

RESEARCH

Open Access



Artificial intelligence classification model for macular degeneration images: a robust optimization framework for residual neural networks

Wen-Hsien Ho^{1,2}, Tian-Hsiang Huang³, Po-Yuan Yang⁴, Jyh-Horng Chou^{1,5,6}, Hong-Siang Huang⁶, Li-Chung Chi^{7,8}, Fu-I Chou^{6*} and Jinn-Tsong Tsai^{1,9*} 

From International Conference on Biomedical Engineering Innovation 2019 Kaohsiung, Taiwan. 15-19 November 2019

*Correspondence:

alvis.cf@gmail.com;
jtsai@mail.nptu.edu.tw

⁶ Department of Electrical Engineering, National Kaohsiung University of Science and Technology, No. 415, Chien-Kung Road, Kaohsiung 807, Taiwan

⁹ Department of Computer Science, National Pingtung University, No. 4-18, Min-Sheng Road, Pingtung 900, Taiwan
Full list of author information is available at the end of the article

Abstract

Background: The prevalence of chronic disease is growing in aging societies, and artificial-intelligence–assisted interpretation of macular degeneration images is a topic that merits research. This study proposes a residual neural network (ResNet) model constructed using uniform design. The ResNet model is an artificial intelligence model that classifies macular degeneration images and can assist medical professionals in related tests and classification tasks, enhance confidence in making diagnoses, and reassure patients. However, the various hyperparameters in a ResNet lead to the problem of hyperparameter optimization in the model. This study employed uniform design—a systematic, scientific experimental design—to optimize the hyperparameters of the ResNet and establish a ResNet with optimal robustness.

Results: An open dataset of macular degeneration images (<https://data.mendeley.com/datasets/rscbjbr9sj/3>) was divided into training, validation, and test datasets. According to accuracy, false negative rate, and signal-to-noise ratio, this study used uniform design to determine the optimal combination of ResNet hyperparameters. The ResNet model was tested and the results compared with results obtained in a previous study using the same dataset. The ResNet model achieved higher optimal accuracy (0.9907), higher mean accuracy (0.9848), and a lower mean false negative rate (0.015) than did the model previously reported. The optimal ResNet hyperparameter combination identified using the uniform design method exhibited excellent performance.

Conclusion: The high stability of the ResNet model established using uniform design is attributable to the study's strict focus on achieving both high accuracy and low standard deviation. This study optimized the hyperparameters of the ResNet model by using uniform design because the design features uniform distribution of experimental points and facilitates effective determination of the representative parameter



combination, reducing the time required for parameter design and fulfilling the requirements of a systematic parameter design process.

Keywords: Residual Neural Network, Uniform experimental design, Hyperparameter optimization, Macular degeneration classification

Background

The macula is located at the center of the retina and is responsible for vision and color identification. Its full name, macula lutea, originates from its dark-yellow color (visible under an ophthalmoscope), which is attributable to its high lutein content. Macular degeneration comprises two forms, dry and wet. Choroidal neovascularization (CNV) and diabetic macular edema (DME) are wet forms, whereas drusen is the dry form. Only 10% of patients with macular degeneration are diagnosed as having a wet form. Because the deterioration of vision is often mistaken as a sign of presbyopia, the symptoms are easily overlooked, which can lead to serious results such as vision loss [1, 2].

The diagnosis of retinal abnormalities requires an examination of the retina by a trained medical professional and retinal image data obtained through optical coherence tomography (OCT). However, determining the form of macular degeneration through medical imaging is a time- and labor-consuming process. Ophthalmologists receive professional training and engage in repeated inspections and discussions. In remote areas or areas with insufficient medical resources, medical technologists are the only front-line medical professionals, and they may not have the diagnostic capacity or confidence to make such a diagnosis. Accordingly, patients usually wait for weeks for the diagnostic result, which delays treatment and involves an enormous amount of labor and social resources [3, 4]. Artificial intelligence (AI) exhibits potential in quick, automatic classification of medical images, acceleration of diagnoses, and reduction of labor [5]. Appropriate applications of AI can greatly accelerate the examination of macular diseases and reduce the costs involved.

Various applications of AI in the diagnosis of eye diseases have been researched. For example, Sambaturu et al. [6] employed a convolutional neural network (CNN) to automatically delineate exudate and hemorrhage in fundus images in DME. In the following year, Gleryz and Ulusoy [7] used a CNN to segment retinal vessels and extract relevant vessel features (e.g., tortuosity, width, and length) for the diagnosis, treatment, and screening of diabetes and vascular diseases such as hypertension. The present study used a residual neural network (ResNet) for modeling, with the aim of distinguishing between macular diseases of various types.

Despite increasing applications of AI in the medical domain, scholars and researchers have predominantly focused their discussion and analysis on the accuracy of models but failed to consider model stability. Accordingly, with consideration of both model accuracy and stability, the present study aimed to construct an AI model with high accuracy and a stable recognition rate. It employed a systematic uniform design method to identify the optimal hyperparameter combination and thus determine the AI model optimal for relieving the diagnostic burden on medical personnel, reassuring patients, and maximizing patient satisfaction.

The uniform experimental design (UED) method developed by Wang and Fang [8–10] uses space filling designs to construct a set of experimental points uniformly scattered

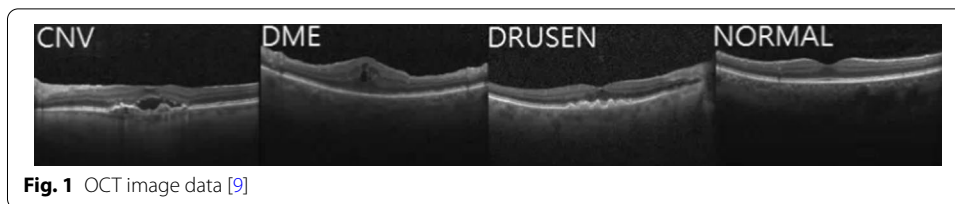


Table 1 Number of images for macular degeneration modeling and testing

Type of macular degeneration	Number of modeling data points	Number of test data points
CNV	37,205	242
DME	11,348	242
Drusen	8616	242
Normal	26,315	242
Total	83,484	968

in a continuous design parameter space. Because it only considers uniform dispersion and not comparable orderliness, UED minimizes the number of experiments needed to acquire all available information. Therefore, the UED is very suitable for solving problems involving multiple factors with multiple levels.

To verify the AI retinal-disease classification model established using a ResNet, the model was compared with CNN, an artificial neural network proposed by Najeeb et al. [11].

Methods

For the diagnosis of macular degeneration, a ResNet model was constructed with hyper-parameters optimized using uniform design. With a focus on model accuracy and low standard deviation, this study proposed a method for establishing an accurate, low-standard deviation AI classification model and applied the model in prediction based on actual medical images to test its performance. The following describes the data collection process and design focus of the proposed method, focusing on the experimental design used to optimize the classification model.

The modeling was conducted using an OCT image dataset published by Kermany et al. [12], which comprised four categories of data: CNV, DME, drusen, and normal data. Figure 1 presents data for the categories, and Table 1 illustrates the numbers of data points in each data category. A total of 83,484 images were used for modeling, with 37,205 CNV, 11,348 DME, 8616 drusen, and 26,315 normal images; 968 images were employed for model testing, with 242 images for each of the categories.

ResNet

ResNets—developed by He et al. [13], a Microsoft team—feature a higher number of layers than previous networks. The deepening of neural networks is crucial to improving network performance; however, deepening the number of network layers in the learning stage of deep learning often hinders the learning process and hence undermines performance. By

contrast, the incorporation of an identity mapping structure into a ResNet creates a direct proportion between the number of layers and the level of performance. Table 2 provides the levels of a ResNet’s six hyperparameters: number of kernel (X_1), kernel size (X_2), pooling size (X_3), layer (X_4), activation function (X_5), and optimizer (X_6).

Note that the learning rate (η) and a very small constant (ϵ) of optimizer Adam in Table 2 were set up to 0.001 and e^{-8} respectively. In addition, a constant was equal to 1.050 and a regulated parameter was set up as 1.6732 for activation function SELU. A gradient parameter (λ) was set up as 0.03 for activation function Leaky ReLU.

Uniform design

Scientific research usually involves multiple experiments, which can be costly and time-consuming. Uniform design improves experimental quality and reduces production time and cost. Uniform design is a multifactor and multilevel experimental design method that features uniform scattering of design points, which enhances the representativeness of each experimental point [8–10, 14]. In this design, a suitable uniform layout is determined for experiments, and a regression analysis is used to obtain the optimal parameters and solution.

A uniform layout is denoted using $U_n(n^s)$, where U refers to uniform layout, n is the number of levels (equal to the number of experiments), and s is the number of hyperparameters. A uniform layout has $(n - 1)$ rows if n is a prime number. An even-number uniform layout is constructed by first establishing an odd-number uniform layout and then removing the last row of the layout. Table 3 describes the $U_{12}(12^6)$ uniform layout established in this study.

Signal-to-noise ratio and model assessment

To maximize the AI model’s accuracy and minimize its standard deviation, signal-to-noise ratio (SNR) was adopted as an indicator of measurement quality. An SNR for robust design is obtained only by repeating each hyperparameter combination three or more times, with a high SNR indicating high quality. Equation (1) was used to calculate an SNR with consideration of variance and the difference between predicted and actual values.

$$SNR_i = -10 \log \left[(\bar{t}_i - m)^2 + \sigma_i^2 \right], \tag{1}$$

Table 2 Levels of a ResNet’s hyperparameters

Hyperparameter	Level			
	1	2	3	4
Number of Kernel (X_1)	8	16		
Kernel size (X_2)	3	5		
Pooling size (X_3)	4	8	16	32
Layer (X_4)	14	20		
Activation function (X_5)	Tanh	SELU	Leaky ReLU	ReLU
Optimizer (X_6)	SGD	Adam		

Table 3 $U_{12}(12^6)$ uniform experiment layout

Number of experiences	Hyperparameter					
	X_1	X_2	X_3	X_4	X_5	X_6
1	8	5	8	20	Tanh	Adam
2	16	5	32	14	Tanh	SGD
3	8	5	4	14	Tanh	Adam
4	16	5	16	20	SELU	SGD
5	8	5	32	14	SELU	SGD
6	16	5	8	14	SELU	Adam
7	8	3	16	20	Leaky ReLU	SGD
8	16	3	4	20	Leaky ReLU	Adam
9	8	3	8	14	Leaky ReLU	Adam
10	16	3	32	20	ReLU	SGD
11	8	3	4	20	ReLU	Adam
12	16	3	16	14	ReLU	SGD

Table 4 Confusion matrix

Actual values	Predicted values			
	Category 1	Category 2	Category 3	Category 4
Category 1	T_{11}	T_{12}	T_{13}	T_{14}
Category 2	T_{21}	T_{22}	T_{23}	T_{24}
Category 3	T_{31}	T_{32}	T_{33}	T_{34}
Category 4	T_{41}	T_{42}	T_{43}	T_{44}

where SNR_i is the SNR of i th hyperparameter combination; \bar{t}_i and σ_i are respectively the mean and standard deviation of accuracy for the i th hyperparameter combination; and m is the targeted accuracy ($m = 1$), $i = 1, 2, \dots, 12$.

$$\bar{t}_i = \frac{\sum_{j=1}^p t_{ij}}{p} \tag{2}$$

and

$$\sigma_i^2 = \frac{\sum_{j=1}^p (t_{ij} - \bar{t}_i)^2}{p - 1}, \tag{3}$$

where t_{ij} is the j th accuracy testing of the i th hyperparameter combination, and $p = 1, 2, 3$. Therefore, the optimal objective is to maximum Eq. (1). A confusion matrix can be used to calculate indicators such as accuracy, sensitivity, and false negative rate in assessing model performance [15]. Therefore, this study employed a confusion matrix for model assessment (Table 4). As depicted in (4), ACC is the overall accuracy of the model; high ACC indicates good model performance.

$$ACC = \frac{\sum_{i=1}^q T_{ii}}{\sum_{i=1}^q \sum_{j=1}^q T_{ij}} \quad (q = 1, 2, 3, 4) \tag{4}$$

Table 5 Accuracy of uniform experimental results obtained using training data

Hyperparameter combinations	Results					
	Accuracy in the 1st experiment	Accuracy in the 2nd experiment	Accuracy in the 3rd experiment	Mean	Standard deviation	SNR
1	0.9340	0.9338	0.9378	0.9352	0.002254	23.76
2	0.9586	0.9463	0.9602	0.9550	0.007606	26.81
3	0.9388	0.9365	0.9398	0.9383	0.001692	24.20
4	0.9678	0.9780	0.9805	0.9754	0.006728	31.87
5	0.9567	0.9551	0.9541	0.9553	0.001312	26.99
6	0.9404	0.9398	0.9418	0.9406	0.001026	24.53
7	0.9557	0.9557	0.9558	0.9557	0.000058	27.07
8	0.9487	0.9544	0.9449	0.9493	0.004782	25.86
9	0.9408	0.9406	0.9408	0.9407	0.000116	24.54
10	0.9528	0.9722	0.9710	0.9653	0.010871	28.79
11	0.9536	0.9527	0.9556	0.9539	0.001484	26.73
12	0.9725	0.9761	0.9662	0.9716	0.005011	30.80

Table 6 SNRs of the uniform experiments

Hyperparameter combinations	Results		
	SNRs obtained using training data	SNR obtained using validation data	Summed SNR
1	23.76	18.48	42.24
2	26.81	24.01	50.83
3	24.20	18.35	42.56
4	31.87	24.31	56.19
5	26.99	22.30	49.30
6	24.53	22.39	46.93
7	27.07	23.96	51.04
8	25.86	22.74	48.62
9	24.54	22.06	46.61
10	28.79	23.68	52.48
11	26.73	24.39	51.13
12	30.80	23.42	54.22

Results and discussion

In the $U_{12}(12^6)$ uniform experiment layout (Table 3), each hyperparameter combination was used for three rounds of random data grouping to construct three ResNet models; the mean and standard deviation of accuracy of each were obtained and the SNR calculated using (1). Subsequently, the performance of all hyperparameter combinations listed in Table 3 was assessed to determine the optimal combination. Table 5 presents the uniform experiment result based on the training data. Table 6 provides the SNR result of the uniform experiment based on the training and validation data. To achieve an accurate, stable model, the SNRs of the training and validation data were summed for each hyperparameter combination to identify the combination with the highest accuracy and stability. Combination 4 had the optimal

Table 7 Test results for hyperparameter combination 4 (test data)

Experiment number	Results	
	Accuracy	False negative rate
1	0.9907	0.0092
2	0.9824	0.0175
3	0.9814	0.0185
Mean	0.9848	0.0150

Table 8 The experimental results of tree neural network models (test data)

Model	Result			
	Best accuracy	Mean accuracy	Standard deviation	Mean false negative rate
ResNet	0.9907	0.9848	0.0051	0.0150
AlexNet	0.9876	0.9803	0.0063	0.0188
CNN	0.9566	–	–	0.0430

hyperparameter combination, with an SNR of 31.87 for the training data and 24.31 for the validation data for a total of 56.19, the largest sum among all combinations (Table 6).

Hyperparameter combination 4 was tested thrice using the 968 test data points specified in Table 1. Table 7 lists the test results. The mean accuracy was 0.9848 and the mean false negative rate 0.0150. The results verified the excellent performance of the hyperparameter combination obtained using the proposed method.

This study subsequently compared a ResNet model with the CNN model proposed by Najeeb et al. [11]. The CNN model had 32 kernels, a kernel size of 3, and a stride of 1 in the convolutional layer; a pooling size of 2 in the pooling layer; a 20% dropout; a ReLU activation function; and two fully connected layers, with one comprising 256 neurons and the other comprising 4. Najeeb et al. [11], who employed the same dataset used in the present study to test their CNN model, obtained an accuracy of 0.9566 and a false negative rate of 0.043.

Ultimately, the present study’s ResNet model compared with an AlexNet model [16]. Refer to [16], there are 4096, 4096 and 4 neurons for three fully connected layers respectively, and extra 50% dropout was added for the first two fully connected layers and all activation functions were set up as ReLU, and optimizers were SGD. Three experiments of the AlexNet model with the same test dataset resulted a mean accuracy of 0.9803 and a mean false negative rate of 0.0188.

Table 8 summarizes the evaluation outcomes for the present study’s model, CNN model of Ref. [11] and AlexNet model of Ref. [16]. In comparison with the CNN model and AlexNet model, the present study’s ResNet model had superior performance, with a mean accuracy of 0.9848 and a mean false negative rate of 0.0150.

Conclusions

The ResNet model established using uniform design exhibited high stability, which is attributable to the study's strict focus on achieving both high accuracy and low standard deviation in the model establishment process. Uniform design was used to optimize the ResNet model's hyperparameters because this design method—with its uniform distribution of experimental points—facilitates effective identification of the representative hyperparameters, reduces the time required for hyperparameter design, and fulfils the requirements of a systematic hyperparameter design process.

To verify the performance of the proposed method in finding the optimal hyperparameter combination, an open dataset was divided into training, validation, and test datasets for the testing of hyperparameter combinations. The test results were then compared with those obtained using the CNN model developed by Najeeb et al. [11]. According to the comparison, the present study's ResNet model, with its optimal hyperparameter combination determined using the proposed method, outperformed the CNN model of Najeeb et al. [11] in both accuracy and false negative rate, which verified the practicality of the proposed method.

Abbreviations

CNV: Choroidal neovascularization; DME: Diabetic macular edema; OCT: Optical coherence tomography; AI: Artificial intelligence; CNN: Convolutional neural network; ResNet: Residual neural network; SNR: Signal-to-noise ratio; ACC: Overall accuracy of the model.

Acknowledgements

Not applicable.

About the supplement

This article has been published as part of BMC Bioinformatics Volume 22 Supplement 5 2021: Proceedings of the International Conference on Biomedical Engineering Innovation (ICBEI) 2019-2020. The full contents of the supplement are available at <https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-22-supplement-5>.

Authors' contributions

WHH, THH, PYY and JHC contributed equally to the algorithm design and theoretical analysis. HSH, LCC, FIC and JTT contributed equally to the quality control and document reviewing. All authors read and approved the final manuscript.

Funding

Publication costs are funded by the Ministry of Science and Technology, Taiwan, under grants MOST 109-2221-E-037-005. The design and part writing costs of the study are funded by the KMU-TC109B08 and costs of collection, analysis and interpretation of data and part writing are funded by MOST 109-2221-E-153-005-MY3 and MOST 109-2222-E-035-002, and the "Intelligent Manufacturing Research Center" (IMRC) from the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.

Availability of data and materials

The datasets analysed during the current study are available in the Large Dataset of Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images 2018, <https://data.mendeley.com/datasets/rscbjbr9sj/3>.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ Department of Healthcare Administration and Medical Informatics, Kaohsiung Medical University, No. 100, Shin-Chuan 1st Road, Kaohsiung 807, Taiwan. ² Department of Medical Research, Kaohsiung Medical University Hospital, No. 100, Shin-Chuan 1st Road, Kaohsiung 807, Taiwan. ³ Center for Big Data Research, Kaohsiung Medical University, No. 100, Shin-Chuan 1st Road, Kaohsiung 807, Taiwan. ⁴ Department of Information Engineering and Computer Science, Feng Chia University, No. 100, Wenhwa Road, Taichung 407, Taiwan. ⁵ Department of Mechanical Engineering, National

Chung-Hsing University, No. 145, Xingda Road, Taichung 402, Taiwan. ⁶ Department of Electrical Engineering, National Kaohsiung University of Science and Technology, No. 415, Chien-Kung Road, Kaohsiung 807, Taiwan. ⁷ Department of Ophthalmology, Kaohsiung Medical University Hospital, No. 100, Shin-Chuan 1st Road, Kaohsiung 807, Taiwan. ⁸ Department of Ophthalmology, Kaohsiung Municipal SiaoGang Hospital, No. 482, Shanming Road, Kaohsiung 812, Taiwan. ⁹ Department of Computer Science, National Pingtung University, No. 4-18, Min-Sheng Road, Pingtung 900, Taiwan.

Received: 15 March 2021 Accepted: 16 March 2021

Published online: 08 November 2021

References

- Lang GE. Diabetic macular edema. *Ophthalmologica*. 2012;227:21–9.
- Perdomo O, Ojalora S, Rodríguez F, Arevalo J, González FA. A novel machine learning model based on exudate localization to detect diabetic macular edema. In: 19th International Conference on Medical Image Computing and Computer Assisted Intervention. 2016; 137–144.
- Naz S, Hassan T, Akram MU, Khan SA. A practical approach to OCT based classification of diabetic macular edema. In: 2017 international conference on signals and systems (ICSigSys). 2017; 217–220.
- Perdomo O, Ojalora S, Gpnzalez FA, Meriaudeau F, Muller H. Oct-Net: a convolutional network for automatic classification of normal and diabetic macular edema using Sd-Oct volumes. In: IEEE 15th international symposium on biomedical imaging (ISBI 2018). 2018; 1423–1426.
- Trivizakis E, Manikis GC, Nikiforaki K, Drevelegas K, Constantinides M, Drevelegas A, Marias K. Extending 2-D convolutional neural networks to 3-D for advancing deep learning cancer classification with application to MRI liver tumor differentiation. *IEEE J Biomed Health Inform*. 2019;23:923–30.
- Sambaturu B, Srinivasan B, Prabhu SM, Rajamani KT, Palanisamy T, Haritz G, Singh D. A novel deep learning based method for retinal lesion detection. In: 2017 international conference on advances in computing, communications and informatics (ICACCI) 2017; 33–37.
- Güteryüz MŞ, Ulusoy I. Retinal vessel segmentation using convolutional neural networks. In: 26th signal processing and communications applications conference (SIU). 2018; 1–4.
- Wang Y, Fang KT. A note on uniform distribution and experimental design. *Chin Sci Bull*. 1981;26:485–9.
- Fang KT. Uniform design and uniform layout. Beijing: Science Press; 1994.
- Tsao H, Lee L. Uniform layout implement using Matlab. *Stat Decis*. 2008;6:144–6.
- Najeeb S, Sharmile N, Khan MS, Sahin I. Classification of retinal diseases from OCT scans using convolutional neural networks. In: 10th International Conference on Electrical and Computer Engineering (ICECE). 2018; 465–468.
- Kermany DS, Zhang K, Goldbaum MH. Labeled optical coherence tomography (OCT) and chest X-ray images for classification. Mendeley Data. 2018. <https://doi.org/10.17632/rscbjbr9sj.2>.
- He K, Zhang X, Sun J. Deep residual learning for image recognition. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR). 2016; 770–778.
- Hickernell FJ. A generalized discrepancy and quadrature error bound. *Math Comput*. 1998;67:299–322.
- Visa S, Ramsay B, Ralescu A, van der Knaap E. Confusion matrix-based feature selection. In: Proceedings of the 22nd midwest artificial intelligence and cognitive science conference. 2011; 120–127.
- Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst*. 2012;25(2):1106–14.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

