicted surgical risk. Analysis was conducted with SPSS 22.0 and R 3.6.3. $P < 0.05$ was statistically significant.

Results: Among 229 OCT images, the traditional omics, Resnet101, and fusion models achieved accuracies of 93%, 94%, and 95%, respectively, in the training set. In the test set, the fusion model and Resnet101 correctly identified 39 images, while the traditional omics model identified 35. The nomogram had a C-index of 0.996, with macular hole diameter most strongly associated with surgical risk.

Conclusion: The deep learning system with transfer learning and multi-feature fusion effectively diagnoses and grades IMH from OCT images.

Keywords: transfer learning, deep learning, optical coherence tomography, idiopathic, macular hole, multi-feature, grading

Introduction

Idiopathic Macular Hole (IMH) refers to a discontinuity in the photoreceptor layer of the macula, with an incidence of approximately 0.1% in adults.¹⁻³ It is primarily caused by pathologic vitreomacular traction.^{4,5}

The diagnostic criteria for idiopathic macular holes have undergone significant evolution. While initial classification by Gass relied on biomicroscopic assessment of full-thickness defects and diameter measurements,⁶ the introduction of high-resolution optical coherence tomography (OCT) - with its remarkable 3.3 μm axial resolution - has established OCT as the current gold standard. This imaging modality provides histology-level visualization of retinal architecture, enabling precise morphological assessment of macular pathology. Altaweel et al⁷ proposed an OCT-derived grading system for IMH diagnosis. Clinical treatment protocols are determined based on the diagnostic classification established

through OCT examination.^{8–11} Furthermore, accumulating clinical trial evidence has led to continuous evolution in surgical approaches for IMH.¹²

However, OCT-based diagnostic and grading approaches present several inherent limitations: Firstly, the current approach relies heavily on empirical observation and demands significant expertise from experienced ophthalmologists. However, severe shortages of eye care specialists persist in many rural and underserved regions. With only approximately 40,000 registered ophthalmologists serving China's vast population, patients frequently experience delayed diagnosis and treatment. This critical gap in care access often leads to preventable vision deterioration and even blindness. Secondly, the current approach lacks standardized quantitative criteria, leading to significant variability in diagnostic accuracy among clinicians of differing experience levels. This inconsistency in assessment directly impacts both disease classification and subsequent treatment decisions. Particularly for small macular holes (<250µm), intravitreal pharmacologic therapy alone often achieves satisfactory anatomical closure rates.^{13,14}

DL technology holds significant advantages in the AI field, especially in medical image analysis. Transfer learning further enhances DL system performance under limited biomedical image conditions. DL systems excel in several aspects of image analysis: their core advantage lies in high accuracy, comparable to expert-level diagnostics across diabetic retinopathy,¹⁵ age-related macular degeneration,¹⁶ glaucoma,¹⁷ and cataract¹⁸ diagnostics. Highly automated, DL models minimize human intervention and subjective errors, enhance diagnostic consistency, and expedite processing speed.¹⁹ Moreover, DL models handle large datasets cost-effectively, reducing manpower and resource demands.²⁰

However, current literature lacks dedicated DL models specifically for diagnosing and grading Idiopathic Macular Hole (IMH), particularly in large-scale clinical applications.^{21–23} Existing studies predominantly employ conventional machine learning approaches, which demonstrate limited generalizability and fail to fully leverage multimodal imaging features, resulting in suboptimal performance metrics. Multimodal data integration combines heterogeneous data from various imaging modalities (OCT, fundus photos, FA) and sources (imaging, clinical, genomic data) to extract complementary features and improve model performance. The combination of DL, transfer learning, and multimodal data integration shows promise for optimizing ophthalmic image analysis, though its application to IMH remains to be clinically validated.

This study aims to fill this gap by developing a DL model based on transfer learning and multi-feature fusion for IMH diagnosis and classification. The following key questions are addressed: whether the diagnostic performance of the fusion model is improved compared with the traditional omics model and ResNet101; We compare its performance with traditional computational models and ResNet101. Visualization of predictive factors through nomograms²⁴ to assist in the formulation of surgical intervention strategies, providing clinicians with personalized decision-making criteria to predict the need for surgical intervention in Idiopathic Macular Hole.

Materials and Methods

Ethical Approval and Patient Selection

This retrospective study was approved by the Ethics Review Committee of First Affiliated Hospital of Nanchang University (Ethics No.(2023) CDYFYLYK(04–040). All procedures were carried out in accordance with the principles of the Declaration of Helsinki. We included patients with macular holes admitted to the Ophthalmology Department of First Affiliated Hospital of Nanchang University from November 2019 to January 2023. Patients with (1) A diagnosis of IMH was made and OCT examination was performed in the outpatient department of our hospital; (2) clearly visualized macular holes on OCT were included in this study. Patients with (1) Macular hole caused by high myopia; (2) received treatment prior to examination; and (3) had other ocular or systemic diseases were excluded. The research roadmap is shown in [Figure 1](#).

A total of 229 OCT images were included. All OCT images were obtained from the same machine, the Heidelberg Spectralis OCT(Heidelberg Engineering, Dossenheim, Germany), and were acquired by the same ophthalmologist. Two ophthalmologists reviewed these images and classified each macular hole into the following grades: grade 0 - without macular hole; grade 1 - impending macular hole without vitreomacular traction; grade 2 - impending macular hole with vitreomacular traction; grade 3 - small full-thickness macular hole (FTMH) <250 µm; grade 4 - medium FTMH >250 µm but <400 µm; and grade 5 - large FTMH >400 µm. Images that were classified as different grades were reviewed by a senior ophthalmologist. All OCT images were randomly assigned to the training and test sets with an 8:2 ratio. As such, 183 and 46 images were assigned

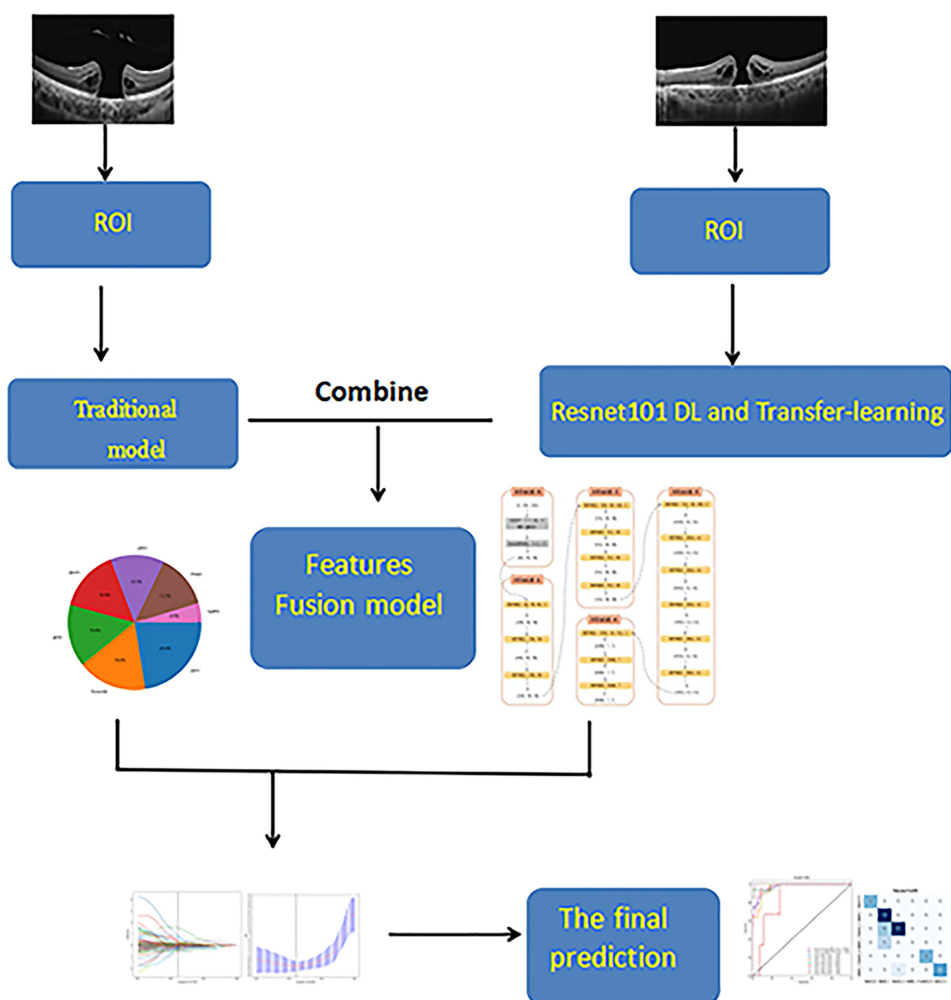


Figure 1 The flowchart of this study.

to the training and test sets, respectively. ITK-SNAP software (Version 3.8.0) was used to manually mark the region of interest (ROI) in the OCT images. The minimal extent of the full-layer macular hole was measured on OCT images, which was defined as “hole diameter” and the diameter of the hole base at the level of the retinal pigment epithelium, which was defined as “basal diameter”, and whether there was a post-vitreous detachment (PVD) was recorded, with 1 and 0, respectively.

Image Acquisition and Classification

Development and Preliminary Feature

In the present investigation, a deep learning (DL) framework has been meticulously engineered to prognosticate the grading of Idiopathic Macular Holes (IMH). Ensuring adherence to data privacy standards, all input variables were anonymized prior to analysis. The dataset was bifurcated into distinct cohorts for training and validation purposes, with the latter serving to evaluate the algorithm’s capability to discriminate among multiple treatment stratagems corresponding to the IMH severity. The predictive efficacy of the proposed DL architecture was benchmarked against conventional omics-based models as well as established convolutional neural network paradigms, specifically the Resnet101 architecture, to ascertain its relative performance in multi-classification tasks within the domain of ophthalmological interventions.

Extraction Advanced Feature Extraction Using

Utilizing the “PyRadiomics” package, an open-source Python library, a comprehensive extraction of radiomic features was performed. Specifically, the Python Software Foundation’s offering facilitated the delineation of seven distinct radiomic

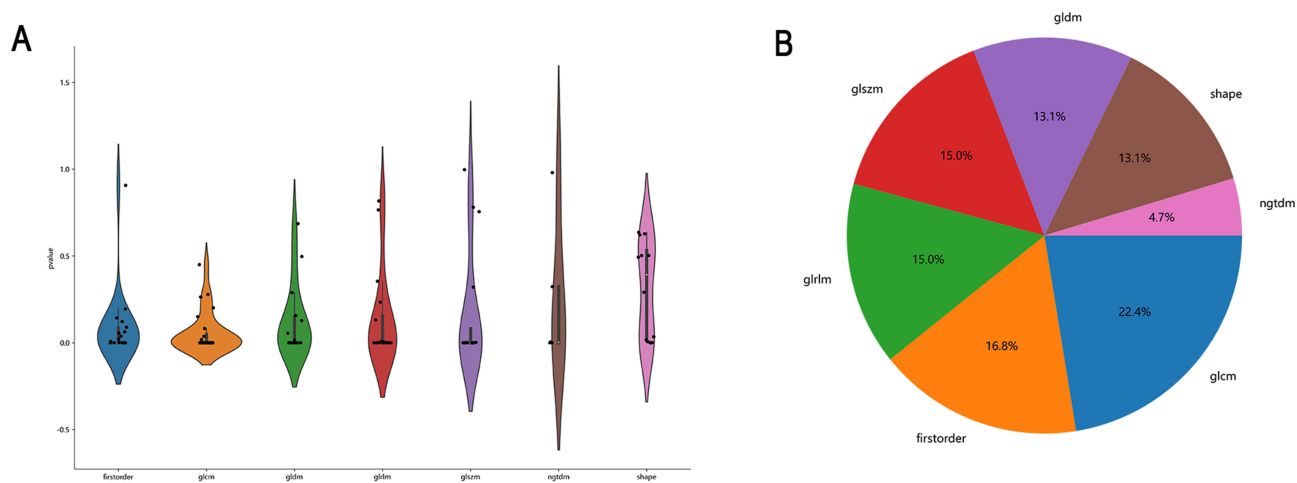


Figure 2 (A) Violin diagram of feature distribution extracted from traditional omics model; (B) Pie chart of feature distribution extracted from traditional omics model.

feature categories. These encompassed first-order statistics, two-dimensional morphological metrics, and higher-order textures, including the Gray-Level Co-Occurrence Matrix (GLCM), Gray-Level Run Length Matrix (GLRLM), Gray-Level Size Zone Matrix (GLSZM), Gray-Level Difference Matrix (GLDM), and Gray-Level Dependence Matrix (GLDM). For an in-depth elucidation of the individual features, interested parties are directed to consult the official PyRadiomics documentation, accessible at (<https://pyradiomics.readthedocs.io/en/latest/features.html>). The spatial distribution and central tendencies of the traditional omics-derived features are graphically represented in Figure 2.

PyRadiomicsResNet101 Model Adaptation and Feature Extraction

ResNet101²⁵ is an instantiation of the ResNet model which comprises 101 layers and utilizes a directed acyclic graph (DAG) structure for deep learning. This model is renowned for its efficacy in reducing error rates in image classification tasks, exemplified by its winning performance in the 2015 ImageNet Contest with a 3.6% error rate. The architecture of ResNet consists of an input stem that applies convolution with a large stride for dimensionality reduction, followed by four stages containing multiple residual learning blocks, each stage typically decreasing resolution and increasing width (channels). The output stem is adaptable to various tasks, marking ResNet101's flexibility and deep compositional capabilities. Firstly, the model was initialized with the parameters obtained from full training on the ImageNet²⁶ dataset. Then, due to the relatively small OCT image data set, a transfer learning program was used to train the pre-training model multiple times (100 epochs). The parameters were fine-tuned by the feedback of cross entropy loss function to improve network performance. The average probability of all images was used to produce DL or DTL features, DTL features means deep learning features for which transfer learning has been performed, and the output of the penultimate FC layer of the convolutional neural network (CNN) was used as DL or DTL features. The structure of convolutional neural network was implemented in Python.

Integrative Approach for Feature Fusion

In our study, features from the Resnet101 model were combined with those from a traditional omics model through a process known as feature fusion.²⁷ This method enhances the discriminative power of the resulting feature vector by improving the representation of image information.

LASSO Regression for Feature Refinement

In our study, computer-generated randomized sampling distributed 80% of the images to the training set and 20% to the test set. We utilized LASSO regression to enhance model efficiency by eliminating less significant features through penalizing larger regression coefficients, effectively shrinking some to zero. This method prioritizes feature screening to refine the model. The selection of the regularization parameter, λ , critical to the LASSO model, was determined through 5-fold cross-validation, choosing λ based on the optimal performance criterion.

Machine Learning Models for Prediction

Each feature group was individually normalized by z-score to combine features of different magnitudes into one. Non-zero coefficients were used as useful predictors in each feature group, and LASSO regression was used for feature selection in the training cohort. Machine learning classifier Support vector machine (SVM) and GradientBoosting were used for predictive classification.

Visualization Techniques for Model Insights

We employed the Grad-CAM technique to produce heat maps that visualize the neural network's response during the grading of Idiopathic Macular Holes (IMH). These heat maps highlight specific regions within the images that are significantly influential in predicting the outcome, thereby providing insights into which areas of the image the neural network focuses on for making its predictions. This visualization aids in understanding the model's decision-making process.

Construction and Evaluation of the Nomogram

Logistic regression was used to analyze clinical factors, and a nomogram was established based on the results of the analysis to predict the relationship between clinical factors and the risk of requiring surgical treatment. "Grade 0–3" was defined as requiring no surgery or only intravitreal drug injection, and "Grade 4–5" was defined as requiring vitrectomy. The evaluation of the prediction model includes three parts, Discrimination: it is obtained by calculating the area under ROC curve (AUC) and the consistency index (C-index); Calibration: Describe by drawing the Calibrate Plot; Clinical Benefit: Evaluated using Decision Curve Analysis(DCA).

Statistical Analysis

All extracted features were dimensionally reduced by LASSO regression to improve the accuracy and fit of the modeling. The extracted features were then used to construct the prediction model with machine-learning methods, such as SVM and GradientBoosting. The performance of the study model was evaluated through AUC, accuracy measurements. Select data were also presented through confusion matrices and heat maps. Logistic regression was performed using the IBM Statistical Package for the Social Sciences (SPSS) 22.0 Windows software package. Nomogram developed and C-index calculated were based on logistic regression of software R 3.6.3. (<http://www.r-project.org>; version3.6.3). P value (two-sided) < 0.05 was considered to be statistically significant.

Results

Characteristics of Participants and OCT Images

A total of 229 OCT images from 229 participants were included in this study. According to the grading criteria, they were classified into grade 0 (27 images), grade 1 (58 images), grade 2 (73 images), grade 3 (15 images), grade 4 (23 images), and grade 5 (33 images). As shown in Table 1, 183 (80%) images comprised the training data set, while 46 (20%) images comprised the test set.

Table 1 Baseline Features of the Study Cohort and OCT Images

Clinical Characteristics	Participants, No.(%)		P value
Group	Train (n = 183)	Test(n = 46)	–
OCT images	183	46	–
Age, median±SD(range), years	50.33± 15.59	53.65 ± 15.26	0.81
Male	86	25	0.37
Female	97	21	

LASSO Model and Feature Screening

LASSO regression was used to screen the features of the fusion model, and regularization was introduced to screen the weight of each feature. Each feature was assigned a coefficient that represented its weight size. The larger the penalty coefficient, λ , the smaller the weight. Through 5-fold cross-validation, the optimal λ value identified for the fusion model was 0.022230 (Figure 3). Through the application of LASSO regression, 12 features with non-zero coefficients were selected, while 56 redundant features were successfully eliminated. This process preserved the features that are strongly associated with the disease, thereby reducing computational time, mitigating the risk of overfitting inherent in small-sample datasets with excessive features, and ultimately enhancing the overall performance of the classifier.

Comparison of Model Classification Performance

The Accuracy

Both classifiers showed the fusion model achieving top test-set accuracy (GradientBoosting: 74%; SVM: 85%). Using the GradientBoosting classification model, our results showed that the accuracy of the traditional omics model, Resnet101 model and the fusion model were 93%, 94% and 95% respectively, in the training set, whereas the scores were 70%, 67% and 74% in the test set, respectively. Comparatively, by using the SVM classification model, the accuracy of the three model were 78%, 93% and 92% respectively, in the training set, whereas the scores were 76%, 85% and 85% in the test set, respectively. As shown in Table 2.

Receiver Operating Characteristic (ROC) Curves and Area Under ROC Curve (AUC)

We utilized ROC curves and AUC values to distinguish IMH grades. The fusion model (SVM-based) achieved superior or comparable AUC values (0.90–1.00 across grades 0–5) for IMH grading compared to traditional omics (0.70–1.00) and ResNet101 (0.88–1.00) models, with particularly improved performance for grade 4 (AUC=0.90 vs 0.70 in omics).

In detail, in the fusion model, using the SVM classification as the classifier, the AUCs for grades 0–5 were calculated as 1.00, 0.97, 0.97, 0.90, 1.00 and 1.00 respectively, whereas the scores were 1.00, 0.92, 0.93, 0.70, 0.98 and 1.00 in the traditional omics model, 1.00, 0.96, 0.98, 0.88, 1.00 and 1.00 in the Resnet101 model. The GradientBoosting-based fusion model demonstrated superior and more consistent performance across IMH grades 0–5 (AUCs: 0.88–1.00), particularly excelling in grade 4 classification (AUC=0.91 vs 0.67 in omics and 0.73 in ResNet101), while maintaining competitive results in other grades compared to both traditional omics and ResNet101 approaches.

In detail, in the fusion model, the AUCs for grades 0–5 were calculated as 1.00, 0.90, 0.93, 0.91, 0.88 and 0.97, respectively. In comparison, the AUCs for grades 0–5 in the traditional omics model were 1.00, 0.88, 0.95, 0.67, 0.93,

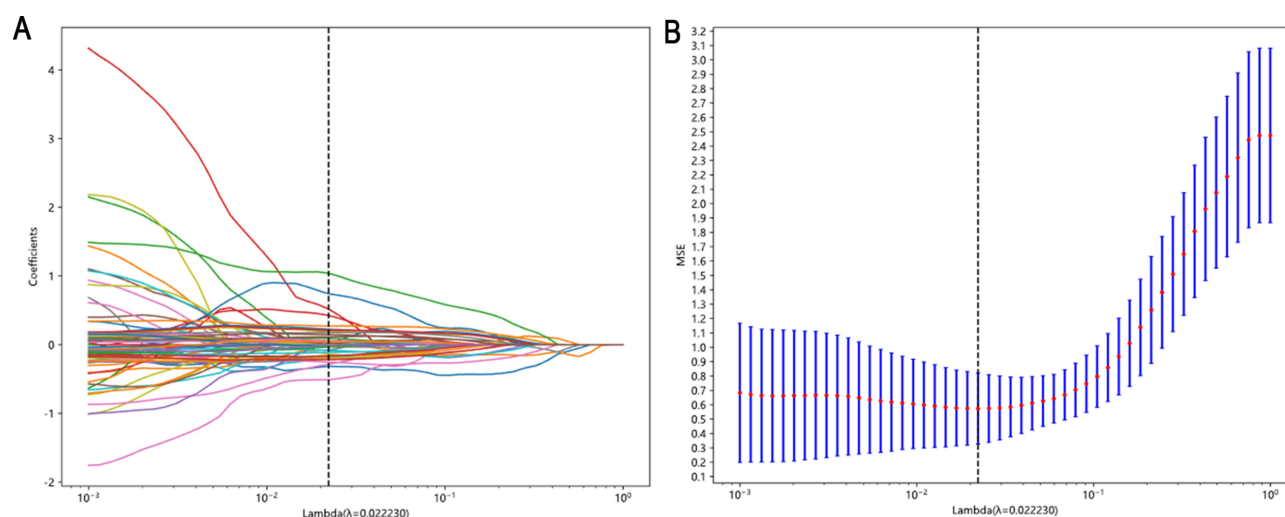


Figure 3 Feature selection by LASSO model. (A) LASSO coefficient profiles of the 75 variables (in Fusion model); (B) The curve of the binomial deviation with respect to the parameter λ , where the vertical dotted line at the right shows the optimal value of λ ($\lambda = 0.022230$) (in Fusion model).

Table 2 The Accuracy of the Three Models for the Train and Test Set

Model	Task	Accuracy	
		Gradient Boosting	SVM
Traditional omics model	Label-train	93%	78%
	Label-test	70%	76%
Resnet101 model	Label-train	94%	93%
	Label-test	67%	85%
Fusion model	Label-train	95%	92%
	Label-test	74%	85%

and 0.96, respectively, whereas the scores were 1.00,0.89,0.88,0.73,0.82 and 0.99, respectively in the Resnet101 model. As shown in Figure 4.

Confusion Matrix

The classification results of the test set were distributed in a 6×6 confusion matrix (Figure 5).

The fusion model demonstrated superior diagnostic performance, attaining the highest correct identification rates under both SVM (39/46) and GradientBoosting (34/46) classification frameworks when compared to traditional omics and ResNet101 models.

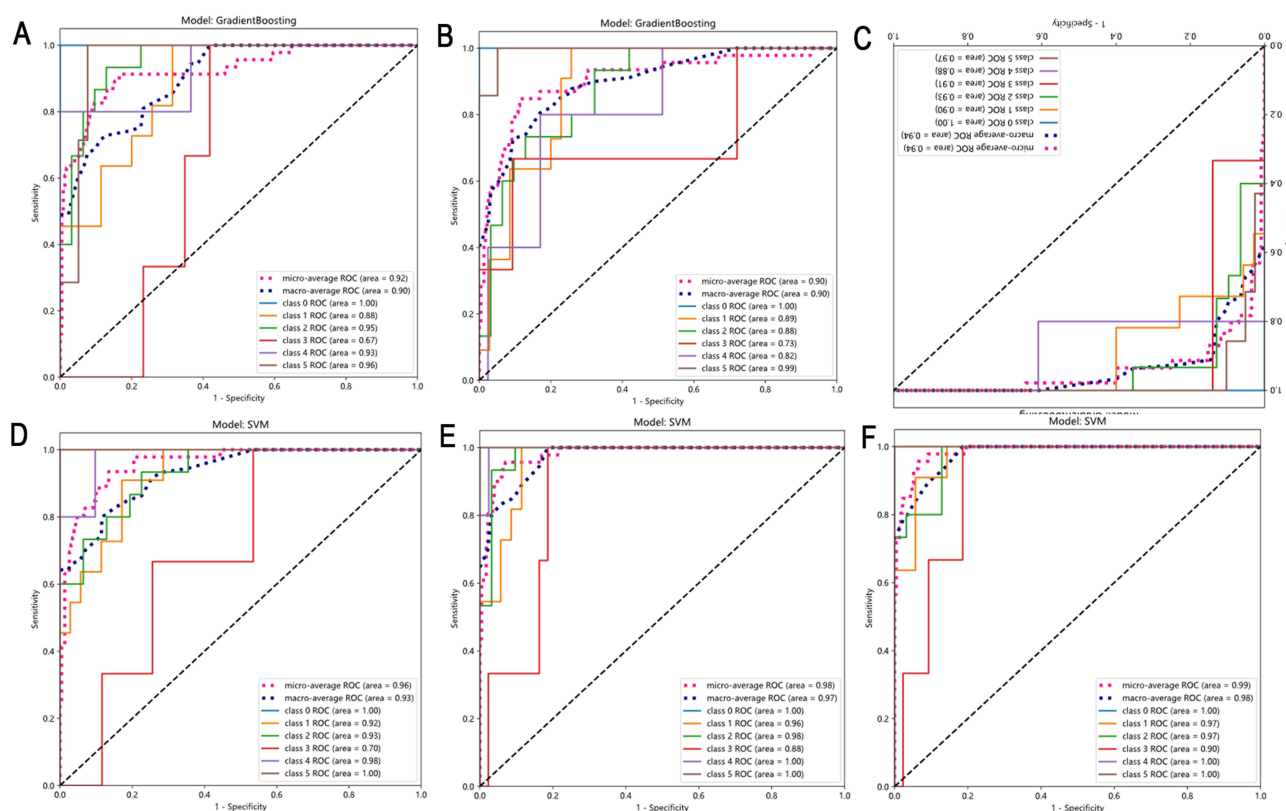


Figure 4 Receiver operating characteristic curves (ROC) and area under curve (AUC): AUC values for models under different classifiers. The five ROC curves in each figure represent the grade 0 (blue), grade 1 (Orange), grade 2 (green), grade 3 (red), grade 4 (purple), grade 5 (brown) respectively. "Class" means "grade". (A and D) Traditional omics model; (B and E) Resnet101 model; (C and F) Fusion model.

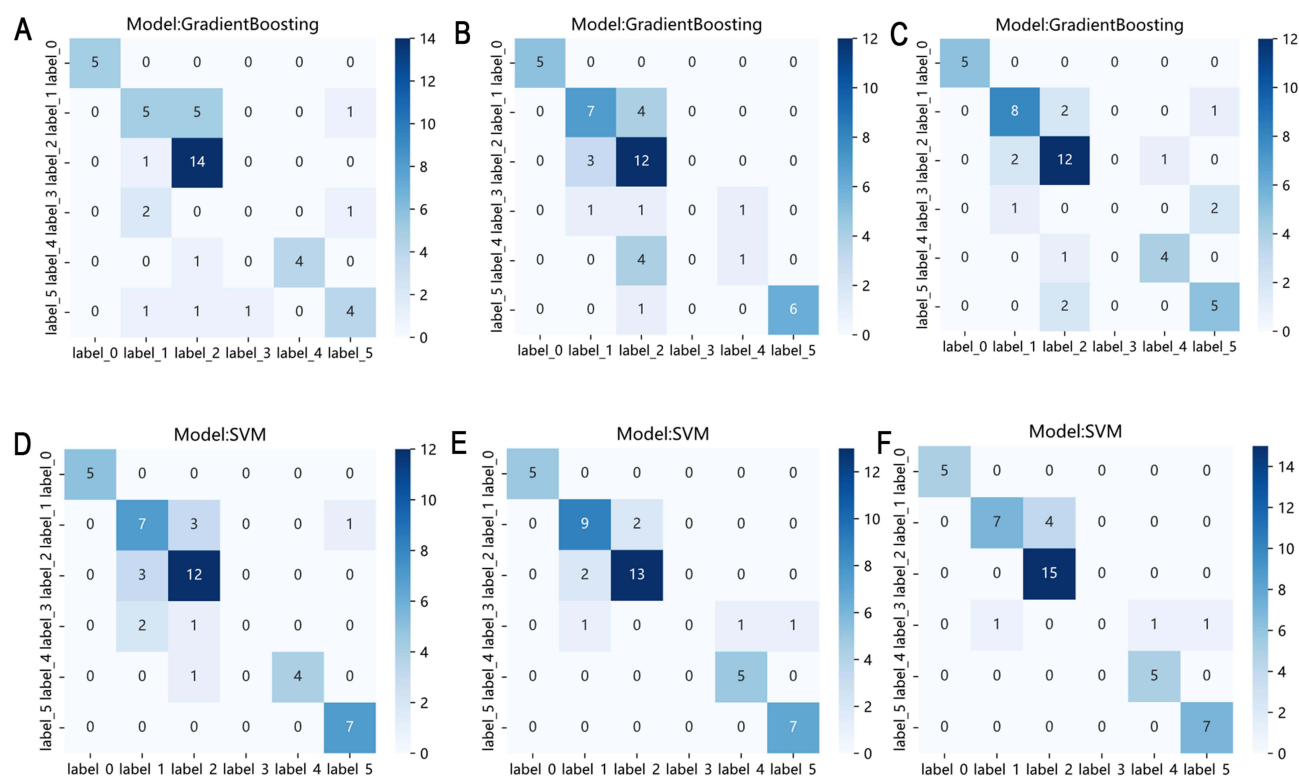


Figure 5 Confusion matrix. "label" means "grade". (A–C) When Gradient Boosting was used as the prediction classification model in the test set, the number of correct predictions by traditional omics model, Resnet101 model and fusion model were 32, 31 and 34 respectively; (D–F) When SVM was used as the prediction classification model in the test set, the number of correct predictions by the three model were 35, 39 and 39 respectively.

To elaborate, when SVM was used for predictive classification, the results showed that the fusion model demonstrated better recognition performance than the traditional omics model. Specifically, the fusion model identified 39 correct images, the same as the Resnet101 model, whereas the traditional omics model identified 35 images. The total number of images in the test set was 46. When GradientBoosting was used as the classification model, the three models demonstrated lower accuracy. In this scenario, the fusion model, the Resnet101 model and the traditional omics model identified only 34, 31 and 32 images correctly, respectively.

Heat Map

As seen in the heat maps, the central region of the macula is the primary region of interest (Figure 6). The blue regions indicate the model's prioritized feature extraction zones, with macular retinal structural characteristics - particularly full-thickness discontinuity and macular hole diameter - emerging as the predominant discriminative features, as visually demarcated in the heatmap.

Nomogram

Based on the results of the Logistic regression model (Table 3), a nomogram (C-index = 0.996, Figure 7) was established to predict the surgical risk. As shown in Figure 7, the hole diameter, basal diameter, and posterior vitreous detachment (PVD) were included as variables in the study. Each variable's corresponding line is marked with a scale representing the possible range of values for that variable, while the length of the line segment reflects the magnitude of that factor's contribution to the clinical outcome event. The "Point" means the individual score, which represents the corresponding single-item score for each variable at different values; and the "Total Point", which is the sum of the individual scores for all variables after their values are determined. "Predicted probability": This indicates the risk of requiring surgical intervention.

The results show that the risk score corresponding to hole diameter is the highest (100 points), PVD (35 points), and basal diameter (25 points). When clinical risk factors are visualized, the risk of surgery can be predicted. First, each independent clinical risk factor was projected upward to the first row of the scale to obtain the score for each factor, and

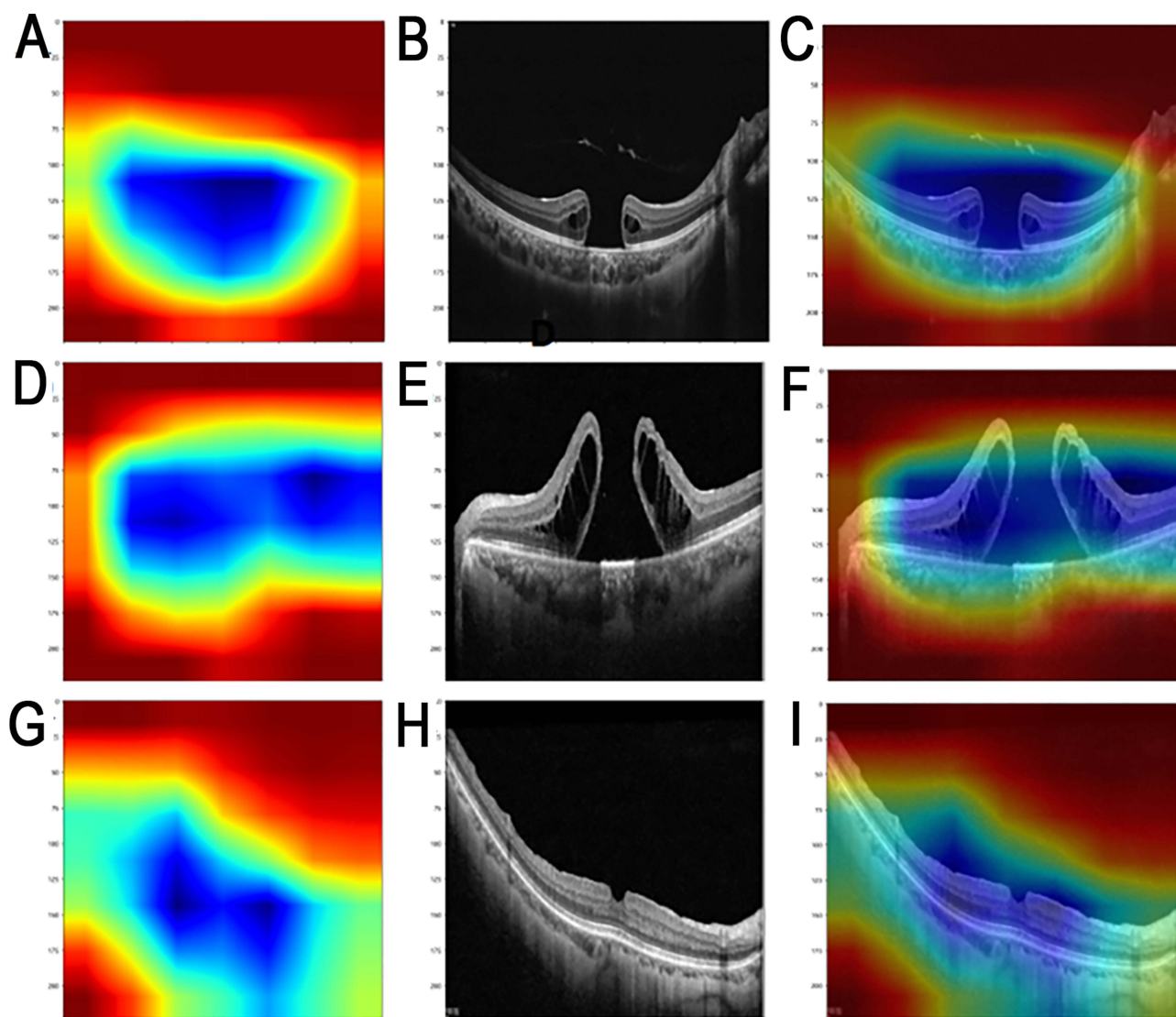


Figure 6 (A, D and G) Heat maps of the OCT images; (B, E and F) OCT images of IMH; (C, F and I) Composite images of the previous two images. The blue area is the synaptic focus of the system. The primary feature extraction region is localized at the macular hole position.

then the scores of the three risk factors were summed to obtain the total score. The ROC curve is shown in Figure 8A, and the AUC is 0.996, which illustrates the high prediction accuracy of the nomogram.

The calibration curve is shown in Figure 8B, where the points are close to the 45-degree line, indicating a high degree of fit between the predicted and observed values. Due to the small sample size in this study, we combined the training and validation cohorts and applied DCA to evaluate the clinical benefit of the nomogram, as shown in Figure 8C, when the threshold was between 0–1, the prediction model was clinically beneficial.

Table 3 Logistic Regression of Clinical Variables

Clinical Variables	Z-Score	P-value
Hole diameter	4.11	4.00E-05(P<0.05)
Basal diameter	2.36	0.02(P<0.05)
PVD	0.01	0.99(P>0.05)

Abbreviations: IMH, Idiopathic Macular Hole; AUC, area under the curve; OCT, optical coherence tomography (OCT).

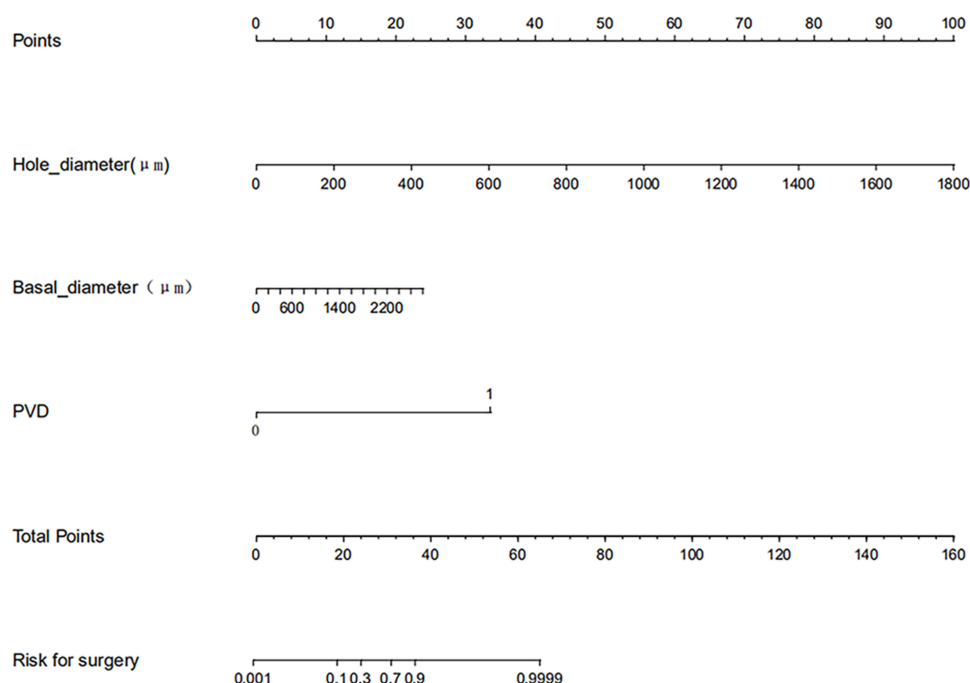


Figure 7 Nomogram for predicting surgical risk of idiopathic macular hole. Surgical decision predictors in IMH: The hole diameter measured at the retinal defect; Basal diameter at the retinal pigment epithelium level; PVD status (arrowheads indicate detached posterior hyaloid).

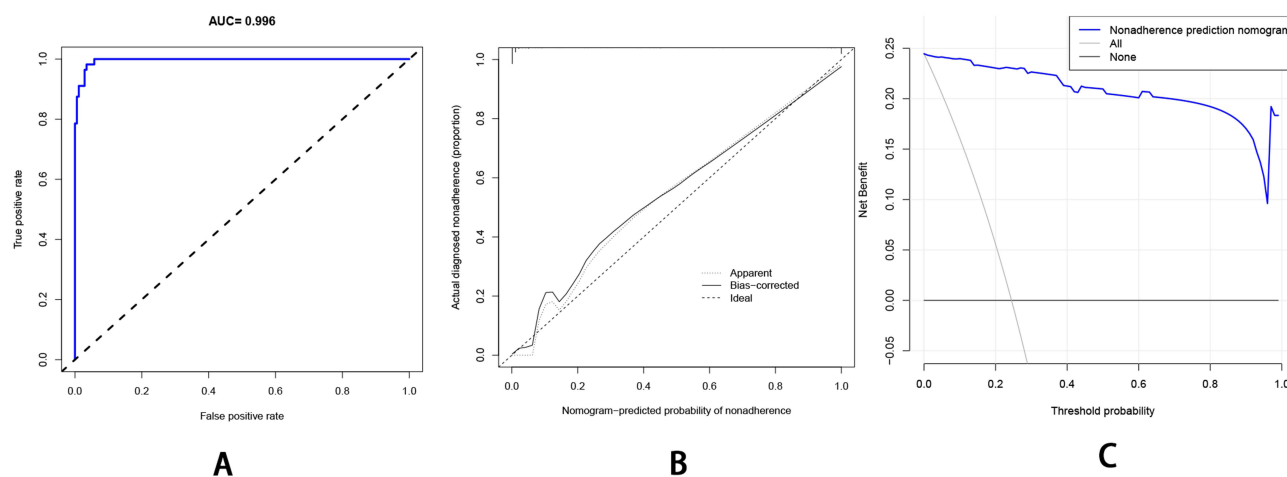


Figure 8 (A) The ROC curve and AUC are used to estimate the accuracy of the nomogram. $AUC=0.996$, which shows good discriminative ability between the IMH group and the normal group. (B) Calibration plot comparing actual vs predicted non-adherence. The x axis is the predicted probability, and the y axis is the actual probability. The 45-degree line (gray line) shows that the prediction matches the reality. (C) The DCA curve of the nomogram assessed the clinical benefit. X-axis: Threshold probability (0%-100%) indicating the clinical intervention cutoff. Y-axis: Standardized net benefit quantifying the model's clinical utility across thresholds. Solid blue line represents the prediction model, compared to "treat-all" (dashed line) and "treat-none" (horizontal line) strategies.

Discussion

The treatment of IMH is closely tied to its OCT-based staging. Several DL models for macular hole diagnosis and postoperative status prediction have been proposed. For example, Dong et al developed an AI system that can screen and diagnose 10 kinds of retinal diseases, including macular holes. Results from this study showed a high sensitivity of 89.8% (95% confidence interval, 89.5–90.1%).²⁸ In comparison, the DL model proposed by Hu et al was shown to predict macular hole status following vitrectomy with internal limiting membrane peeling with an accuracy, sensitivity, and specificity of 84.6%, 85.37%, and 81.99%, respectively.²⁹ Meanwhile, Lachance et al combined clinical features with a DL model to predict visual improvement after macular hole surgery. The AUC for this model was 81.9 ± 5.2 .³⁰

DL technology has allowed researchers to continuously develop and upgrade existing DL models. First proposed in 2015, ResNet³¹ was a milestone in the development of DL. ResNet is a residual neural network that mainly serves CNNs. CNNs can analyze matrices of images and gradually extract underlying features from highly abstract images. The more layers in the network, the richer the abstract features that can be extracted from each level. The more abstract the features, the larger the amount of semantic information that can be contained within them. Research has also shown that networks with more layers did not mean achieve better optimization, sometimes with more degradation. ResNet can also integrate residual mapping between each layer, which significantly increases model efficiency.

Transfer learning³² is another example of efficient machine-learning technology. Transfer learning, defined as the influence of one kind of learning on another, has been increasingly applied in DL. This particular algorithm is pre-trained with a specific data set containing millions of ordinary images. The resulting output is a high-precision model with a relatively smaller training data set. Comparatively, multi-mode fusion technology refers to the combination of multiple data sets in various forms, such as images and text, for target prediction. The resulting model can summarize the features of various modes. The primary benefit of this particular model is that it can extract the most valuable information from the obtained features. However, there may be some redundancy in the extracted information. As such, there remains room for improvement with this model.³³

Machine learning (ML) algorithms have been used for key feature training and recognition as well as group classification. Support Vector Machine (SVM)³⁴ learning is one of many machine learning methods. Compared to other ML methods, SVM is very powerful in identifying subtle patterns in complex data sets. The main idea of support vector machine regression is to map the data x to the high level feature space F through a nonlinear mapping Φ , and carry out linear regression in this space. Support vector machines create separate hyperplanes to give class-level predictions, which is a powerful method for constructing classifiers. GradientBoosting³⁵ is another ML, which is a combination of regression and classification tree models that apply progressively improved estimation to boost predictive power, helping to improve the accuracy of the tree. Each new tree constructed approximates the negative gradient of the empirical loss function, which can correct the mistakes made by the previous trees in the collection. Both of these classifiers have been widely used in medicine. The comparison and method improvement of various classifiers and the application of integration technology make the performance of classifiers improve constantly.^{36,37} This study aimed to develop a transfer learning and multi-feature fusion-based AI system that can classify IMH from OCT images and help guide ophthalmologists regarding patient treatment. In addition, we compared the diagnostic performance of the two classifiers (SVM and GradientBoosting) for IMH classification. Our results demonstrate that the SVM model achieved higher diagnostic accuracy than Gradient Boosting in the confusion matrix analysis. We hypothesize that Gradient Boosting underperforms SVM because its tree-based additive model struggles to capture critical non-linear feature interactions (hole diameter and PVD) that are explicitly modeled by SVM's kernel trick.

To the best of our knowledge, there are no previous reports on DL models for OCT-based macular hole grading. However, other DL models have been proposed for the diagnosis of other ophthalmologic conditions. For example, Hung et al utilized three other anterior segment-based AI models to construct a model that can grade pterygium. The accuracy values of Hung's model was 86.67%.²² Son et al compared the application of ResNet18, WideResNet50-2, ResNext50, and their fusion models for cataract grading. The results from their study demonstrated that their fusion model had the best predictive performance, with an accuracy of more than 90%.²³ Bhardwaj et al applied the InceptionResnet-V2 deep neural network framework to propose a quadrant-integrated automatic diabetic retinopathy classification method with an accuracy of 93.33%.²¹ Li et al adopted the CNN-combined transfer learning method of Inceptionresnet V2 to screen patients at risk of vision loss from high myopia. The resulting model had sensitivity and specificity values over 90%.³⁸ These studies have all demonstrated the excellent diagnostic performance of deep learning models for multi-class data. Our study is the first to apply deep learning models to the grading and screening of idiopathic macular holes. By combining transfer learning and multi-feature fusion methods, we further explored the use of multimodal data to improve feature extraction techniques. Additionally, we compared the screening performance of traditional radiomics with that of the ResNet101 model. This represents a novel attempt in the application of deep learning models for classification-based diagnostic screening of ophthalmic diseases. Our results show that, the fusion model in this study also showed exceptional results, with accuracy 95%, using the classification model of GradientBoosting. Comparatively, the traditional omics model and the Resnet101 model had accuracy scores of 93% and 94%, respectively. Similar results were found with classification model of SVM. Notably, both models performed not

particularly well when it came to small macular holes. As such, future research can focus on add the clinical measurement results of macular hole diameter into the model as a training parameter. Through comparative analysis, we observed that the deep learning models demonstrated exceptional performance in terms of both accuracy and AUC values. Specifically, the Fusion model and Resnet101 model achieved the highest test accuracy and AUC values under the SVM algorithm, highlighting their outstanding capability in handling complex data. High AUC values indicate that the models have excellent discriminatory power across various thresholds, enabling more accurate predictions of positive and negative samples.

Specifically, with the SVM algorithm, the Fusion model and Resnet101 model achieved AUC values of 0.99 and 0.98, respectively, while the traditional omics model had an AUC value of 0.96. Under the Gradient Boosting algorithm, the AUC values were 0.94, 0.90, and 0.92 for the Fusion model, Resnet101 model, and traditional omics model, respectively. These high AUC values further support the superior performance of the models.

AUC values are crucial for evaluating model performance, especially with imbalanced datasets. Higher AUC values indicate that the model has better discriminatory power, meaning it can accurately distinguish between positive and negative samples across various thresholds. In this study, the deep learning models, particularly the Resnet101 model and the Fusion model, achieved high AUC values, underscoring their potential in practical applications. A higher AUC value means the model can maintain good performance in recognizing positive and negative samples across different thresholds, which is particularly important when handling imbalanced datasets, as it ensures the model does not favor the majority class.

When SVM was used as the classifier, the confusion matrix showed the maximum number of correct identifiers, and the three models (the fusion model, the Resnet101 model and the traditional omics model) were 39, 39 and 35, respectively. However, with GradientBoosting, the correct scores were only 34, 31, 32, respectively. Therefore, when SVM was used as the classifier, our model showed better classification effect.

In summary, the deep learning models, especially the Resnet101 model and Fusion model, demonstrated high accuracy and AUC values in IMH classification, proving their superior capability in handling complex datasets. These results not only provide effective tools for clinical practice but also guide future research directions.

Research has shown that the minimum diameter of the macular hole (hole diameter)¹⁰ and the basal diameter³⁹ of the macular hole are closely associated with the surgical outcomes of macular hole treatment. In addition, the latest macular hole classification also takes the diameter of macular hole as one of the important criteria, and PVD is another important classification index.⁸ According to the latest grading criteria and combined with the results of our regression analysis (Table 3), it can be seen that macular hole diameter is the most relevant to predict the possibility of surgery, but the correlation of PVD did not show statistical significance in our results, which may be related to our small sample size, and a larger sample is still needed to confirm in the future. The macular hole diameter, basal diameter and PVD were analyzed as clinical variables of the nomogram, and the results showed that the prediction model had good diagnostic performance (C-index = 0.996). Calibration Plot and DCA were used for model evaluation, and the results also showed good performance of the prediction model.

Image grading analysis has broad medical applications, including disease diagnosis, condition monitoring, and treatment evaluation. This study, to the authors' knowledge, is the first to report on macular hole grading diagnosis. Current research in image grading primarily focuses on deep learning techniques, such as convolutional neural networks (CNNs) and pre-trained models, along with multimodal data fusion. Key challenges in this field include effectively extracting key features from large datasets, enhancing model generalization, and validating models in clinical settings.

This study introduces a novel approach in ophthalmic disease diagnosis, using deep learning and multi-feature fusion techniques. By employing CNNs and transfer learning, the need for large-scale data is reduced, easing data labeling, lowering costs, and improving model accuracy. LASSO regression was used to efficiently extract features from imaging data, optimizing the fusion strategy and enhancing model generalization.

Additionally, this study incorporated clinical factors (eg, posterior vitreous detachment and macular hole diameter) and developed a nomogram to predict the relationship between these factors and the risk of requiring surgery. This innovative approach validated the model in clinical practice, with results demonstrating excellent predictive capabilities.

Our study demonstrates the high accuracy and AUC values of deep learning models in IMH classification, indicating their superior performance in handling complex medical imaging data. Deep learning models, particularly Resnet101 and the fusion model, possess robust feature extraction capabilities that extend beyond OCT images to other types of medical

imaging, such as X-rays, CT scans, and MRI images.⁴⁰ The use of transfer learning in our study has shown the potential to achieve high-precision classification with smaller datasets, a method applicable to other medical imaging tasks like breast cancer screening and skin lesion detection.^{41,42} Additionally, the successful application of multimodal fusion technology in this study suggests that combining different types of imaging data can provide more comprehensive diagnostic information.⁴³ For instance, in oncology, combining PET and CT scans can enhance diagnostic accuracy.⁴⁴ The importance of high AUC values lies in their performance on imbalanced datasets, ensuring that models remain efficient in identifying minority classes (such as abnormal cases), which is crucial for practical clinical applications.¹⁹ Therefore, deep learning models exhibit generalizability and effectiveness across various medical imaging tasks. Future research should further validate the performance of these models in other medical imaging fields, thereby advancing the comprehensive development of medical imaging analysis.

Our study has limitations. Firstly, our dataset was limited in size and derived from a single center. Small sample size is a key factor contributing to model overfitting. When the training data volume is insufficient relative to model complexity, the model cannot adequately learn generalizable patterns. This manifests as the model excessively memorizing training set noise (specific scanning artifacts) rather than learning pathological features, ultimately compromising generalization capability. Secondly, image quality plays a crucial role, as poor-quality image data can significantly impair the model's ability to learn generalizable patterns. In our study design, we implemented several measures to mitigate overfitting risks: (1) applied regularization techniques to reduce the influence of non-essential features, (2) employed cross-validation to evaluate model performance, and (3) excluded poor-quality OCT images through rigorous quality control. These combined measures effectively minimize the impact of potential overfitting.

Future work should focus on: 1. Expanding the dataset, especially with data from multiple centers, to validate the model's generalization and robustness. 2. Enhancing multimodal data fusion, integrating additional imaging like OCT-A and Fundus Photography to improve accuracy and diagnostic capability. 3. Developing real-time diagnostic tools that integrate the model into existing medical workflows. 4. Researching personalized treatment plans by incorporating patient differences and histories for more precise recommendations. 5. Improving model interpretability, such as enhancing heatmap displays, to help clinicians better understand and trust the model's outputs for clinical use. 6. Enhancing the verification and supervision processes for clinical physicians and promptly eliminating the noise and unimportant features of learning characteristics will further improve the performance of the model.

Conclusion

Our study shows that the DL model based on multi-feature fusion and transfer learning can be used to diagnose and grade macular holes from OCT images. The performance of fusion model is better than that of traditional omics model and Resnet101 model. Among the classifiers tested, SVM demonstrated superior classification performance. The nomogram based on macular hole diameter has high accuracy in predicting surgical risk and the indicator with the highest risk is the hole diameter.

Data Sharing Statement

The data associated with this study has not been deposited into a publicly available repository, but it can be obtained from the authors upon reasonable request.

Ethics Statement

The studies involving human participants were reviewed and approved by institutional review board of The First Affiliated Hospital of Nanchang University (Ethics No.(2023) CDYFYLK(04-040)). This study obtained written informed consent from the patients, who agreed to the publication of all images, clinical data, and other data included in the manuscript.

Consent for Publication

We statement that the details of any images, videos, recordings, etc can be published, and that the persons providing consent have been shown the article contents to be published.

Acknowledgment

The authors declare that this manuscript has been submitted as a preprint on SSRN, available at [https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4845388].

Funding

This work was supported by National Natural Science Foundation of China [grant numbers 82260211, 81460092]; The Central Government Guides Local Science and Technology Development Foundation [grant number 20211ZDG02003]; The Key research and development project in Jiangxi Province [grant number 20203BBG73058 and 20192BBGL70033]; The Chinese medicine science and technology project in Jiangxi province [grant number 2020A0166]; The funding sources: National Natural Science Foundation of China; Ministry of Science and Technology of the People's Republic of China; Science and Technology Department of Jiangxi Province, China.

Disclosure

The authors declare that the research was conducted in the absence of any financial or non-financial relationships that could be construed as a potential conflict of interest.

References

1. Thapa SS, Thapa R, Paudyal I, et al. Prevalence and pattern of vitreo-retinal diseases in Nepal: the Bhaktapur glaucoma study. *BMC Ophthalmol*. 2013;13(1):9. doi:10.1186/1471-2415-13-9
2. Wang S, Xu L, Jonas JB. Prevalence of full-thickness macular holes in urban and rural adult Chinese: the Beijing eye study. *Am J Ophthalmol*. 2006;141(3):589–591. doi:10.1016/j.ajo.2005.10.021
3. Nangia V, Jonas JB, Khare A, Lambat S. Prevalence of macular holes in rural central India. The central India eye and medical study. *Graefes Arch Clin Exp Ophthalmol*. 2012;250(7):1105–1107. doi:10.1007/s00417-012-1933-8
4. Russell SR. What we know (and do not know) about vitreoretinal adhesion. *Retina*. 2012;32(Suppl 2):S181–S186.
5. Sonmez K, Capone A, Trese MT, Williams GA. Vitreomacular traction syndrome: impact of anatomical configuration on anatomical and visual outcomes. *Retina*. 2008;28(9):1207–1214. doi:10.1097/IAE.0b013e31817b6b0f
6. Gass JD. Reappraisal of biomicroscopic classification of stages of development of a macular hole. *Am J Ophthalmol*. 1995;119(6):752–759. doi:10.1016/S0002-9394(14)72781-3
7. Altaweel M, Ip M. Macular hole: improved understanding of pathogenesis, staging, and management based on optical coherence tomography. *Semin Ophthalmol*. 2003;18(2):58–66. doi:10.1076/soph.18.2.58.15858
8. Duker JS, Kaiser PK, Binder S, et al. The international vitreomacular traction study group classification of vitreomacular adhesion, traction, and macular hole. *Ophthalmol*. 2013;120(12):2611–2619. doi:10.1016/j.ophtha.2013.07.042
9. Ip MS, Baker BJ, Duker JS, et al. Anatomical outcomes of surgery for idiopathic macular hole as determined by optical coherence tomography. *Arch Ophthalmol*. 2002;120(1):29–35. doi:10.1001/archophth.120.1.29
10. Ullrich S, Haritoglou C, Gass C, Schaumberger M, Ulbig MW, Kampik A. Macular hole size as a prognostic factor in macular hole surgery. *Br J Ophthalmol*. 2002;86(4):390–393. doi:10.1136/bjo.86.4.390
11. Trese MT, Williams GA, Hartzer MK. A new approach to stage 3 macular holes. *Ophthalmol*. 2000;107(8):1607–1611. doi:10.1016/S0161-6420(00)00210-4
12. Rezende FA, Ferreira BG, Rampakakis E, et al. Surgical classification for large macular hole: based on different surgical techniques results: the CLOSE study group. *Int J Retina Vitreous*. 2023;9(1):4. doi:10.1186/s40942-022-00439-4
13. Stalmans P, Benz MS, Gandorfer A, et al. Enzymatic vitreolysis with ocriplasmin for vitreomacular traction and macular holes. *N Engl J Med*. 2012;367(7):606–615. doi:10.1056/NEJMoa1110823
14. Dugel PU, Tolentino M, Feiner L, Kozma P, Leroy A. Results of the 2-year ocriplasmin for treatment for symptomatic vitreomacular adhesion including macular hole (OASIS) randomized trial. *Ophthalmol*. 2016;123(10):2232–2247. doi:10.1016/j.ophtha.2016.06.043
15. Lu W, Tong Y, Yu Y, Xing Y, Chen C, Shen Y. Deep learning-based automated classification of multi-categorical abnormalities from optical coherence tomography images. *Transl Vis Sci Technol*. 2018;7(6):41. doi:10.1167/tvst.7.6.41
16. Li F, Chen H, Liu Z, et al. Deep learning-based automated detection of retinal diseases using optical coherence tomography images. *Biomed Opt Express*. 2019;10(12):6204–6226. doi:10.1364/BOE.10.006204
17. Yang H, Ahn Y, Askaruly S, You JS, Kim SW, Jung W. Deep learning-based glaucoma screening using regional RNFL thickness in fundus photography. *Diagnostics*. 2022;12(11):2894. doi:10.3390/diagnostics12112894
18. Jiang J, Lei S, Zhu M, et al. Improving the generalizability of infantile cataracts detection via deep learning-based lens partition strategy and multicenter datasets. *Front Med*. 2021;8:664023. doi:10.3389/fmed.2021.664023
19. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115–118. doi:10.1038/nature21056
20. Lee CS, Baughman DM, Lee AY. Deep learning is effective for the classification of OCT images of normal versus age-related macular degeneration. *Ophthalmol Retina*. 2017;1(4):322–327. doi:10.1016/j.oret.2016.12.009
21. Bhardwaj C, Jain S, Sood M. Deep learning-based diabetic retinopathy severity grading system employing quadrant ensemble model. *J Digit Imaging*. 2021;34(2):440–457. doi:10.1007/s10278-021-00418-5

22. Hung KH, Lin C, Roan J, et al. Application of a deep learning system in pterygium grading and further prediction of recurrence with slit lamp photographs. *Diagnostics*. 2022;12(4):888. doi:10.3390/diagnostics12040888
23. Son KY, Ko J, Kim E, et al. Deep learning-based cataract detection and grading from slit-lamp and retro-illumination photographs: model development and validation study. *Ophthalmol Sci*. 2022;2(2):100147. doi:10.1016/j.xops.2022.100147
24. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. *Lancet Oncol*. 2015;16(4):e173–80. doi:10.1016/S1470-2045(14)71116-7
25. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *Proc IEEE Conf Comput Vis Pattern Recognit*. 2016;770–778.
26. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. ImageNet: a large-scale hierarchical image database. *Proc IEEE Conf Comput Vis Pattern Recognit*. 2009;248–255.
27. Liu J, Sun W, Zhao X, Zhao J, Jiang Z. Deep feature fusion classification network (DFFCNet): towards accurate diagnosis of COVID-19 using chest X-rays images. *Biomed Signal Process Control*. 2022;76:103677. doi:10.1016/j.bspc.2022.103677
28. Dong L, He W, Zhang R, et al. Artificial intelligence for screening of multiple retinal and optic nerve diseases. *JAMA Network Open*. 2022;5(5):e229960. doi:10.1001/jamanetworkopen.2022.9960
29. Hu Y, Xiao Y, Quan W, et al. A multi-center study of prediction of macular hole status after vitrectomy and internal limiting membrane peeling by a deep learning model. *Ann Transl Med*. 2021;9(1):51. doi:10.21037/atm-20-1789
30. Lachance A, Godbout M, Antaki F, et al. Predicting visual improvement after macular hole surgery: a combined model using deep learning and clinical features. *Transl Vis Sci Technol*. 2022;11(4):6. doi:10.1167/tvst.11.4.6
31. He K, Zhang X, Ren S, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans Pattern Anal Mach Intell*. 2015;37(9):1904–1916. doi:10.1109/TPAMI.2015.2389824
32. Ma J, Bao L, Lou Q, Kong D. Transfer learning for automatic joint segmentation of thyroid and breast lesions from ultrasound images. *Int J Comput Assist Radiol Surg*. 2022;17(2):363–372. doi:10.1007/s11548-021-02505-y
33. Chen M, Jin K, Yan Y, et al. Automated diagnosis of age-related macular degeneration using multi-modal vertical plane feature fusion via deep learning. *Med Phys*. 2022;49(4):2324–2333. doi:10.1002/mp.15541
34. Noble WS. What is a support vector machine? *Nat Biotechnol*. 2006;24(12):1565–1567. doi:10.1038/nbt1206-1565
35. Oguz BU, Shinohara RT, Yushkevich PA, Oguz I. Gradient boosted trees for corrective learning. *Mach Learn Med Imaging*. 2017;10541:203–211.
36. Golpour P, Ghayour-Mobarhan M, Saki A, et al. Comparison of support vector machine, naïve Bayes and logistic regression for assessing the necessity for coronary angiography. *Int J Environ Res Public Health*. 2020;17(18):6449. doi:10.3390/ijerph17186449
37. Ebrahimi M, Mohammadi-Dehcheshmeh M, Ebrahimie E, Petrovski KR. Comprehensive analysis of machine learning models for prediction of sub-clinical mastitis: deep learning and gradient-boosted trees outperform other models. *Comput Biol Med*. 2019;114:103456. doi:10.1016/j.combiomed.2019.103456
38. Li Y, Feng W, Zhao X, et al. Development and validation of a deep learning system to screen vision-threatening conditions in high myopia using optical coherence tomography images. *Br J Ophthalmol*. 2022;106(5):633–639. doi:10.1136/bjophthalmol-2020-317825
39. Kim SH, Kim HK, Yang JY, Lee SC, Kim SS. Visual recovery after macular hole surgery and related prognostic factors. *Korean J Ophthalmol*. 2018;32(2):140–146. doi:10.3341/kjo.2017.0085
40. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60–88. doi:10.1016/j.media.2017.07.005
41. Shin HC, Roth HR, Gao M, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imaging*. 2016;35(5):1285–1298. doi:10.1109/TMI.2016.2528162
42. Valindria VV, Pawlowski N, Rajchl M, et al. Multi-modal learning from unpaired images: application to multi-organ segmentation in CT and MRI. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). 2018:547–556.
43. Townsend DW. Physical principles and technology of clinical PET imaging. *Ann Acad Med Singap*. 2004;33(2):133–145. doi:10.47102/annals-acadmedsg.V33N2p133
44. He H, Ma Y. *Imbalanced Learning: Foundations, Algorithms, and Applications*. Wiley-IEEE Press; 2013.

Clinical Ophthalmology

Publish your work in this journal

Clinical Ophthalmology is an international, peer-reviewed journal covering all subspecialties within ophthalmology. Key topics include: Optometry; Visual science; Pharmacology and drug therapy in eye diseases; Basic Sciences; Primary and Secondary eye care; Patient Safety and Quality of Care Improvements. This journal is indexed on PubMed Central and CAS, and is the official journal of The Society of Clinical Ophthalmology (SCO). The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/clinical-ophthalmology-journal>

Dovepress
Taylor & Francis Group