

Article



Hybrid Integent Staging of Age-Related Macular Degeneration for Decision-Making on Patient Management Tactics

Ekaterina A. Lopukhova ^{1,*,†}, Ernest S. Yusupov ^{1,†}, Rada R. Ibragimova ^{2,†}, Gulnaz M. Idrisova ^{2,†}, Timur R. Mukhamadeev ^{2,†}, Elizaveta P. Grakhova ^{1,†} and Ruslan V. Kutluyarov ^{1,†}

School of Photonics Engineering and Research Advances (SPhERA), Ufa University of Science and Technology, 32 Z. Validi Street, 450076 Ufa, Russia; ernest.math.yu@gmail.com (E.S.Y.); grakhova.ep@ugatu.su (E.P.G.)

² Department of Ophthalmology, Bashkir State Medical University, 3 Lenin Street, 450008 Ufa, Russia;

⁺ These authors contributed equally to this work.

Abstract: Treatment efficacy for age-related macular degeneration relies on early diagnosis and precise determination of the disease stage. This involves analyzing biomarkers in retinal images, which can be challenging when handling a large flow of patients and can compromise the quality of healthcare services. Clinical decision support systems offer a solution to this issue by employing intelligent algorithms to recognize biomarkers and specify the age-related macular degeneration stage through the analysis of retinal images. However, different stages of age-related macular degeneration may exhibit similar biomarkers, complicating the application of intelligent algorithms. This article presents a hybrid and hierarchical classification method for solving these problems. By leveraging the hybrid structure of the classifier, we can effectively manage issues commonly encountered with medical datasets, such as class imbalance and strong correlations between variables. The modifications to the intelligent algorithm proposed in this work for staging age-related macular degeneration resulted in an increase in average accuracy, sensitivity, and specificity of 20% compared to initial values. The Cohen's Kappa coefficient, used for consistency estimation between the regression model and expert assessments of the intermediate class severity, was 0.708, indicating a high level of agreement.

Keywords: age-related macular degeneration; optical coherence tomography; staging; computer vision; deep learning; hierarchical classification; semi-supervised learning

1. Introduction

Age-related macular degeneration (AMD) is a socially significant disease associated with the risk of central vision loss. According to 2020 data, the prevalence of AMD worldwide is about 200 million cases [1]. It is important to note that AMD is a chronic disease that tends to progress gradually [2,3]. Timely detection and treatment of AMD can help slow its progression and improve patients' quality of life [4,5]. Monitoring the development of AMD is essential for addressing this issue [6].

AMD has three stages: early, intermediate, and late [7], each with distinct clinical presentations and biomarkers. The early stage is identified by druses up to 125 μ m in size and is often asymptomatic, making it challenging to diagnose [8]. Signs of the intermediate stage include druses with a diameter of more than 125 μ m, drusenoid detachment of the pigment epithelium, and atrophy of the pigment epithelium outside the center. Complaints may also be absent at this stage [9]. The late stage is marked by deterioration of central



Academic Editor: Pedro Couto

Received: 22 December 2024 Revised: 28 January 2025 Accepted: 3 February 2025 Published: 13 February 2025

Citation: Lopukhova, E.A.; Yusupov, E.S.; Ibragimova, R.R.; Idrisova, G.M.; Mukhamadeev, T.R.; Grakhova, E.P.; Kutluyarov, R.V. Hybrid Integent Staging of Age-Related Macular Degeneration for Decision-Making on Patient Management Tactics. *Appl. Sci.* 2025, *15*, 1945. https://doi.org/ 10.3390/app15041945

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/ licenses/by/4.0/).

ibragimova.rada2016@yandex.ru (R.R.I.); idguma@mail.ru (G.M.I.); photobgmu@gmail.com (T.R.M.) Correspondence: lopukhova.ea@ugatu.su

vision, visual distortions, and changes in color perception. At this stage, geographic atrophy (GA) or macular neovascularization (MNV) may develop [10–12]. GA is the outcome of the early stage of AMD. MNV, if untreated, can lead to subretinal fibrosis (SF) [12,13].

The tactics of managing patients with different stages of AMD differ significantly [14]. There is no specific treatment for the early stage of AMD; preventive measures are used to eliminate risk factors for the development and progression of AMD [15,16]. Currently, drugs are being introduced to treat GA, but this therapy has yet to see widespread use [17]. In turn, treatment of the late form utilizes intravitreal injections to inhibit the vascular endothelial growth factor (anti-VEGF). These injections are widely utilized and have proven their effectiveness.

It is crucial to identify an intermediate stage that separates the early and late stages to make timely adjustments to patients' diagnostic and treatment plans. This can help reduce or neutralize the negative factors in the development of the disease [18,19]. A key aspect of analyzing the progression of AMD is tracking the moment when intravitreal injections of anti-VEGF drugs are needed to help slow down the progression of the late form of AMD [20].

However, distinguishing the intermediate stage of AMD from the early and late stages (geographic atrophy and macular neovascularization) can be difficult due to the similarity of biomarkers. The diagnostician needs to invest extra time and effort in visually identifying each biomarker and measuring its dimension, before comparing it with the evaluation scale. A large flow of patients can reduce efficiency due to human factors [21,22].

The most informative and standard diagnostic method for AMD is optical coherence tomography [23]. Currently, computer vision (CV) and machine learning (ML) methods are widely used to automate optical coherence tomography (OCT) visual analysis [24]. Many studies use ML to identify the intermediate stage [25]. In this case, two main approaches to the implementation of staging algorithms can be distinguished:

- 1. Direct classification of stages without identifying disease biomarkers [26–29].
- Classification of AMD biomarkers identified in a retinal image using an additional CV algorithm. The biomarker extraction algorithm can be implemented using retinal segmentation based on unsupervised learning [30–32] or using supervised learning algorithms by comparing medical images and a segmented set [33–35].

Biomarker-based AMD classification offers several advantages. It provides a better understanding of the algorithm and expands its scope of application. Biomarker segmentation can also be used separately in the image classification pipeline to extract the position and shape of detected biomarkers. This information can be valuable for a quantitative and statistical analysis of pathologies [36,37].

However, obtaining a labeled dataset can be a complex task requiring significant time investment for high-quality labeling [38]. Open-labeled datasets are only sometimes suitable for training ML models since the available OCT image datasets may not correspond to the imaging specifics of different tomographs [39,40]. Unsupervised learning segmentation methods also require expert participation to verify the results, which can be difficult [41].

In cases where obtaining a labeled dataset for segmentation is difficult or economically unfeasible, biomarker-based classification can be achieved by directly applying classifiers to images using additional predictor analysis algorithms. This approach does not involve identifying biomarker boundaries to assess their progression, but instead focuses the diagnostician's attention on the presence of a group of biomarkers. It allows the expert to concentrate on specific areas of the image to confirm or refute the hypotheses proposed by the algorithm. This approach aligns with the concept of decision support systems. An algorithm error is less likely to result in an incorrect decision by an expert than the allocation Addressing several associated challenges is essential to creating clinical decision support systems (CDSSs) for adjusting the treatment approach for retinal diseases using the AMD stage classifier. Identifying the intermediate stage of AMD is particularly challenging because the OCT images at this stage can resemble those from both the early and late stages of AMD. One effective approach to tackling this issue is to utilize the relationship between the identified classes of disease stages when implementing a hybrid classification [44]. Hybrid classification involves combining multiple ML models or methods to improve classification efficiency. This approach takes advantage of different algorithms to address specific shortcomings or limitations that one model may have [45]. Hierarchical classification can help address issues like class imbalance and overfitting when analyzing the structure of medical data. This method organizes the class space as a hierarchy, often represented as a tree or a directed acyclic graph [46]. Once the hierarchical tree is constructed, the ML model's work can be divided into tasks for each tree branch. This approach allows for breaking down a complex problem into more manageable subtasks.

Thus, ref. [47] demonstrated that organizing classes into a hierarchy can significantly enhance the scalability of the classification process. This breakthrough paves the way for decision-making at different levels and simplifies the complexity of distinguishing many classes simultaneously. Moreover, ref. [48] showed that hybrid and hierarchical classification can effectively organize the feature space, mainly when dealing with many classes. In addition, the authors of [49] demonstrated that hierarchical classification allows for identifying complex relationships between classes, which are often present, including in biological data. This leads to more meaningful and accurate classification results. The results of the studies show promise in addressing the various challenges that arise during the development process due to the unique nature of medical datasets and clinical decision support systems. This work aims to develop an OCT-based AMD classification algorithm, which will become part of the CDSS, monitor changes in AMD, and adjust patient management plans accordingly.

The main focus of the work was on the intermediate stage of AMD and the ability to assess its progression compared to the early and late stages of the disease. Implementing such algorithms involves preprocessing images and organizing data for training and testing a machine learning model [50]. However, data collection in clinical practice presents several challenges. These include visual similarities and overlaps among visual predictors in medical images [51] and the uneven frequency of disease stage detection in patients. These factors contribute to class imbalances in the generated dataset [52]. Such issues can significantly increase the resources and time required to develop computer vision applications for medical diagnostics.

The main issues are the need for a labeled OCT dataset and the need to address the class imbalance problem when identifying GA and SF. Additionally, an algorithm for detecting the intermediate stage without biomarker analysis is necessary, considering the high correlation of visual features in medical images of the intermediate, early, and late stages. The steps outlined in this paper for addressing the specific problem of AMD staging can be generalized into a method for developing a hybrid classifier. This approach minimizes the time diagnosticians spend during its creation, as it eliminates the need for a detailed markup of OCT data and the long-term accumulation of a sufficient number of examples for less common stages of AMD.

2. Development of the Algorithm

This study created an approach to diagnose different stages of AMD by analyzing macular images. It identified four main stages: no disease, early, intermediate, and late. Additionally, two late-stage development scenarios were identified: GA and SF. A series of OCT images was then generated and pre-processed. A basic classifier model was then trained and tested on these images to identify various challenges that affect the accurate classification of AMD stages and development options.

2.1. Dataset Structure

To develop the algorithm, we utilized a diverse set of 1928 OCT images of the macular region of patients with AMD. These images covered a wide range of AMD stages and were obtained using OCT on the Avanti XR (Optovue; Fremont, CA, USA) and REVO NX (Optopol; Zawiercie, Poland) devices at the Optimed Laser Vision Restoration Center in Ufa, Russia. Optical coherence tomography imaging of the macular region was performed in all patients using the Radial line mode. The study included OCT images from patients without concomitant pathology of the macular region and with sufficient optical transparency. The analysis was conducted with optimal signal strength and without any artifacts.

The original image files obtained from the Avanti X-ray tomograph were screenshots of the main screen of the Optovue application, displaying an OCT B-scan out of fifteen included in Radial Liens, saved in JPEG format. The diagnosticians selected the visualized B-scan measuring 958 by 404 pixels by the criterion of the greatest severity of the characteristic visual signs defining the stage of AMD by their expert experience. Similarly, the REVO NX tomograph produced B-scans of OCT measuring 860 by 580 pixels in JPEG format. The diagnosticians also selected these images based on the presence of radiation lesions, following the same process used for the Avanti X-ray tomograph.

The list of classes under consideration included the following cases with the corresponding code designations:

- 1. No disease (N) (23%);
- 2. Early AMD (S) (18%);
- 3. Intermediate stage (P) (18%);
- 4. Late AMD:

- Geographic atrophy (SI) (5%);
- Macular neovascularization (V) (26%) and subretinal fibrosis (VI) (10%).

To enhance the accuracy of our proposed classification model for Class P, we implemented a new process for analyzing input OCT images. These images can be categorized as Class P, and the nearest classes, S and V. This analysis aims to evaluate how closely the OCT image, initially classified as Class P, relates to Classes S or V.

We also decided to enhance the algorithm by enabling it to identify the progression of stage P based on its proximity to stages S and V. By labeling only a portion of the class, users can utilize their expert judgment to personalize the analysis of the intermediate class, reducing potential disagreements among experts. To determine the minimum threshold for the number of examples of class P, we considered several options for the proportion of labeled and unlabeled data: 1/6, 1/4, and 1/3. In the section on algorithm implementation, the choice of the minimum proportion of labeled data was analyzed in terms of its accuracy in identifying the progression of intermediate-class AMD on test samples.

The analysis of OCT B-scans obtained from two different tomographs revealed significant differences in the images. These differences were evident in size, brightness histograms, and the positioning of retinal layers within the B-scans. To streamline the process of identifying key predictors for the stages of AMD while filtering out irrelevant data, the following stages of image preprocessing were implemented:

- Normalizing pixel brightness levels to remove any color distortions;
- Generating new image samples by randomly flipping them horizontally, ensuring that all possible C-scan image positions were accounted for;
- Standardizing and reducing image size to ensure uniformity in visualizing images obtained from various tomographs, considering the retina's distinct horizontal orientation. The optimal image size for input data into the developed hybrid neural network was determined iteratively during the first stage of encoder training, as detailed in Section 2.2. Throughout this process, the width of the images was systematically reduced from 600 pixels to 404 pixels, considering the most diminutive dimensions of each image variant and the aspect ratio for the REVO NX. Ultimately, an image size of 64 by 128 pixels was selected, demonstrating the best accuracy for the trained model with minimal signs of low generalization ability. This was assessed by measuring deviations in the learning error graph for training and validation samples.

2.2. Developing a Classifier Structure and Identifying the Problem of Direct Classification of Medical Data

The modern approach to image analysis relies on processing images at the level of individual pixels. To analyze an image, the brightness of each pixel is determined, as

well as its location in the entire image and other clusters of pixels [53]. These features, obtained through analysis, form a vector representation of the image in a feature space. The more distinct the vectors representing one class are from those of others, the greater the classification effectiveness will be [54].

Convolutional neural networks (CNNs) are effective for extracting features from images. These networks can separate image vectors in feature space when there are enough representative examples for each class in classification problems [55,56]. However, additional steps are required when the number of images is limited or there is a class imbalance to ensure the stability of deep computer vision algorithms [57,58]. To address these challenges, we selected a base classification model and then modernized it step by step, evaluating the effectiveness of each change.

We selected a four-layer CNN-based encoder model as our classifier structure to implement the direct classification of AMD stages. This model has demonstrated effectiveness in tackling classification problems in computer vision, mainly when dealing with a limited number of classes [59]. We conducted an ablative analysis to evaluate whether the numbers of layers and parameters were sufficient to address the problem. The results of this analysis enabled us to draw the following conclusions:

- 2. When layers 3 and 4 were removed, the average precision level dropped below 0.5 for all classes except for class N, which maintained an average precision level of 0.65.
- 4. Incorporating a fifth layer into the encoder structure led to a slight improvement in the accuracy of classifying class N. In contrast, the accuracy for the remaining classes either remained at the same level as the model with four layers or decreased. This decline is attributed to model retraining, as indicated by the divergence in the error rates between the training and validation samples throughout the training process.

Metrics	Ν	Р	S	SI	V	VI
Precision	0.9725	0.5741	0.618	0.4565	0.5606	0.7358
	(0.96, 0.98)	(0.54, 0.6)	(0.58, 0.64)	(0.42, 0.48)	(0.52, 0.59)	(0.7, 0.76)
Sensitivity	0.6698	0.6526	0.6471	0.6774	0.6379	0.65
	(0.64, 0.69)	(0.62, 0.68)	(0.61, 0.67)	(0.64, 0.7)	(0.61, 0.66)	(0.62, 0.67)
Specificity	0.9939	0.8647	0.9029	0.9381	0.9231	0.9627
	(0.98, 0.99)	(0.84, 0.88)	(0.88, 0.92)	(0.92, 0.95)	(0.90, 0.93)	(0.95, 0.97)
F1-score	0.7933	0.6108	0.6322	0.5455	0.5968,	0.6903
	(0.76, 0.81)	(0.58, 0.64)	(0.6, 0.66)	(0.51, 0.57)	(0.56, 0.62)	(0.66, 0.72)

Table 1. Performance indicators of the primary encoder model with confidence interval.

To ensure consistent training conditions for all ML models throughout the study, the following training parameters were established:

- Number of model training epochs: 80 (the average time spent on training models was 25 min);
- Optimizer: Adam algorithm;
- Error function: cross entropy.

During the first training stage, as described in Section 2.2, we conducted an iterative process to determine the optimal hyperparameter values to maximize the encoder's accuracy. After fixing these values for the primary encoder model, they were applied to train subsequent model versions. The specific hyperparameter values are detailed below:

- 1. The Adam optimizer has a learning rate of 3×10^4 .
- 2. The size of the latent space in the encoder was 512.
- 3. The batch size was 128.

The technical specifications of the apparatus employed for training the models were as follows:

- Processor: AMD (Santa Clara, CA, USA) Ryzen 5700C;
 - Cores/Threads: 8 cores/16 threads;
 - Clock Speed: 3.8–4.6 GHz;
 - Cache: 4 MB L2, 16 MB L3;
- System Memory: 32 GB DDR4 (2 × 16 GB modules).

We trained and tested the primary encoder model using the transformed dataset to classify six classes, which include the main stages and two late-stage scenarios. To assess the performance of our classifiers, we utilized the following set of metrics:

- Precision measures the proportion of correct positive predictions among all positive predictions made by the model, including false positives. High precision indicates a significant probability that the answer is correct in the case of positive predictions for a given class [60];
- Sensitivity is the ratio of correctly identified positive cases to the total number of
 positive cases, including false negative cases. A high sensitivity value indicates that
 the model is more efficient at correctly identifying positive cases [60];
- Specificity measures the ratio of correctly identified negative cases to the total number of negative cases. A high specificity indicates that the model accurately identifies negative cases [60];
- F1-Score is the harmonic mean of precision and recall. It measures the model's overall performance, considering both false positives and false negatives, especially when dealing with class imbalance [60].

Figure 1 and Table 1, with confidence intervals for a 95% confidence level, present the encoder testing results in an error matrix.

The analysis of the confusion matrix and evaluation metrics for the binary classifier indicated that the model's performance was suboptimal. This was primarily due to its difficulty distinguishing the intermediate state and the imbalance between the SI and VI categories, which exhibit similar characteristics to AMD's early and advanced stages. The model demonstrated a conservative approach characterized by high precision but low recall. This indicated that while it was unlikely to make errors in classifying categories N and VI, there was a significant chance of missing many actual instances of these categories. In clinical practice, misclassifications between the absence of disease (N), early-stage AMD (S), and intermediate-stage AMD (P) could result in a risk of overlooking the onset of AMD. This situation may require adjustments in patient management strategies. Additionally, misclassifying geographic atrophy cases as earlier stages, with eight cases incorrectly categorized as stage S, could delay essential interventions.

To improve the efficiency of CNN, the parameters of convolutional layers and their impact on classification accuracy were studied. Since biomarkers in different stages of AMD vary in size, the impact of convolution kernel size was investigated in terms of accuracy in separating S and V classes. A neural network with convolutional and fully connected



Figure 1. Error matrix of the primary encoder model.



Figure 2. Sensitivity of the model for classes S and V when using different sizes of convolution kernels.

We explored how the encoder's performance varies when employing parallel convolution layers at each stage. The results of this study are presented in a confusion matrix, as shown in Figure 3 and summarized in Table 2.

We used statistical significance tests and effect size calculations to compare the performance of two models: a primary encoder and an encoder with parallel convolutional layers. To evaluate the effectiveness of parallel convolutional layers, we employed a nonparametric McNemar test to compare the two models on the same dataset [61]. We used a Cohen effect size measure, h, to estimate the extent of changes in the models' results for each class [62].

	_						
	z	75	20	11	0	0	0
	∟ -	1	65	13	4	2	10
label	თ -	1	13	65	14	1	2
True	<u>-</u>	0	0	8	23	0	0
	> .	0	6	3	2	41	6
	2.	0	0	0	0	20	40
		N	P	S	SI SI	V	VI
				Pred	licted lab	ei	

The results of the statistical significance tests, including the effect size, are presented in Table 3.

Figure 3. Encoder error matrix with the inclusion of parallel convolutional layers in the architecture.

Table 2. Performance indicators of the primary encoder model with the inclusion of parallel convolutional layers in the architecture with confidence interval.

Metrics	Ν	Р	S	SI	V	VI
Precision	0.974	0.6915	0.6465	0.5349	0.6406	0.6897
	(0.96, 0.98)	(0.66, 0.72)	(0.61, 0.67)	(0.51, 0.56)	(0.61, 0.67)	(0.66, 0.71)
Sensitivity	0.7075	0.6842	0.7529	0.7419	0.7069	0.6667
	(0.67, 0.73)	(0.65, 0.71)	(0.72, 0.78)	(0.71, 0.76)	(0.67, 0.73)	(0.63, 0.69)
Specificity	0.9939	0.9147	0.900	0.9505	0.939	0.952
	(0.98, 1.0)	(0.89, 0.93)	(0.88, 0.91)	(0.93, 0.96)	(0.92, 0.95)	(0.93, 0.96)
F1-score	0.8197	0.6878	0.6957	0.6216	0.6721	0.678
	(0.79, 0.84)	(0.65, 0.71)	(0.66, 0.72)	(0.59, 0.65)	(0.64, 0.7)	(0.64, 0.72)

Table 3. Statistical significance tests comparing the primary encoder and a primary encoder model, with the inclusion of parallel convolutional layers in the architecture.

ه () () () () () () () () () (Ν	Р	S	SI	V	VI
s s s s s s s s s s s s s s s s s s s s	0.3184 (0.573)	4.9130 (0.027)	0.2785 (0.598)	2.0313 (0.154)	2.4309 (0.119)	0.6163 (0.432)
ि () () () () () () () () () () () () ()	0.087	0.237	0.064	0.156	0.170	0.091

The analysis of the test results from the statistical evaluation of the two models shows that integrating parallel convolutional layers into the encoder structure resulted in a statistically significant improvement for class P. This is supported by the findings of the McNemar test (p < 0.05) and an average Cohen coefficient of h = 0.237. However, further modifications to the classifier were necessary to ensure a consistent improvement in model performance for the other classes. These adjustments also aimed to address the issue of significant class imbalance and improve the accuracy of class P differentiation from classes S and V.

Analyzing the data presented in Figure 3, we can conclude that the complexity of the trained model in separating classes S and SI requires additional measures to eliminate this disadvantage. The error matrix indicators indicate that class SI, from the point of view of the trained model, is closer to class S than to classes V and VI, which would be more logical from a clinical standpoint. Furthermore, the error matrix indicates that the model struggles to distinguish between classes V and VI, highlighting the need for a solution. An error analysis of the current version of the encoder model prompted us to take the following steps:

- 1. We applied a hierarchical classification approach, in which we identified the base classes of the senior hierarchy, and then defined the classes subordinate to them.
- 2. We consciously changed the clinical hierarchy of AMD stages given in Section 2.1, adapting it to solve the problems of class overlap for the trained encoder model.

In clinical practice, the development of a model to recognize essential hierarchies or create a global classifier should focus on four main categories: absence of disease, early stage of AMD, intermediate stage, and late stage of AMD. However, an analysis of results from different encoder versions has shown difficulty distinguishing between the GA stage (SI class) and the early stage of AMD (S class). This challenge is mainly due to the presence of overlapping features in OCT images, a phenomenon that has also been noted in other studies [63]. A similar issue arises when differentiating between the SF stage (VI class) and the late stage of AMD (V class) because of overlapping specific features, which have also been reported in other research [64].

Due to the overlap between classes S and SI and V and VI, we adjusted the hierarchical structure to simplify the classification task for the trained model. In the new hierarchical structure, the initial stage of AMD, referred to as class S, has been replaced with class S', which now encompasses both S and SI. Similarly, the later stage of AMD is classified as V', which includes only V and VI. To ensure users receive accurate information about the stages of AMD according to true classification, a deterministic set of rules has been developed that transforms AMD output data.

The allocation of a new base class, S', which includes the SI class and has a lower tendency to overlap with the P class, can help the model better distinguish the class within the whole of S'. Incorporating regression models into the structure of the hybrid classifier will further enhance this distinction. In cases where the definitions of the class between P and S are unclear, the regression models will still provide valuable insights.

In order to carry out a hierarchical classification in accordance with the above principles, we divided the difficult task of classifying six non-equilibrium classes into more manageable subtasks. As a result, the following components of the developed hierarchical classifier were identified:

- 1. Global classifier for the classification of primary classes of AMD stages: N, P, S' and V';
- 2. Binary classifiers working within the S' and V' classes to separate S and SI and V and VI, respectively;
- 3. Regression model for estimating the degree of proximity of class P to classes S and V.

It is important to note that the image features extracted by the global model using different convolutional layers can act as input to binary classifiers. This significantly reduces the computational complexity required for high accuracy, sensitivity, and specificity. This is achieved by passing information directly from the global model to these classifiers from intermediate layers. This approach was also utilized to create input data for the regression algorithm, which determines the degree of development of stage P.

Thus, a hybrid classifier integrates several models, including a global classifier, two binary classifiers for processing S' and V' subclasses, and a regressor for processing the intermediate stage P. The general scheme of its structure is shown in Appendix A.1. Common information data were allocated for the S' subclass and the regression of class P, effectively identifying small and large predictors. Only large predictors were allocated for the V' subclass, which proved to be the most effective approach.

The hybrid classifier's training process involved several stages. First, only the global classifier was trained. Then, the binary classifiers and the regressor were trained, and the global classifier processed the input data. To train the global classifier, classes S' and V' included examples from both their samples and from classes SI and VI, respectively. At the next stage, the binary classifiers S' and V' separated the classes S and SI, as well as V and VI.

The algorithm for determining the severity of intermediate AMD included a regressor operating in tandem with the Label Propagation (LPA) algorithm [65]. The training was conducted in three iterations using three different versions of the labeled dataset (1/6, 1/4, and 1/3). Accuracy was evaluated using a predetermined test dataset, which accounted for 10% of the total labeled data volume.

Thus, a method was developed to create a hybrid classifier to address the issues related to the flat classification of AMD stages, particularly in situations involving class imbalance and the presence of overlapping visual predictors for different classes, such as those observed in classes P and S. For example, the results obtained from the four-layer encoder we used in flat classification were unsatisfactory. The flowchart depicting the developed method is shown in Figure 4, where the main steps of the process are presented in plain text. At the same time, specific examples related to our classification of AMD stages are highlighted in italics.



Figure 4. A flowchart illustrating a method for creating a hybrid classifier to address flat classification problems.

3. Results

The hybrid classifier's initial stage of training involved training the global classifier using four selected base classes: N, P, S', and V'. The results from testing the global classifier are presented in the error matrix displayed in the accompanying Figure 5 and Table 4.



Figure 5. Error matrix of the global classifier.

Table 4. Performance indicators of the global classifier with confidence intervals.

Metrics	ة "	оста ста ста ста ста ста ста ста ста ста	ა ,	ब
Precision	ه: هـ	ಿ 1	0.8350 (0.812, 0.858)	甜
Sensitivity	笃 ¹	งсьсьсьсь примерии прим	0.7414 (0.714, 0.769)	陆
Specificity	聒音ाぢぢぢぢぢぢぢぢぢぢぢぢぢぢぢぢぢぢぢぢぢぢぢぢぢぢぢぢ57777777777	s s s s s s s s s s s s s s s s s s s s	0.9467 (0.933, 0.961)	さいいいいいいいいいいいいいいいいいいいいいいいいいいいいいいいいいいいいいいい
F1-score	उ	অ	0.7854 (0.761, 0.818)	उ

Testing of the global classifier revealed a significant reduction in class imbalance. It increased accuracy in P stage detection, suggesting that integrating the global classifier and regressor responses further improves the P detection process.

The values of the regressor were divided into three regions:

- P-free region (0–0.3);
- Region of initial P progression (0.3–0.67);
- Region of late P progression (0.67–1).

If the regressor's response falls outside the first interval and the global classifier indicates that the class S' closest to P is the correct one, then the final response is determined to be P.

In the next stage of developing a hybrid classifier, we trained binary classifiers to separated the classes S and SI, as well as V and VI. The results of testing these classifiers

are illustrated as error matrices in Figures 6 and 7, and their performance indicators are provided in Table 5.



Figure 6. Error matrices for binary classifier S and SI.



Figure 7. Error matrices for binary classifier V and VI.

The obtained data indicate that binary classifiers are effectively separate classes, S and SI and V and VI, as the metrics show consistently high values, even for adjacent classes.

The regressor was trained using a partially labeled dataset, with 80% of the data used for training. The dataset included three states: 0—no P, 0.5—the initial stage of P, and 1—the extreme P stage. The Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) were used as metrics to evaluate the regression model. MAE calculates the average error in a set of forecasts and is less affected by outliers. On the other hand, RMSE calculates

the square root of the average of the squared differences between the predicted and actual values. It is more sensitive to large deviations in the predicted value, allowing it to clearly emphasize significant discrepancies between the predicted and actual values [66].

»»	S	अ	V	み
ओ	0.9882 (0.98, 0.99)	줞µµµµµµµµµµµµµµµµµµµµµµµµµµµµµµµµµµµµµµµ	0.9492 (0.93, 0.96)	"
જ	0.9871 (0.97, 0.99)	ि	0.9655 (0.95, 0.97)	ओ
ծ 1	0.9531 (0.94, 0.96)	अ	0.954 (0.94, 0.96)	み
জmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm	0.9895 (0.98, 0.99)	尊	0.9573 (0.94, 0.96)	み

Table 5. Performance indicators of the joint work of the global classifier and binary classifiers.

First, the regressor was trained on these examples, and then the LPA model. The results of testing the regressor are presented in Table 6.

Table 6. Performance indicators of the regressor for assessing the degree of affinity between class P and classes S and V with confidence intervals.

""	1/6	1/4	1/3
じ	0.2297	0.0882	0.0484
	(0.14, 0.31)	(0.05, 0.12)	(0.03, 0.06)
ले ति	0.2827	0.1243	0.0551
	(0.17, 0.38)	(0.07, 0.17)	(0.03, 0.0757)

Concerning the boundaries we set for intermediate AMD progression, the outcome achieved using a 1/4 dataset can be minimally acceptable. For subsequent work, we utilized a model trained on a dataset with a 1/3 labeled data ratio.

The final stage of our study involved a comprehensive testing of the entire hybrid classifier. This testing was crucial in evaluating the performance of our system, and the results are presented in the error matrix in Figure 8 and Table 7.

Table 7. Performance of the hybrid classifier with confidence intervals.

Metrics	Ν	Р	S	SI	V	VI
Precision	0.9902	0.9381	0.9059	0.8611	0.9825	1.0000
	(0.98, 0.99)	(0.92, 0.95)	(0.88, 0.92)	(0.83, 0.88)	(0.97, 0.99)	(1.0, 1.0)
Sensitivity	0.9528	0.9579	0.9059	1.0000	0.9655	0.9667
	(0.93, 0.96)	(0.94, 0.97)	(0.88, 0.92)	(1.0, 1.0)	(0.95, 0.97)	(0.95, 0.97)
Specificity	0.997	0.9824	0.9771	0.9876	0.9973	1.0000
	(0.99, 1.00)	(0.97, 0.99)	(0.96, 0.98)	(0.98, 0.99)	(0.99, 1.00)	(1.0, 1.0)
F1-score	0.9712	0.9479	0.9059	0.9254	0.9739,	0.9831
	(0.96, 0.98)	(0.93, 0.96)	(0.88, 0.92)	(0.9, 0.94)	(0.96, 0.98)	(0.97, 0.99)

To assess progress in improving the classification of the entire set of AMD stage classes, the McNemar test was conducted, and the Cohen effect size measure h was applied to compare the outcomes of the source encoder and the hybrid classifier, as presented in Table 8.



Figure 8. Error matrix of hybrid classifier.

Table 8. Statistical significance tests comparing primary encoder and hybrid classifier.

VI
V I
7.8067 <0.001)
1.092
)

Statistical significance tests comparing the original and hybrid encoder models showed that the hybrid model was superior to the original model for classes P, S, SI, V, and VI. This was evident from the results of the McNemar test, which had a high degree of confidence (p < 0.001) and medium to huge Cohen's h values (h > 0.6). Because the N classification results were initially high, there were no significant improvements in its detection.

A remarkable improvement in classification accuracy across all stages of AMD was noted. The most advanced stages, including stages IV (SI), V, and VI, now show an accuracy exceeding 96%. The classification of geographic atrophy (GA) at stage I has been accomplished with perfect accuracy, reaching a remarkable 100%. There has been a significant reduction in misclassifications between adjacent stages, such as $N \rightarrow S \rightarrow P$. The latter stages of AMD, SI, V, and VI have been more accurately differentiated, with a minimum number of cross-classification errors. The model exhibits a consistent performance across all stages of AMD, addressing the previously identified imbalance where the late-stage classes performed poorly, ensuring a more equitable assessment. Despite the improvements, some bugs persist, which require attention for further stages of CDSS development:

- 1. A total of 5% of early-stage AMD cases were mistakenly classified as healthy retina, which is important to pay attention to for the possibility of early intervention. Meanwhile, 6% of early-stage AMD was misdiagnosed as geographical atrophy.
- 2. A total of 4% of intermediate-stage AMD cases were mistakenly attributed to the early stage, and 2% of early stage AMD cases were mistakenly attributed to the middle stage. These errors, although less critical, can lead to minor delays.

Using Cohen's Kappa coefficient, we assessed the agreement between the hybrid classifier and expert opinion for a statistical measure of their agreement [67]. The testing focused on diagnosing the most challenging cases of AMD, involving thorough analysis and comparison with an assessment scale. A set of 305 examples was given to an expert ophthalmologist, along with the developed algorithm, for evaluation. They were tasked with deciding whether to change the patient's treatment plan after diagnosing the early stage of the disease during a prior examination. Cohen's Kappa coefficient is calculated using the following formula:

$$k = \frac{p_o - p_e}{1 - p_e},$$
 (1)

$$p_o = \frac{N - (c+d)}{N},\tag{2}$$

$$\nu_e = \frac{(a+b)(a+c) + (c+d)(b+d)}{N^2}.$$
(3)

where *a* is the number of times both test participants answered affirmatively; *b* is the number of times test participant 1 answered affirmatively and participant 2 answered negatively; *c* is the number of times test participant 1 answered negatively and participant 2 answered affirmatively; *d* is the number of times both test participants answered negatively; *and N* is the total number of test examples.

During the testing of the algorithm and the expert, the following results were obtained:

$$p_o = \frac{305 - (20 + 20)}{305} = 0.86885,\tag{4}$$

$$p_e = \frac{(181+20)(181+20) + (20+84)(20+84)}{305^2} = 0.55057,$$
(5)

$$k = \frac{0.86885 - 0.55057}{1 - 0.55057} = 0.708191 \tag{6}$$

An analysis of Equation (5) demonstrates that the degree of concordance between the solutions provided by clinicians and the hybrid classifier reached 87% of the total test cases, which is considered a very high indicator. Additionally, the number of instances in which the hybrid classifier adjusted the AMD stage was comparable to its precautionary conclusions based on expert opinions, accounting for 13% of the total test set. According to clinicians, it would be advisable for the algorithm to show increased sensitivity to the stages of AMD that pose the greatest threat to the patient's visual function. This would enable the diagnostician working with the CDSS to assess potential risks or dismiss them in cases of apparent false alarms. We will focus on these aspects in the future to enhance the proposed method for developing a hybrid classifier.

Cohen's Kappa coefficient analysis shows a high agreement between expert assessments and the algorithm's responses, demonstrating its effectiveness as a clinical decisionsupport tool.

The results indicate that the issues of class imbalance and complexity in identifying the intermediate stage of AMD have been effectively resolved. Additionally, the conservative classification for classes N and V has been minimized. This framework can be seamlessly integrated into the CDSS to help determine when to adjust a patient's treatment plan. The hybrid classifier offers an accurate and interpretable assessment of the intermediate stage and early and late AMD outcomes.

4. Discussion

The developed hybrid classifier has demonstrated high efficiency and aligns well with expert assessments. The method's potential clinical applicability goes beyond the initial testing environment, opening up prospects for use in various ophthalmological conditions and diseases. This method is desirable for diseases with stages with different manifestation frequencies in patients, leading to a significant class imbalance during the training phase of classifier development. Integrating these classifiers into the CDSS element will help diagnosticians enhance diagnostic accuracy and optimize workflow management. This approach eliminates the need for a detailed marking of medical image sets while also emphasizing the importance of gathering sufficient examples of rare disease stages. Furthermore, the classifier's ability to accurately differentiate between sub-stages of or variations in ophthalmic conditions can facilitate earlier interventions, potentially preventing the progression to more severe forms of the disease and ultimately improving patient treatment outcomes.

However, expanding clinical applicability requires solving several problems related to integration into clinical practice. One major challenge is ensuring an equal or comparable amount of data from various tomograph models, as OCT images have different visualization features. Machine learning models included in the hybrid classifier must operate with a sufficient set of visual features that differ from each other to generalize them into one class. This presents a significant limitation for developing the hybrid classifier, as it cannot analyze OCT images obtained from tomographs that were not part of the training sample. The same applies to the visual quality of OCT images, which can decrease the performance of the hybrid classifier due to the suboptimal image quality often encountered in real clinical settings. Reducing the criticality of this factor can be achieved by introducing reliable preprocessing methods and including mechanisms for evaluating and managing image quality. This can increase the classifier's stability, something to which future research efforts will be devoted in further improving the presented method of developing hybrid classifiers.

One of the challenges in integrating the proposed method for developing a hybrid classifier is the "black box" nature of the deep learning models it utilizes. This lack of transparency can reduce trust among clinicians. Although the classifier demonstrates high efficiency, its interpretability issues may make clinicians hesitant to rely on automated recommendations entirely. To address this concern, future work will focus on iterating the algorithms to enhance the proposed method, incorporating techniques from explainable artificial intelligence. This approach will help clinicians better understand the classifier's decision-making process.

When comparing the effectiveness and methodological approaches of the presented hybrid classifier for identifying AMD stages with previous studies, several notable differences and similarities emerge. The proposed hybrid classifier exhibits impressive performance metrics, achieving an accuracy of over 96%, along with sensitivity and specificity levels surpassing those reported in prior research [29]. Additionally, the hybrid classifier demonstrates advanced recognition capabilities, as evidenced by its high F1-score and AUC values. While the authors of [29] achieved commendable accuracy with CNN-based screening, they encountered limitations when working with unbalanced classes, often resulting in biased predictions favoring the majority classes. Unlike reference [29], where traditional deep convolutional neural networks were used and designed mainly for processing fundus images, the method described in this article combines hybrid and hierarchical classification approaches, effectively addressing the class imbalance characteristic of AMD datasets. This issue was also noted in [57], where the class imbalance was addressed by accumulating additional data and creating repeated samples. This process required extra time from diagnosticians to develop the classification model presented. Notably, the overall performance indicators of our hybrid classifier surpassed those reported in [57].

In conclusion, this study demonstrates the possibility and potential of using the CDSS based on the hybrid classifier for monitoring changes in the management plan for patients with AMD. With further development and clinical validation, this system may simplify the treatment of AMD and enable earlier intervention, thereby improving the quality of healthcare services.

5. Conclusions

In this paper, we propose a hybrid approach for developing a CDSS aimed at determining the appropriate timing for changing the treatment plan for patients with AMD based on the analysis of OCT images. The architecture of our hybrid classifier includes a global classifier, several local binary classifiers, and a regressor. This developed architecture effectively addresses challenges commonly faced by machine learning models that utilize medical datasets collected during clinical practice. Specifically, it deals with visual similarities and overlaps among predictors in medical images, particularly the difficulty in detecting the intermediate stage of AMD amid early and late stages. It also addresses class imbalance, which presents challenges in identifying the stages of GA and SF. As a result of our hybrid approach to CDSS development, we achieved significant improvements in accuracy, sensitivity, and specificity, and an F1-score exceeding 0.90 across all AMD stages. This performance enhancement surpassed the average results of the original CNN-based four-layer encoder architecture by more than 20%. Additionally, we calculated Cohen's Kappa coefficient to evaluate the level of agreement between the algorithm's outcomes and expert assessments regarding the severity of the intermediate stage of AMD. The calculated value of 0.708 indicates a high level of agreement between the algorithm's results and the experts' evaluations.

Author Contributions: Conceptualization, E.A.L.; methodology, E.A.L. and E.S.Y.; software, E.A.L. and E.S.Y.; validation, R.R.I. and G.M.I.; formal analysis, G.M.I.; investigation, E.A.L., E.S.Y., R.R.I. and G.M.I.; resources, T.R.M., R.R.I. and G.M.I.; data curation, R.R.I. and G.M.I.; writing—original draft preparation, E.A.L.; writing—review and editing, E.A.L., E.P.G. and R.V.K.; visualization, E.A.L. and E.S.Y.; supervision, E.P.G. and G.M.I.; project administration, T.R.M. and R.V.K.; funding acquisition, E.P.G. and R.V.K. All authors have read and agreed to the published version of the manuscript.

Funding: The research is supported by the Ministry of Science and Higher Education of the Russian Federation within the state assignment for UUST (agreement № 075-03-2024-123/1 dated 15 February 2024) and conducted in the research laboratory "Sensor systems based on integrated photonics devices" of the Eurasian Scientific and Educational Center.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Restrictions apply to the availability of these data. Data were obtained from Optimed Laser Vision Restoration Center (Ufa, Russia) and are available from the authors with the permission of Optimed.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AMD	닾ͳͳͳͳͳͳͳͳ
GA	ই) 1999 1999 1999 1999 1999 1999 1997 1977
MNV	"
SF	ள்
anti-VEGF	ه المالي مالي

CV	ಯ
ML	""
OCT	ё
CDSS	ë """"
CNN	""
LPA	銹

Appendix A

Appendix A.1. Structure of the Hybrid Classifier



References

- Stahl, A. The Diagnosis and Treatment of Age-Related Macular Degeneration. *Dtsch. Ärzteblatt Int.* 2020, 117, 513–519. [CrossRef]
 [PubMed]
- Bressler, N.M. Antiangiogenic Approaches to Age-Related Macular Degeneration Today. *Ophthalmology* 2009, 116, S15–S23. [CrossRef]
- 3. Yu, J.; Ba, J.; Peng, R.; Xu, D.; Li, Y.; Shi, H.; Wang, Q. Intravitreal anti-VEGF injections for treating wet age-related macular degeneration: A systematic review and meta-analysis. *Drug Des. Dev. Ther.* **2015**, 2015, 5397–5405. [CrossRef] [PubMed]

- 4. Lee, Y.; Ahn, E.J.; Hussain, A. Saponin-Mediated Rejuvenation of Bruch's Membrane: A New Strategy for Intervention in Dry Age-Related Macular Degeneration (AMD). *Recent Adv. New Perspect. Manag. Macular Degener.* **2022**. [CrossRef]
- Arrigo, A.; IRCCS San Raffaele Scientific Institute, Vita-Salute San Raffaele University, Milan, Italy.; Bandello, F.; IRCCS San Raffaele Scientific Institute, Vita-Salute San Raffaele University, Milan, Italy. Towards the Development of Longer and More Efficacious Therapies for Wet and Dry Age-related Macular Degeneration. US Ophthalmic Rev. 2022, 16, 30. [CrossRef]
- Wong, T.Y.; Lanzetta, P.; Bandello, F.; Eldem, B.; Navarro, R.; Lövestam-Adrian, M.; Loewenstein, A. Current Concepts and Modalities for Monitoring the Fellow Eye in Neovascular Age-Related Macular Degeneration: An Expert Panel Consensus. *Retina* 2020, 40, 599–611. [CrossRef]
- Ferris, F.L.; Wilkinson, C.; Bird, A.; Chakravarthy, U.; Chew, E.; Csaky, K.; Sadda, S.R. Clinical Classification of Age-related Macular Degeneration. *Ophthalmology* 2013, 120, 844–851. [CrossRef]
- 8. Waugh, N.; Loveman, E.; Colquitt, J.; Royle, P.; Yeong, J.L.; Hoad, G.; Lois, N. Treatments for Dry Age-Related Macular Degeneration and Stargardt Disease: A Systematic Review. *Health Technol. Assess.* **2018**, *22*, 27. [CrossRef]
- 9. Garcia-Layana, A.; Cabrera-López, F.; García-Arumí, J.; Arias-Barquet, L.; Ruiz-Moreno, J.M. Early and intermediate age-related macular degeneration: Update and clinical review. *Clin. Interv. Aging* **2017**, *12*, 1579–1587. [CrossRef]
- 10. Morris, B.; Imrie, F.; Armbrecht, A.M.; Dhillon, B. Age-related macular degeneration and recent developments: New hope for old eyes? *Postgrad. Med. J.* **2007**, *83*, 301–307. [CrossRef]
- LlorenteGonzález, S.; Hernandez, M.; GonzálezZamora, J.; BilbaoMalavé, V.; FernándezRobredo, P.; SaenzdeViteri, M.; BarrioBarrio, J.; RodríguezCid, M.J.; Donate, J.; Ascaso, F.J.; et al. The role of retinal fluid location in atrophy and fibrosis evolution of patients with neovascular agerelated macular degeneration longterm treated in real world. *Acta Ophthalmol.* 2022, 100, e521–e531. [CrossRef] [PubMed]
- Regillo, C.; Nijm, L.; Shechtman, D.; Kaiser, P.; Karpecki, P.; Ryan, E.; Ip, M.; Yeu, E.; Kim, T.; Rafieetary, M.; et al. Considerations for the Identification and Management of Geographic Atrophy: Recommendations from an Expert Panel. *Clin. Ophthalmol.* 2024, 18, 325–335. [CrossRef] [PubMed]
- 13. Zhang, J.; Sheng, X.; Ding, Q.; Wang, Y.; Zhao, J.; Zhang, J. Subretinal fibrosis secondary to neovascular age-related macular degeneration: Mechanisms and potential therapeutic targets. *Neural Regen. Res.* **2025**, *20*, 378–393. [CrossRef]
- 14. Fabre, M.; Mateo, L.; Lamaa, D.; Baillif, S.; Pagès, G.; Demange, L.; Ronco, C.; Benhida, R. Recent Advances in Age-Related Macular Degeneration Therapies. *Molecules* **2022**, *27*, 5089. [CrossRef] [PubMed]
- 15. Lombardo, M.; Serrao, S.; Lombardo, G. Challenges in Age-Related Macular Degeneration: From Risk Factors to Novel Diagnostics and Prevention Strategies. *Front. Med.* **2022**, *9*, 887104. [CrossRef] [PubMed]
- Hagag, A.M.; Kaye, R.; Hoang, V.; Riedl, S.; Anders, P.; Stuart, B.; Traber, G.; Appenzeller-Herzog, C.; Schmidt-Erfurth, U.; Bogunovic, H.; et al. Systematic review of prognostic factors associated with progression to late age-related macular degeneration: Pinnacle study report 2. *Surv. Ophthalmol.* 2024, 69, 165–172. [CrossRef]
- 17. Danzig, C.J.; Khanani, A.M.; Loewenstein, A. C5 inhibitor avacincaptad pegol treatment for geographic atrophy: A comprehensive review. *Immunotherapy* **2024**, *16*, 779–790. [CrossRef] [PubMed]
- 18. Xu, Y.; Feng, Y.; Zou, R.; Yuan, F.; Yuan, Y. Silencing of YAP attenuates pericytemyofibroblast transition and subretinal fibrosis in experimental model of choroidal neovascularization. *Cell Biol. Int.* **2022**, *46*, 1249–1263. [CrossRef]
- Ho, A.C.; Kleinman, D.M.; Lum, F.C.; Heier, J.S.; Lindstrom, R.L.; Orr, S.C.; Chang, G.C.; Smith, E.L.; Pollack, J.S. Baseline Visual Acuity at Wet AMD Diagnosis Predicts Long-Term Vision Outcomes: An Analysis of the IRIS Registry. *Ophthalmic Surg. Lasers Imaging Retin.* 2020, *51*, 633–639. [CrossRef] [PubMed]
- Shamsnajafabadi, H.; Daftarian, N.; Ahmadieh, H.; Soheili, Z.S. Pharmacologic Treatment of Wet Type Age-related Macular Degeneration; Current and Evolving Therapies. *Arch. Iran. Med.* 2017, 20, 525–537.
- Flores, R.; Carneiro, Â.; Tenreiro, S.; Seabra, M.C. Retinal Progression Biomarkers of Early and Intermediate Age-Related Macular Degeneration. *Life* 2021, 12, 36. [CrossRef]
- Lad, E.M.; Finger, R.P.; Guymer, R. Biomarkers for the Progression of Intermediate Age-Related Macular Degeneration. *Ophthalmol. Ther.* 2023, 12, 2917–2941. [CrossRef] [PubMed]
- 23. Boopathiraj, N.; Wagner, I.V.; Dorairaj, S.K.; Miller, D.D.; Stewart, M.W. Recent Updates on the Diagnosis and Management of Age-Related Macular Degeneration. *Mayo Clin. Proc. Innov. Qual. Outcomes* **2024**, *8*, 364–374. [CrossRef] [PubMed]
- Sheeba, T.M.; Raj, S.A.A.; Anand, M. Computational Intelligent Techniques and Analysis on Automated Retinal Layer Segmentation and Detection of Retinal Diseases. In Proceedings of the 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 17 March 2023; IEEE: New York, NY, USA, 2023; pp. 2213–2219. [CrossRef]
- 25. Koseoglu, N.D.; Grzybowski, A.; Liu, T.Y.A. Deep Learning Applications to Classification and Detection of Age-Related Macular Degeneration on Optical Coherence Tomography Imaging: A Review. *Ophthalmol. Ther.* **2023**, *12*, 2347–2359. [CrossRef]

- Burlina, P.M.; Joshi, N.; Pekala, M.; Pacheco, K.D.; Freund, D.E.; Bressler, N.M. Automated Grading of Age-Related Macular Degeneration From Color Fundus Images Using Deep Convolutional Neural Networks. *JAMA Ophthalmol.* 2017, 135, 1170. [CrossRef] [PubMed]
- Leingang, O.; Riedl, S.; Mai, J.; Reiter, G.S.; Faustmann, G.; Fuchs, P.; Scholl, H.P.N.; Sivaprasad, S.; Rueckert, D.; Lotery, A.; et al. Automated deep learning-based AMD detection and staging in real-world OCT datasets (PINNACLE study report 5). *Sci. Rep.* 2023, 13, 19545. [CrossRef] [PubMed]
- Tan, J.H.; Bhandary, S.V.; Sivaprasad, S.; Hagiwara, Y.; Bagchi, A.; Raghavendra, U.; Krishna Rao, A.; Raju, B.; Shetty, N.S.; Gertych, A.; et al. Age-related Macular Degeneration detection using deep convolutional neural network. *Future Gener. Comput. Syst.* 2018, *87*, 127–135. [CrossRef]
- Govindaiah, A.; Hussain, M.A.; Smith, R.T.; Bhuiyan, A. Deep convolutional neural network based screening and assessment of age-related macular degeneration from fundus images. In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018; IEEE: New York, NY, USA, 2018; pp. 1525–1528. [CrossRef]
- Abd El-Khalek, A.A.; Balaha, H.M.; Alghamdi, N.S.; Ghazal, M.; Khalil, A.T.; Abo-Elsoud, M.E.A.; El-Baz, A. A concentrated machine learning-based classification system for age-related macular degeneration (AMD) diagnosis using fundus images. *Sci. Rep.* 2024, 14, 2434. [CrossRef]
- Bogunovic, H.; Montuoro, A.; Baratsits, M.; Karantonis, M.G.; Waldstein, S.M.; Schlanitz, F.; Schmidt-Erfurth, U. Machine Learning of the Progression of Intermediate Age-Related Macular Degeneration Based on OCT Imaging. *Investig. Ophthalmol. Vis. Sci.* 2017, *58*, BIO141–BIO150. [CrossRef]
- Yellapragada, B.; Hornhauer, S.; Snyder, K.; Yu, S.; Yiu, G. Unsupervised Deep Learning for Grading of Age-Related Macular Degeneration Using Retinal Fundus Images. *arXiv* 2020, arXiv:2010.11993. [CrossRef]
- 33. Yim, J.; Chopra, R.; Spitz, T.; Winkens, J.; Obika, A.; Kelly, C.; Askham, H.; Lukic, M.; Huemer, J.; Fasler, K.; et al. Predicting conversion to wet age-related macular degeneration using deep learning. *Nat. Med.* **2020**, *26*, 892–899. [CrossRef] [PubMed]
- Liefers, B.; Taylor, P.; Alsaedi, A.; Bailey, C.; Balaskas, K.; Dhingra, N.; Egan, C.A.; Rodrigues, F.G.; Gonzalo, C.G.; Heeren, T.F.; et al. Quantification of Key Retinal Features in Early and Late Age-Related Macular Degeneration Using Deep Learning. *Am. J. Ophthalmol.* 2021, 226, 1–12. [CrossRef] [PubMed]
- 35. Borrelli, E.; Serafino, S.; Ricardi, F.; Coletto, A.; Neri, G.; Olivieri, C.; Ulla, L.; Foti, C.; Marolo, P.; Toro, M.D.; et al. Deep Learning in Neovascular Age-Related Macular Degeneration. *Medicina* **2024**, *60*, 990. [CrossRef]
- Holland, R.; Kaye, R.; Hagag, A.M.; Leingang, O.; Taylor, T.R.; Bogunović, H.; Schmidt-Erfurth, U.; Scholl, H.P.; Rueckert, D.; Lotery, A.J.; et al. Deep Learning–Based Clustering of OCT Images for Biomarker Discovery in Age-Related Macular Degeneration (PINNACLE Study Report 4). *Ophthalmol. Sci.* 2024, 4, 100543. [CrossRef]
- Yildirim, K.; Al-Nawaiseh, S.; Ehlers, S.; Schießer, L.; Storck, M.; Brix, T.; Eter, N.; Varghese, J. U-Net-Based Segmentation of Current Imaging Biomarkers in OCT-Scans of Patients with Age Related Macular Degeneration. *Stud. Health Technol. Inform.* 2023, 302, 947–951. [CrossRef]
- Biloborodova, T.; Skarga-Bandurova, I.; Koverha, M.; Skarha-Bandurov, I.; Yevsieieva, Y. A Learning Framework for Medical Image-Based Intelligent Diagnosis from Imbalanced Datasets. *Stud. Health Technol. Inform.* 2021, 287, 13–17. [CrossRef]
- Prudhomme, C.; Schaffert, M.; Ponciano, J.J. Deep Learning Datasets Challenges for Semantic Segmentation—A Survey. In Proceedings of the Lecture Notes in Informatics, Berlin, Germany, 26–29 September 2023.
- Yang, S.; Zhou, X.; Wang, J.; Xie, G.; Lv, C.; Gao, P.; Lv, B. Unsupervised Domain Adaptation for Cross-Device OCT Lesion Detection via Learning Adaptive Features. In Proceedings of the 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), Iowa City, IA, USA, 3–7 April 2020; IEEE: New York, NY, USA, 2020; pp. 1570–1573. [CrossRef]
- Kim, W.; Kanezaki, A.; Tanaka, M. Unsupervised Learning of Image Segmentation Based on Differentiable Feature Clustering. IEEE Trans. Image Process. 2020, 29, 8055–8068. [CrossRef]
- 42. Brunekreef, J.; Marcus, E.; Sheombarsing, R.; Sonke, J.J.; Teuwen, J. Kandinsky Conformal Prediction: Efficient Calibration of Image Segmentation Algorithms. *arXiv* 2023, arXiv:2311.11837. http://arxiv.org/abs/2311.11837.
- 43. Orzeł, Z.; Kosior-Romanowska, A.; Bednarczuk, P. Methods of medical image segmentation analysis to improve the effectiveness of diagnosing lung diseases. *J. Mod. Sci.* 2024, *57*, 609–621. [CrossRef]
- 44. Sajjadi, Z.; Esmaeili, M.; Ghobaei-Arani, M.; Minaei, B. A Hybrid Clustering Approach for link prediction in Heterogeneous Information Networks. *Knowl. Inf. Syst.* 2023, *65*, 4905–4937. [CrossRef]
- 45. Mavaie, P.; Holder, L.; Skinner, M.K. Hybrid deep learning approach to improve classification of low-volume high-dimensional data. *BMC Bioinform.* **2023**, *24*, 419. [CrossRef]
- 46. Chou, Y.Y.; Shapiro, L.G. A hierarchical multiple classifier learning algorithm. Pattern Anal. Appl. 2003, 6, 150–168. [CrossRef]
- Silla, C.N.; Freitas, A.A. A survey of hierarchical classification across different application domains. *Data Min. Knowl. Discov.* 2011, 22, 31–72. [CrossRef]

- Lin, Y.; Liu, H.; Zhao, H.; Hu, Q.; Zhu, X.; Wu, X. Hierarchical Feature Selection Based on Label Distribution Learning. *IEEE Trans. Knowl. Data Eng.* 2022, 35, 5964–5976. [CrossRef]
- 49. Rezende, P.M.; Xavier, J.S.; Ascher, D.B.; Fernandes, G.R.; Pires, D.E.V. Evaluating hierarchical machine learning approaches to classify biological databases. *Briefings Bioinform.* 2022, 23, bbac216. [CrossRef] [PubMed]
- 50. Ojo, M.O.; Zahid, A. Improving Deep Learning Classifiers Performance via Preprocessing and Class Imbalance Approaches in a Plant Disease Detection Pipeline. *Agronomy* **2023**, *13*, 887. [CrossRef]
- Long, Q.Y.; Wang, F.Y.; Hu, Y.; Gao, B.; Zhang, C.; Ban, B.H.; Tian, X.B. Development of the Interpretable Typing Prediction Model for Osteosarcoma and Chondrosarcoma Based on Machine Learning and Radiomics: A Multicenter Retrospective Study. *Front. Med.* 2024, 11, 1497309. [CrossRef] [PubMed]
- 52. Li, Y.; Qian, G.; Jiang, X.; Jiang, Z.; Wen, W.; Zhang, S.; Li, K.; Lao, Q. Hierarchical-Instance Contrastive Learning for Minority Detection on Imbalanced Medical Datasets. *IEEE Trans. Med. Imaging* **2023**, *43*, 416–426. [CrossRef]
- 53. Qi, J.; Yang, H.; Kong, Z. A review of traditional image segmentation methods. In Proceedings of the 5th International Conference on Computer Information Science and Application Technology (CISAT 2022), Chongqing, China, 20 October 2022; Zhao, F., Ed.; SPIE: Washington, DC, USA; p. 127. [CrossRef]
- Lindeberg, T. Scale-Space. In Wiley Encyclopedia of Computer Science and Engineering; John Wiley & Sons: Hoboken, NJ, USA, 2009; pp. 2495–2504.
- 55. Archana, R.; Jeevaraj, P.S.E. Deep learning models for digital image processing: A review. Artif. Intell. Rev. 2024, 57, 11. [CrossRef]
- Zeng, S.; Zhao, Y.; Li, S. Comparison Between the Traditional and Deep Learning Algorithms on Image Matching. In Proceedings of the 2022 IEEE/ACIS 22nd International Conference on Computer and Information Science (ICIS), Zhuhai, China, 26 June 2022; IEEE: New York, NY, USA, 2022; pp. 182–186. [CrossRef]
- 57. Bria, A.; Marrocco, C.; Tortorella, F. Addressing class imbalance in deep learning for small lesion detection on medical images. *Comput. Biol. Med.* **2020**, 120, 103735. [CrossRef] [PubMed]
- 58. Brigato, L.; Barz, B.; Iocchi, L.; Denzler, J. Image Classification With Small Datasets: Overview and Benchmark. *IEEE Access* 2022, 10, 49233–49250. [CrossRef]
- Diab, M.S.; Rodriguez-Villegas, E. Performance Evaluation of Embedded Image Classification Models Using Edge Impulse for Application on Medical Images. In Proceedings of the 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Scotland, UK, 11–15 July 2022; pp. 2639–2642. [CrossRef]
- 60. Jude Chukwura Obi. A comparative study of several classification metrics and their performances on data. *World J. Adv. Eng. Technol. Sci.* 2023, *8*, 308–314. [CrossRef]
- 61. Westfall, P.H.; Troendle, J.F.; Pennello, G. Multiple McNemar Tests. Biometrics 2010, 66, 1185. [CrossRef]
- 62. Lee, D.K. Alternatives to P Value: Confidence Interval and Effect Size. Korean J. Anesthesiol. 2016, 69, 555–562. [CrossRef]
- 63. Abdin, A.D.; Devenijn, M.; Fulga, R.; Langenbucher, A.; Seitz, B.; Kaymak, H. Prevalence of Geographic Atrophy in Advanced Age-Related Macular Degeneration (AMD) in Daily Practice. *J. Clin. Med.* **2023**, *12*, 4862. [CrossRef] [PubMed]
- 64. Cheng, A.M.S.; Chalam, K.V.; Brar, V.S.; Yang, D.T.Y.; Bhatt, J.; Banoub, R.G.; Gupta, S.K. Recent Advances in Imaging Macular Atrophy for Late-Stage Age-Related Macular Degeneration. *Diagnostics* **2023**, *13*, 3635. [CrossRef] [PubMed]
- 65. Zhu, X.; Ghahramani, Z. *Learning from Labeled and Unlabeled Data with Label Propagation*; Technical Report CMU-CALD-02-107; Carnegie Mellon University: Pittsburgh, PA, USA, 2002.
- 66. Hodson, T.O. Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geosci. Model Dev.* 2022, 15, 5481–5487. [CrossRef]
- Pacurari, A.C.; Bhattarai, S.; Muhammad, A.; Avram, C.; Mederle, A.O.; Rosca, O.; Bratosin, F.; Bogdan, I.; Fericean, R.M.; Biris, M.; et al. Diagnostic Accuracy of Machine Learning AI Architectures in Detection and Classification of Lung Cancer: A Systematic Review. *Diagnostics* 2023, 13, 2145. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.