Age Prediction and Categorization from Retinal Fundus Images Using Residual Neural Networks

Mehmet Aytuğ Yürük
Dept. of Artificial Intelligence and Data Engineering
İstanbul University
İstanbul, Türkiye
myuruk1@ogr.iu.edu.tr

Abbas Memiş
Department of Computer Engineering
İstanbul University
İstanbul, Türkiye
abbas.memis@istanbul.edu.tr

***Abstract*—High-accuracy age prediction from retinal images is an important research topic in biometric and clinical analysis, but it is also a challenging task. This paper presents a deep learning-based study to predict the age of individuals and automatically categorize them into different age groups using retinal fundus images. Within the scope of the study, a comparative analysis was performed using variants of the Residual Neural Network (ResNet) architecture on an open and recent retinal image dataset. In the experimental studies, five different architectures were tested: ResNet-18, ResNet-34, ResNet-50, ResNet-101, and ResNet-152. Furthermore, various data imbalance mitigation approaches and image preprocessing strategies were employed to improve and optimize model performance. The experimental findings revealed that while ResNet variants did not show significant performance differences, combining the models with appropriate preprocessing methods could lead to a modest improvement in performance. As a result of the performance tests carried out, while promising results were observed in general, the best results were achieved with the ResNet-101 architecture (mean absolute error (MAE) of 5.02) in age prediction and the ResNet-34 architecture (F-measure of 0.7165) in age categorization.**

***Keywords*—Age prediction, age categorization, retinal images, fundus images, residual neural network, deep learning.**

I. Introduction

Retinal imaging has a significant role in the analysis of conditions occurring in the retina, either directly or indirectly due to various health-related factors such as diabetes [1]. In the diagnosis, prognosis and follow up of retinal vascular diseases, a wide variety of medical imaging modalities have been used, including color fundus photography, optical coherence tomography, dynamic diagnostic options such as fluorescein angiography, and optical coherence tomography angiography [2]. Retinal images are used not only in the analysis of retinal pathologies [3], but also in biometric authentication [4] and health risk factors [5]. In this context, retinal images have a wide range of applications.

The remarkable progress in deep learning over the past decade has led to significant advancements in medical image analysis. Particularly, image-based analyses make it possible to automatically extract and evaluate the physiological characteristics of individuals quantitatively. In this context, age prediction is an emerging field of research within biometric and clinical analysis. However, high anatomical variability among individuals and variations in image quality make it difficult to directly correlate learning representations with age and limit the generalization ability of models. This makes the task more difficult and limits accurate age prediction. Retinal fundus images are a kind of medical image resource that allows for direct and non-invasive observation of microvascular structures [5]. The presence of age-related structural and functional changes reflected in the retina makes retinal images significant for the task of age prediction.

In this paper, a deep learning-based study is presented for the prediction of individuals' ages and the automatic identification of age categories using retinal images. Within the scope of the study, a comparative analysis was performed using variants of the Residual Neural Network (ResNet) architecture on an open and recent retinal image dataset. Within the scope of the experimental studies, five different architectures were tested: ResNet-18, ResNet-34, ResNet-50, ResNet-101, and ResNet-152. In addition, various data imbalance mitigation approaches and image preprocessing strategies were also employed to improve and optimize model performance. The rest of the paper is organized as follows: Section II addresses related works. Section III describes the dataset used in the study. Methodological details are explained in Section IV. The results observed in the experimental studies are reported in Section V. Finally, the paper is concluded in Section VI by highlighting the findings.

II. Related Works

This section discusses some outstanding and recent age prediction studies using images of the retinal fundus. Although the literature on age prediction using retinal fundus images is limited, it is notable that there has been a significant increase in studies on this topic in recent years.

In [6], Liu et al. carried out a deep learning-based pilot study to investigate the potential of using fundus images to estimate biological age. Using a VGG-19 network, they achieved a mean absolute error (MAE) of 3.73 years on a dataset consisting of 12,000 fundus images from healthy subjects. In [7], Kim et al. proposed a ResNet-152-based model to analyze the effects of hypertension, diabetes, and smoking on age and sex prediction from retinal fundus images. They built age and sex prediction models using 219,302 fundus

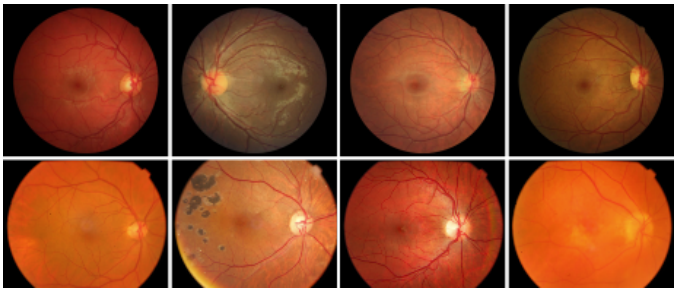

Fig. 1. Image samples from the retina dataset [18].

TABLE I
NUMBER OF SAMPLES PER AGE CATEGORY IN THE TRAIN, VALIDATION, AND TEST SETS.

Age category (age range)	Train	Validation	Test	Total
Pediatric (0-17)	202	44	45	291
Young Adult (18-39)	1015	218	214	1447
Middle Age (40-59)	2061	443	442	2946
Senior (60-74)	2443	528	513	3484
Elderly (75+)	1181	260	248	1689
Total	6902	1493	1462	9857

images from normal participants, and trained models were evaluated using fundus images from normal participants and abnormal individuals (hypertension, diabetes and smoking). In the related study, the ages of the subjects were predicted with a MAE ranging from 2.5 to 4.0. In another study [8], Wen et al. aimed to assess deep learning models for age prediction from fundus images of normal patients and patients with ophthalmic diseases. For this purpose, they adopted the CNN architecture of ShuffleNet-V2 and used an inhouse color fundus image dataset, which consists of 14,957 images of Chinese ophthalmic patients. In experimental studies, they observed MAE values ranging from 3 to 6.5. In addition to the related works summarized in the previous lines, some other notable research on age estimation from retinal fundus images has also been conducted in recent years [9]–[14]. Furthermore, there are also studies being conducted to predict ages from Optical Coherence Tomography (OCT) images [15]–[17].

III. MATERIALS

The proposed study was performed on a recent and publicly available retinal fundus image dataset [18]. The dataset consists of 9,857 color retinal fundus images and the corresponding chronological age values and age category labels (Pediatric (0-17), Young Adult (18-39), Middle Age (40-59), Senior (60-74), Elderly (75+)) for each image. It contains a total of 9,857 retinal fundus images, which have been pre-split into training (6,902), validation (1,493), and test (1,462) subsets on a patient basis. The right and left eye images belonging to the same individual are prevented from being in different subsets, aiming to evaluate the generalization performance of the models in a more reliable manner. Fig. 1 shows sample retinal fundus images from the related dataset.

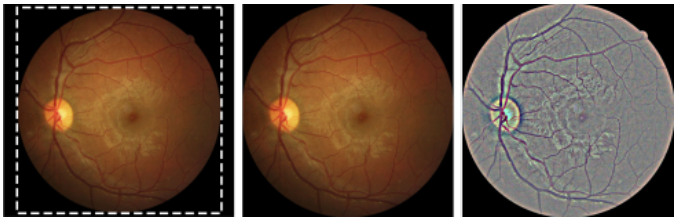

Fig. 2. Preprocessing steps on retinal images: 1) Cropping retina region (**left**), 2) Zero padding and resizing (**center**), 3) Graham filtering (**right**).

In Table I, the distribution of samples in the training, validation, and test sets are given according to age category. The ages of individuals in the dataset range from 5 to 94 years old, and as can be followed from Table I, the age distribution is not uniform. This non-uniform distribution, as might be expected, can also lead to the data imbalance problem.

IV. METHODS

A. Image Preprocessing

A three-stage preprocessing procedure was applied to the retinal fundus images in the dataset being used. All retinal images were scaled to a resolution of 224×224 pixels to ensure that the experimental results could be analyzed, compared, and interpreted within a consistent framework. However, the scaling was not performed directly by means of resizing. In order to preserve the geometric integrity of the retinal structure, the retinal region was positioned (aligned) to fill a 224×224 pixel canvas maximally without distorting its aspect ratio. The empty areas resulting from the scaling were filled with black padding. After resizing and alignment, the retinal images were processed with the Graham filter [19]. This filter, which is frequently used in the retinal image analysis literature, aims to make retinal vessel structures and fine details more distinct and apparent. The image preprocessing steps are also illustrated in Fig. 2.

B. Deep Learning Model

In the proposed study, Residual Neural Network (ResNet) [20], a Convolutional Neural Network (CNN) based architecture, was used to perform age prediction. ResNet is widely used in deep learning-based image analysis thanks to its structure, which enables deep networks to be trained more stably through residual connections. Within the scope of the study, five different variants of the related architecture were analyzed: ResNet-18, ResNet-34, ResNet-50, ResNet-101, and ResNet-152.

The training of all ResNet variant models was initiated using pre-trained weights on the ImageNet dataset [21]. In order to enable the models to learn retina-specific representations end-to-end, all layers were unfrozen and full fine-tuning was performed. In this way, the learning capacity of each architecture variant was optimized directly on the retina data rather than on predefined features.

Due to the structural characteristics of the ResNet architecture, the size of the feature vectors generated by models

of different depths in the final stage varies: the size is 512 for ResNet-18/34 and 2048 for ResNet-50/101/152. In this regard, the regression layer was specifically customized to be adapted to the size of the feature vector produced by the backbone architecture. Other than that, the structural design was preserved in all models. By using this approach, it was aimed to ensure that the performance differences obtained were primarily due to the depth of the ResNet architecture, rather than the regression layer or the training strategy.

During the training process, all images were normalized using the mean and standard deviation values from the ImageNet dataset [21]. Additionally, data augmentation was performed to improve the generalization ability of the model and reduce the potential risk of overfitting. In this context, random horizontal flipping, small-angle rotation, limited brightness-contrast changes in the color space, and Gaussian blurring operations were used.

C. Label Distribution Smoothing (LDS)

The data imbalance problem, which is also common in medical datasets, is likely to cause models to overfocus on classes with too many samples and therefore fail to learn optimally. As stated in Section III, the class distribution of samples in the dataset used in the proposed study is not well-balanced. To address the potential issue that this imbalance may cause, the Label Distribution Smoothing (LDS) strategy [22] has been adopted for the loss functions of the models.

Unlike the conventional class-based weighting approaches, the LDS method models the relationships between neighboring age values through a smoothed density function, taking into account the label distribution in the target space. In this context, instance-based weights are derived using the inverse of the density values obtained for each age label. During the training process, the loss function is rescaled using these weights, so that the contribution of samples belonging to less frequent age ranges to the learning process is increased, aiming for a more balanced and stable optimization.

In LDS approach, age labels are considered as discrete values, and an empirical label density $p(y)$ is defined for each age. The density is smoothed using a symmetric Gaussian kernel to obtain the effective label density:

$$\tilde{p}(y) = \sum_{y'} p(y')\, \mathcal{G}_\sigma(y - y'), \tag{1}$$

where $\mathcal{G}_\sigma(\cdot)$ denotes a one-dimensional Gaussian kernel with normalized values whose sum is equal to 1 and a standard deviation of σ. This operation aims to smooth sharp frequency variations in the label distribution by taking into account the contiguity between neighboring age values. Using the derived smoothed label density $\tilde{p}(y)$, the loss function is reweighted for each training sample according to the inverse density principle. Accordingly, the weight used for each sample is defined as follows:

$$w(y) = \frac{c}{\tilde{p}(y) + \epsilon}, \tag{2}$$

where c is a scaling factor used to balance the average magnitude of the weights, and ϵ refers to a small positive value added to ensure numerical consistency. Finally, the mini-batch weighted regression loss with LDS is calculated as follows:

$$\mathcal{L}_{\text{LDS}} = \frac{1}{m} \sum_{i=1}^{m} w(y_i)\, \mathcal{L}(\hat{y}_i, y_i), \tag{3}$$

where m denotes the mini-batch size, $\mathcal{L}(\hat{y}_i, y_i)$ refers to the basic regression loss function used, and $\hat{y}_i$ represents the age value predicted by the model.

D. Age Prediction and Categorization

In this study, estimation of individuals' chronological ages from retinal fundus images was addressed as a regression task. In the ResNet architectures used, last layer of the network was configured as a single-output linear regression head, and the model was trained to make age predictions directly. In this context, no additional model or classification head was used for categorization. However, to provide a more interpretable analysis of the regression outputs and to examine the model's ability to distinguish between different age groups, the predicted age values were subsequently categorized according to specific age ranges. The age ranges for categorization were defined as follows: 0-17 years old (pediatric), 18-39 years old (young adult), 40-59 years old (middle age), 60-74 years old (senior), and 75 years old and over (elderly). This categorical mapping was only performed during the evaluation phase; no classification loss or decision mechanism was incorporated into the training process of the model. The specified age range labels are already defined for all samples in the dataset.

E. Evaluation Metrics

In the study, age prediction was addressed as a regression task and age categorization as a classification task. Therefore, distinct evaluation metrics were employed for each task. To measure the difference between the age values predicted by the models and the actual age values, the mean absolute error (MAE) metric was utilized. The MAE formula is expressed in Equation 4.

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i| \tag{4}$$

In Equation 4, y_i represents the actual age value, $\hat{y}_i$ indicates the age value estimated by the model, and N denotes the total number of samples.

In addition to regression-based analyses, classification-based analyses were performed to analyze the prediction performance of the model in different age groups in more detail. For this purpose, the numerical age predictions produced by the model were categorized into age groups predefined in the dataset and converted into categorical labels. The aforementioned analyses do not indicate that the model was trained for a classification task, but are intended to evaluate the consistency of regression outputs across age groups. In this regard, common metrics such as accuracy (A), precision (P),

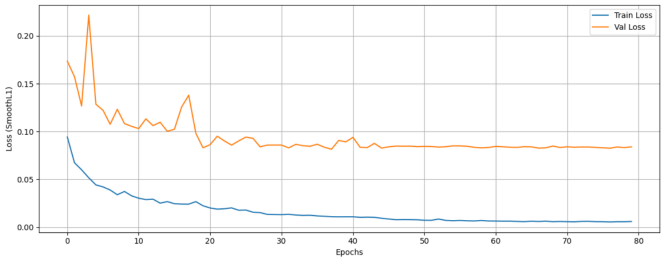

Fig. 3. The loss rates in the training and validation sets during the model training (ResNet-101 architecture and unfiltered color fundus images).

recall (R), and F-measure (F) were employed in classification-based evaluations. These metrics can be calculated via a class confusion matrix using the true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) as follows: A = (TP+TN)/(TP+TN+FP+FN), P = (TP)/(TP+FP), R = (TP)/(TP+FN) and F = (2*P*R)/(P+R).

F. Environmental Settings

The training and testing processes of the models used in the study were performed on hardware with Apple M4 chip architecture using the PyTorch deep learning library. The training was run for a total of 80 epochs, and the mini-batch size was set to 32. The number of epochs was determined based on the convergence of the models on the validation loss. Fig. 3 shows the variations in the loss function depending on the number of epochs in a typical training performed with the ResNet-101 architecture on unfiltered fundus images.

The AdamW algorithm [23] was used as the optimization method. Unlike the classical Adam optimization method, AdamW employs the weight decay parameter separately from the gradient-based parameter updates, thereby providing a more consistent and balanced regularization. This approach was particularly preferred for its ability to reduce the overfitting tendency of deep neural networks.

The Smooth L1 loss [24] has been used as the loss function. This loss function has an adaptive structure depending on the absolute value of the error of the predicted value. It enables more sensitive learning by penalizing errors quadratically for small prediction errors, while limiting the magnitude of gradients for large errors by applying a linear penalty. This feature enables a more stable and balanced learning process by preventing the training process from being dominated by the rare but large errors occurring in the age prediction task.

The initial learning rate is set to 1×10^{-4}. To increase the stability of the training process and adapt to plateaus in validation performance, the ReduceLROnPlateau learning rate scheduler is used, which reduces the learning rate by a factor of 0.5 if no improvement is observed in the validation loss for five consecutive epochs. This approach allows the model to learn quickly in the early stages and make finer adjustments in the later stages.

V. Results

In experimental studies, the performances of the ResNet architecture variants ResNet-18, ResNet-34, ResNet-50, ResNet-101, and ResNet-152 were analyzed comparatively. The results were reported separately for both unfiltered fundus images and those pre-processed with the Graham filter. Furthermore, to evaluate the performance of the models in more detail, (i) class-based metrics calculated for each age group and (ii) overall performance metrics calculated from all test samples were systematically reported. Hence, the performance of the models on different age groups and their overall performance were analyzed comprehensively.

A. Results on Non-filtered Images

The class-based performance results observed using unfiltered retinal images are presented in Table II. The μ values reported in this table were calculated by weighting the performance metrics for each age group using the ratio of samples in the test set. The class weights were calculated based upon the age distribution of the images in the test set and are given in detail in Table VI. Additionally, Table III reports the overall performance results on the entire test data.

The lowest error score on unfiltered images was achieved by the ResNet-101 architecture with a MAE of 5.09. Compared to shallow architectures, ResNet-101 yielded the lowest error values, while the deeper ResNet-152 architecture saturated in performance and resulted in an increase in error values. This finding suggests that model depth does not provide a linear performance increase in age estimation problem.

In Fig. 4, an age scatter plot illustrating the distribution of predicted ages for the ResNet-101 architecture is presented along with a class confusion matrix constructed for age categorization.

B. Results on Graham-filtered Images

Class-based performance results obtained using retina images processed with the Graham filter are reported in Table IV, and performance results calculated over the entire test dataset are reported in Table V.

The results in the related tables clearly show that the effect of filtering on model performance also varies according to the architectural depth. The lowest error value on filtered images was achieved by the ResNet-101 architecture with a MAE of 5.02. This value is also the best result among all experiments reported in the study. This finding indicates that retinal inputs with enhanced vascular structures and local tissue patterns can be leveraged more effectively.

However, in shallow models, the contribution of the Graham filter to performance appears to be limited. In these models, no significant improvement in error values is observed in some age groups, while small increases are observed in some cases. In the ResNet-152 architecture, it is seen that, similar to unfiltered images, no better performance is achieved compared to ResNet-101.

TABLE II

CLASS-BASED SUCCESS RATES FOR AGE PREDICTION AND AGE CATEGORIZATION IN NON-FILTERED IMAGES.

	Age category	MAE	Accuracy	Precision	Recall	F-measure
ResNet-18	Pediatric (0-17)	5.48	0.9863	0.9032	0.6222	0.7368
	Young Adult (18-39)	5.78	0.9419	0.8644	0.7150	0.7826
	Middle Age (40-59)	5.04	0.8276	0.6979	0.7579	0.7267
	Senior (60-74)	4.89	0.7736	0.6643	0.7173	0.6898
	Elderly (75+)	5.43	0.9001	0.7318	0.6492	0.6880
	μ	5.17	0.8426	0.7225	0.7148	0.7157
ResNet-34	Pediatric (0-17)	5.30	0.9836	0.8387	0.5778	0.6842
	Young Adult (18-39)	5.48	0.9405	0.8191	0.7617	0.7893
	Middle Age (40-59)	5.14	0.8269	0.7114	0.7195	0.7154
	Senior (60-74)	4.84	0.7633	0.6462	0.7193	0.6808
	Elderly (75+)	5.23	0.8988	0.7336	0.6331	0.6797
	μ	5.11	0.8382	0.7120	0.7066	0.7071
ResNet-50	Pediatric (0-17)	5.63	0.9850	0.9259	0.5556	0.6944
	Young Adult (18-39)	5.54	0.9432	0.8394	0.7570	0.7961
	Middle Age (40-59)	5.18	0.8297	0.7102	0.7376	0.7236
	Senior (60-74)	5.13	0.7599	0.6517	0.6784	0.6648
	Elderly (75+)	5.19	0.8926	0.6827	0.6855	0.6841
	μ	5.23	0.8373	0.7106	0.7052	0.7060
ResNet-101	Pediatric (0-17)	4.60	0.9808	0.7297	0.6000	0.6585
	Young Adult (18-39)	5.33	0.9384	0.8100	0.7570	0.7826
	Middle Age (40-59)	4.80	0.8242	0.6907	0.7579	0.7228
	Senior (60-74)	5.35	0.7770	0.6904	0.6608	0.6753
	Elderly (75+)	4.93	0.9063	0.7229	0.7258	0.7243
	μ	5.09	0.8431	0.7147	0.7134	0.7132
ResNet-152	Pediatric (0-17)	6.10	0.9822	0.8800	0.4889	0.6286
	Young Adult (18-39)	5.09	0.9419	0.8116	0.7850	0.7981
	Middle Age (40-59)	5.01	0.8297	0.6957	0.7760	0.7337
	Senior (60-74)	5.74	0.7599	0.6731	0.6140	0.6422
	Elderly (75+)	4.76	0.8912	0.6654	0.7218	0.6925
	μ	5.27	0.8368	0.7053	0.7025	0.7008

TABLE III

MEAN (WEIGHTED) SUCCESS RATES FOR AGE PREDICTION AND AGE CATEGORIZATION IN NON-FILTERED IMAGES.

Model	MAE	Accuracy	Precision	Recall	F-measure
ResNet-18	5.17	0.8426	0.7225	0.7148	0.7157
ResNet-34	5.11	0.8382	0.7120	0.7066	0.7071
ResNet-50	5.23	0.8373	0.7106	0.7052	0.7060
ResNet-101	5.09	0.8431	0.7147	0.7134	0.7132
ResNet-152	5.27	0.8368	0.7053	0.7025	0.7008

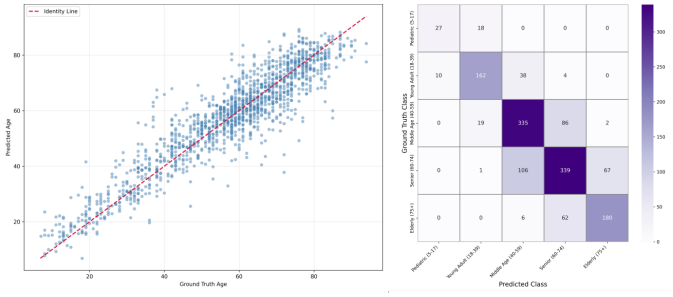

Fig. 4. Age scatter plot (ResNet-101) and class confusion matrix (ResNet-101) for non-filtered retinal images.

In Fig. 5, an age scatter plot illustrating the distribution of predicted ages using the ResNet-101 architecture on Graham-filtered images is presented along with a class confusion matrix constructed for age categorization.

TABLE IV

CLASS-BASED SUCCESS RATES FOR AGE PREDICTION AND AGE CATEGORIZATION IN IMAGES PROCESSED WITH THE GRAHAM FILTER.

	Age category	MAE	Accuracy	Precision	Recall	F-measure
ResNet-18	Pediatric (0-17)	5.17	0.9781	0.6857	0.5333	0.6000
	Young Adult (18-39)	5.32	0.9371	0.8177	0.7336	0.7734
	Middle Age (40-59)	4.93	0.8420	0.7221	0.7760	0.7481
	Senior (60-74)	5.41	0.7695	0.6648	0.6920	0.6781
	Elderly (75+)	5.39	0.8551	0.6770	0.6169	0.6456
	μ	5.24	0.8369	0.7072	0.7059	0.7053
ResNet-34	Pediatric (0-17)	4.88	0.9808	0.6889	0.6889	0.6889
	Young Adult (18-39)	5.70	0.9425	0.8385	0.7523	0.7931
	Middle Age (40-59)	4.82	0.8365	0.7051	0.7896	0.7449
	Senior (60-74)	5.53	0.7729	0.6851	0.6530	0.6687
	Elderly (75+)	5.24	0.9008	0.7137	0.6935	0.7035
	μ	5.27	0.8451	0.7185	0.7168	0.7165
ResNet-50	Pediatric (0-17)	5.80	0.9802	0.7353	0.5556	0.6329
	Young Adult (18-39)	5.27	0.9419	0.8449	0.7383	0.7880
	Middle Age (40-59)	4.66	0.8201	0.6660	0.8122	0.7319
	Senior (60-74)	5.41	0.7531	0.6502	0.6413	0.6457
	Elderly (75+)	5.62	0.8906	0.7245	0.5726	0.6396
	μ	5.21	0.8313	0.6987	0.6929	0.6912
ResNet-101	Pediatric (0-17)	5.18	0.9802	0.7353	0.5556	0.6329
	Young Adult (18-39)	5.03	0.9371	0.8050	0.7523	0.7778
	Middle Age (40-59)	4.65	0.8338	0.7002	0.7873	0.7412
	Senior (60-74)	5.06	0.7722	0.6705	0.6901	0.6801
	Elderly (75+)	5.55	0.8953	0.7340	0.6008	0.6608
	μ	5.02	0.8422	0.7119	0.7083	0.7081
ResNet-152	Pediatric (0-17)	5.82	0.9795	0.7419	0.5111	0.6053
	Young Adult (18-39)	5.39	0.9378	0.8220	0.7336	0.7753
	Middle Age (40-59)	4.74	0.8386	0.7077	0.7941	0.7484
	Senior (60-74)	5.35	0.7709	0.6745	0.6706	0.6725
	Elderly (75+)	5.26	0.8865	0.6752	0.6371	0.6556
	μ	5.17	0.8418	0.7083	0.7066	0.7056

TABLE V

MEAN (WEIGHTED) SUCCESS RATES FOR AGE PREDICTION AND AGE CATEGORIZATION IN IMAGES PROCESSED WITH THE GRAHAM FILTER.

Model	MAE	Accuracy	Precision	Recall	F-measure
ResNet-18	5.24	0.8369	0.7072	0.7059	0.7053
ResNet-34	5.27	0.8451	0.7185	0.7168	0.7165
ResNet-50	5.21	0.8313	0.6987	0.6929	0.6912
ResNet-101	5.02	0.8422	0.7119	0.7083	0.7081
ResNet-152	5.17	0.8418	0.7083	0.7066	0.7056

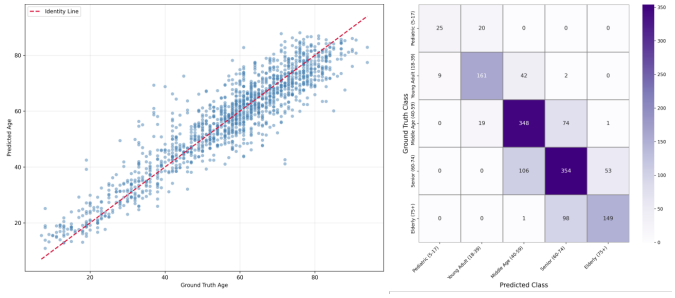

Fig. 5. Age scatter plot (ResNet-101) and class confusion matrix (ResNet-101) for Graham-filtered retinal images.

VI. CONCLUSIONS

In this paper, a deep learning-based study is presented for the prediction of individuals' ages and the automatic identification of age categories using retinal fundus images. Within

TABLE VI
DISTRIBUTION OF THE SAMPLES AND CLASS WEIGHTS IN THE TEST SET.

Age category (age range)	Number of test samples (S)	Class weight (S/T)
Pediatric (0–17)	45	0.0308
Young Adult (18–39)	214	0.1464
Middle Age (40–59)	442	0.3023
Senior (60–74)	513	0.3509
Elderly (75+)	248	0.1696
Total (T)	1462	1.0000

the scope of the study, a comparative analysis was performed using the ResNet-18, ResNet-34, ResNet-50, ResNet-101, and ResNet-152 on a publicly available and recent retinal image dataset.

The experimental results observed in the study clearly demonstrate that the increase in model depth improves performance in the age prediction task up to a certain point, but the relationship is not linear. While models with shallow depths yielded relatively high error values due to limited representation capacity, the medium-depth ResNet-101 architecture achieved the best results. In contrast, the deeper ResNet-152 architecture failed to yield any improvement with increased depth, and the error values re-increased.

In addition to pure fundus images, an analysis was also performed to evaluate the effect of the Graham filter. The tests carried out in this context revealed that the contribution of the Graham filter to the performance varies depending on the architectural depth. It was observed that ResNet architectures with sufficient depth achieved more successful results with retinal image inputs filtered by the Graham filter.

In conclusion, this study contributes to the literature by demonstrating the impact of ResNet architecture variants with different depths on age prediction performance in a systematic manner. The best age prediction was achieved with the ResNet-101 architecture trained on images processed with the Graham filter, with a MAE value of 5.02. In age categorization, the most successful results were observed in the ResNet-34 architecture trained on images processed with the Graham filter, with an F-measure of 0.7165. Future studies are planned to focus on analyzing different models and utilizing different demographic information.

REFERENCES

[1] M. Fogel-Levin, S. R. Sadda, P. J. Rosenfeld, N. Waheed, G. Querques, B. K. Freund, and D. Sarraf, "Advanced retinal imaging and applications for clinical practice: A consensus review," *Survey of ophthalmology*, vol. 67, no. 5, pp. 1373–1390, 2022.

[2] N. U. Häner, C. Dysli, and M. R. Munk, "Imaging in retinal vascular disease: A review," *Clinical & Experimental Ophthalmology*, vol. 51, no. 3, pp. 217–228, 2023.

[3] A. C. Oganov, I. Seddon, S. Jabbehdari, O. E. Uner, H. Fonoudi, G. Yazdanpanah, O. Outani, and J. F. Arevalo, "Artificial intelligence in retinal image analysis: Development, advances, and challenges," *Survey of ophthalmology*, vol. 68, no. 5, pp. 905–919, 2023.

[4] R. M. Devi, P. Keerthika, P. Suresh, P. P. Sarangi, M. Sangeetha, C. Sagana, and K. Devendran, "Retina biometrics for personal authentication," in *Machine Learning for Biometrics*. Elsevier, 2022, pp. 87–104.

[5] R. Poplin, A. V. Varadarajan, K. Blumer, Y. Liu, M. V. McConnell, G. S. Corrado, L. Peng, and D. R. Webster, "Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning," *Nature Biomedical Engineering*, Feb 2018.

[6] C. Liu, W. Wang, Z. Li, Y. Jiang, X. Han, J. Ha, W. Meng, and M. He, "Biological age estimated from retinal imaging: a novel biomarker of aging," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2019, pp. 138–146.

[7] Y. D. Kim, K. J. Noh, S. J. Byun, S. Lee, T. Kim, L. Sunwoo, K. J. Lee, S.-H. Kang, K. H. Park, and S. J. Park, "Effects of hypertension, diabetes, and smoking on age and sex prediction from retinal fundus images," *Scientific reports*, vol. 10, no. 1, p. 4623, 2020.

[8] Y. Wen, L. Chen, L. Qiao, Y. Deng, and C. Zhou, "On the deep learning-based age prediction of color fundus images and correlation with ophthalmic diseases," in *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2020, pp. 1171–1175.

[9] H. Kang, M. Jang, and Y. J. Kim, "Development of age prediction model using retinal fundus image and analysis of the results in patients with chronic disease," *Investigative Ophthalmology & Visual Science*, vol. 64, no. 8, pp. 2396–2396, 2023.

[10] J. Wang, Y. Gao, F. Wang, S. Zeng, J. Li, H. Miao, T. Wang, J. Zeng, D. Baptista-Hon, O. Monteiro *et al.*, "Accurate estimation of biological age and its application in disease prediction using a multimodal image transformer system," *Proceedings of the National Academy of Sciences*, vol. 121, no. 3, p. e2308812120, 2024.

[11] Y. Wang, T. Liu, D. Williamson, R. Struyven, Y. Zhou, D. Romero-Bascones, M. G. Lozano, K. Balaskas, M. C. Borja, J. Rahi *et al.*, "Age prediction from retinal fundus images and segmented vessel images using deep learning," *Investigative Ophthalmology & Visual Science*, vol. 64, no. 8, pp. 1105–1105, 2023.

[12] T. Yamashita, H. Terasaki, R. Asaoka, A. Iwase, H. Sakai, T. Sakamoto, and M. Araie, "Age prediction using fundus parameters of normal eyes from the kumejima population study," *Graefe's Archive for Clinical and Experimental Ophthalmology*, vol. 262, no. 10, pp. 3393–3401, 2024.

[13] X. Xue, W. Du, Y. Fujiwara, M. Miyake, K. Sado, and J. Wang, "Age prediction from fundus images by a multi-branch swin-transformer model," in *2025 11th International Conference on Computing and Artificial Intelligence (ICCAI)*. IEEE, 2025, pp. 122–127.

[14] M. Mani, Y.-T. Chen, and C.-L. Lin, "Age prediction from retinal fundus using deep learning," *Available at SSRN 5312819*, 2025.

[15] O. N. Hassan, M. J. Menten, H. Bogunovic, U. Schmidt-Erfurth, A. Lotery, and D. Rueckert, "Deep learning prediction of age and sex from optical coherence tomography," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2021, pp. 238–242.

[16] R. Chen, S. Zhang, G. Peng, W. Meng, G. Borchert, W. Wang, Z. Yu, H. Liao, Z. Ge, M. He *et al.*, "Deep neural network-estimated age using optical coherence tomography predicts mortality," *Geroscience*, vol. 46, no. 2, pp. 1703–1711, 2024.

[17] K.-M. Chueh, Y.-T. Hsieh, H. H. Chen, I.-H. Ma, and S.-L. Huang, "Identification of sex and age from macular optical coherence tomography and feature analysis using deep learning," *American Journal of Ophthalmology*, vol. 235, pp. 221–228, 2022.

[18] R. Kamran, "Retina age analysis dataset," 2025. [Online]. Available: https://huggingface.co/datasets/ramankamran/retina-age-analysis

[19] B. Graham, "Kaggle diabetic retinopathy detection competition report," University of Warwick, Coventry, UK, Tech. Rep., Aug 2015.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[21] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Miami, FL, USA, 2009, pp. 248–255.

[22] Y. Yang, K. Zha, Y. Chen, H. Wang, and D. Katabi, "Delving into deep imbalanced regression," in *Proceedings of the 38th International Conference on Machine Learning (ICML)*, 2021, pp. 11 842–11 851.

[23] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *Proceedings of the International Conference on Learning Representations (ICLR)*, New Orleans, LA, USA, 2019.

[24] P. J. Huber, "Robust estimation of a location parameter," *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, Mar 1964.