An Intelligent Diagnostic System for Thyroid-Associated Ophthalmopathy Based on Facial Images

Xiao Huang1*, Lie Ju2,3, Jian Li1, Linfeng He1, Fei Tong1, Siyu Liu1,4, Pan Li1, Yun Zhang1, Xin Wang2, Zhiwen Yang2, Jianhao Xiong2, Lin Wang2, Xin Zhao2, Wanji He2, Yelin Huang2, Zongyuan Ge2,3, Xuan Yao1, Wei Hua Yang5* and Ruili Wei1*

1 Department of Ophthalmology, Changzheng Hospital, Naval Medical University, Shanghai, China, 2 Airdoc LLC, Beijing, China, 3 Faculty of Engineering, Monash University, Clayton, VIC, Australia, 4 Department of Ophthalmology, Naval Medical Center of the People’s Liberation Army, Naval Medical University, Shanghai, China, 5 The Laboratory of Artificial Intelligence and Bigdata in Ophthalmology, The Affiliated Eye Hospital of Nanjing Medical University, Nanjing, China

Background: Thyroid-associated ophthalmopathy (TAO) is one of the most common orbital diseases that seriously threatens visual function and significantly affects patients’ appearances, rendering them unable to work. This study established an intelligent diagnostic system for TAO based on facial images.

Methods: Patient images and data were obtained from medical records of patients with TAO who visited Shanghai Changzheng Hospital from 2013 to 2018. Eyelid retraction, ocular dyskinesia, conjunctival congestion, and other signs were noted on the images. Patients were classified according to the types, stages, and grades of TAO based on the diagnostic criteria. The diagnostic system consisted of multiple task-specific models.

Results: The intelligent diagnostic system accurately diagnosed TAO in three stages. The built-in models pre-processed the facial images and diagnosed multiple TAO signs, with average areas under the receiver operating characteristic curves exceeding 0.85 (F1 score > 0.80).

Conclusion: The intelligent diagnostic system introduced in this study accurately identified several common signs of TAO.

Keywords: artificial intelligence, diagnosis, facial images, machine learning, medical data analysis, thyroid-associated ophthalmopathy

INTRODUCTION

Thyroid-associated ophthalmopathy (TAO) is a common orbital disease (1). Several quality-of-life surveys have shown that the visual function, mental health, and social function of most patients with moderate-to-severe TAO are severely impeded (2–4). Although its clinical manifestations are complex and variable, there are clear diagnostic criteria (5) and management guidelines (6) for this disease. Experienced ophthalmologists are able to rapidly diagnose the type, stage, and grade of TAO and develop a treatment plan to prevent disease progression based on simple interrogation, visual examinations, and basic eye examinations, such as exophthalmometry. However, few ophthalmologists in developing countries have experience diagnosing and treating TAO. In addition, TAO is a complicated disease that can easily be misdiagnosed and mistreated, especially in its early stages.
According to the TAO diagnostic criteria (5), exophthalmos is an important diagnostic sign of TAO. Therefore, the use of facial images can significantly reduce the time required for patient visits and referrals. However, manual diagnosis of TAO via facial images is time-consuming and inaccurate (7–13). Approximately 26% of patients receive a final diagnosis after more than 12 months, and many patients with TAO do not undergo treatment at specialized centers or undergo delayed treatment in the late stages of the disease, resulting in an unfavorable disease course (14).

In recent years, the application of artificial intelligence (AI), including deep learning technology, into the field of medical imaging has significantly improved the diagnostic accuracy and efficiency of several diseases, including eye diseases. However, there are no reports regarding the use of computer-aided AI tools for the diagnosis of TAO based on facial images.

A novel AI-based system that incorporates deep learning and machine learning techniques to detect signs of TAO based on facial images is presented in this study.

MATERIALS AND METHODS

Data Collection

The study was conducted in accordance with the principles of the Declaration of Helsinki and was approved by the Medical Ethics Committee of Naval Medical University. Written informed consent for the publication of this study was obtained from all patients.

The data of consecutive patients with TAO treated at Shanghai Changzheng Hospital from January 2013 to January 2018 were included in this study. All facial images were captured...
using a Sony ILCE-7M2 camera (SONY China Co., Ltd.) [lighting: NG CN-576 (Guangdong Nanguang Film & Television Equipment Co. Ltd.)] during routine examinations. The patient diagnoses were extracted from medical records. No patient had comorbidities that would affect their facial expressions or images. Photographs of the front, right, and left of each patient’s face were obtained (Figure 1). Nine eye positions were photographed from the front: superior right, superior, superior left, right, front, left, inferior right, inferior, and inferior left. Seven common signs of TAO were identified in the photographs: eyelid retraction, eyelid congestion, eyelid edema, conjunctival congestion, chemosis, corneal ulcer, and ocular dyskinesia.

The contours of the cornea and sclera were manually drawn on each image using the open-source interactive software tool LabelMe (15). The annotated areas were then mapped using one-hot encoding in the annotation maps (Figure 1).

Only horizontal and vertical flips were used for data augmentation. As the eye position changes with data augmentation, the corresponding eye position label was also changed when the flipped images were used.

The test population comprised 20% of the total patient population in this study. The remaining 80% of the patient population served as the training set, including 10% that was used for internal cross-validation to determine the best diagnostic model.

**Diagnostic Modules**

The diagnostic methods used in this study included modules based on eye location (Module I), ocular dyskinesia (Module II), and other signs (Module III), as shown in Figure 2.

In Module I, an image of the entire face of the patient was input into a trained detection network that analyzed the location of the eyes. The image was cropped, retaining only the eye area.

Module II aimed to diagnose ocular dyskinesia based on segmentation of the cornea and palpebral fissures in the eye area images obtained in Module I and specific calculation rules (16). Impaired eye movement was based on left/right eye rotation when the cornea deviated vertically from the vertex of the canthus in the frontal photographs (Figure 3A). When the cornea was tangent to or intersected the line at the vertex of the canthus,
Figure 3 | Diagnostic rules of eye movement disorders (A). Diagnostic rules of eye movement disorders. Frontal images of left/right eye rotation show the cornea (green circle in A) deviating from the vertex of the canthus (yellow line in A), indicating impaired eye movement (left eye in A). When the cornea is tangent to or intersects the vertex of the canthus, no eye movement disorders are present (right eye in A). Frontal images of superior/inferior eye movement show the cornea (green circle in B,C) intersecting the line between the inner and outer canthus (white line in B,C), representing impaired eye movement (right eye in B). When the cornea is tangent to or separated from the line between the inner and outer canthus, no eye movement disorders are present (left eye in B and both eyes in C).

Deep Learning Networks

Preprocessing is required to remove irrelevant information from facial images and retain only the eye area. A single-shot multibox detector (SSD) (17) is a simple, effective framework for object detection that is easy to train and straightforward to integrate into systems that require a detection component. The ResNet-50 (18) detection component was trained to detect multiple signs of TAO. One network was used to detect each sign. A semantic segmentation network (U-Net) (19) was trained to identify eye movement disorders in the areas of the cornea and sclera to improve diagnostic accuracy. The experimental environment was built using Ubuntu version 18.04.4 LTS 64-bit with GPU 1080Ti and 11 GB memory. The implementation of deep neural networks was based on PyTorch version 1.6.0 (20).

Implementation

The captured images were resized to 512 × 512 pixels prior to their use in the SSD (Module I). The cropped eye area images used in U-Net and ResNet were 224 × 224 pixels. The pixel values of each image were normalized from (0, 255) to (0, 1) prior to the training. To expand the available training data and prevent the model from overfitting, the images were augmented using horizontal and vertical flips, translation (−30, 30), and scaling (0.9, 1.1). The stochastic gradient descent optimizer (21) was used for backpropagation to minimize the objective function (cross-entropy loss) in Module I. The learning rate ranged from $1 \times 10^{-3}$ to $1 \times 10^{-5}$ with the division by 10 on epoch 10.
and epoch 20. Thirty epochs were trained. The Adam optimizer (22) and a cosine-shaped learning rate ranging from $1 \times 10^{-3}$ to $1 \times 10^{-6}$ were used in Modules II and III. Fifty epochs were trained for Module II and 100 epochs were trained for Module III.

**Evaluation Metrics**

Individual models were trained to identify specific signs, and the outcome of the model was based on a yes/no binary classification task. The area under the receiver operating curve (AUROC) quantified the capability of each model to conduct the binary classification, with 0.5 indicating a random chance and 1.0 indicating a perfect model of the validated data (23). The sensitivity and specificity were calculated to evaluate the performance of the diagnostic model.

The accuracy of the eye locations and corneal and scleral segmentation were important for Modules I and II. Accurate eye locations help filter irrelevant information, such as the patient’s face and background. Intersection-over-union (IoU; the Jaccard index) was used to evaluate the detection and segmentation models. IoU, one of the most commonly used metrics in object detection and semantic segmentation tasks, was calculated as:

$$\text{IoU} = \frac{\text{Area of Overlap}(A_1, A_2)}{\text{Area of Union}(A_1, A_2)}$$

where $A_1$ is the ground-truth area, and $A_2$ is the prediction of the model.

**RESULTS**

**Patient Characteristics**

A total of 21,840 images from 1,560 patients (3,120 eyes) were used in this study (Table 1). Eye movement disorders were identified in 77.50% of patients, conjunctival congestion...
in 62.63%, chemosis in 69.49%, and corneal ulcers in 7.44% (Table 2).

### Modules I and II
Module I had an accuracy of 0.98 using an IoU threshold of >0.5 (Table 3). Module II had accuracies of 0.93 for corneal segmentation and 0.87 for scleral segmentation.

### Module III
The AUROC for the detection of eyelid edema was 0.90 (95% confidence interval (CI): 0.88–0.93), while that for the detection of chemosis was 0.60 (95% CI: 0.53–0.68) (Table 4 and Figure 4). The AUROC for conjunctival congestion was 0.91 (95% CI: 0.86–0.96) and for eye movement disorders was 0.93 (95% CI: 0.89–0.96). Of the signs present in less than 30% of patients, eyelid congestion had an AUROC of 0.95 (95% CI: 0.90–0.98), eyelid retraction had an AUROC of 0.93 (95% CI: 0.89–0.96), and corneal ulcer had an AUROC of 0.79 (95% CI: 0.76–0.82). The mean AUROC of the seven signs of TAO was 0.85, with a mean sensitivity of 0.80, mean specificity of 0.79, and mean F1 score of 0.80.

The use of the ResNet-101 network backbone achieved an AUROC of 0.91 (95% CI: 0.89–0.93) (Table 5). The use of the InceptionV3 network backbone achieved an AUROC of 0.89 (95% CI: 0.87–0.92).

### DISCUSSION
Thyroid-associated opthalmopathy is typically accompanied by Graves’ disease and is often missed or misdiagnosed, especially in the early stage.

In the automated diagnostic system presented in this study, Module I located the patients’ eyes and filtered irrelevant

---

**Figure 4** | Receiver operating characteristic (ROC) curves for the detection of signs of thyroid-associated ophthalmopathy.

**Table 5** | Performance of different backbones.

| Backbone     | AUROC (95% CI) | Training time | Parameters |
|--------------|----------------|---------------|------------|
| ResNet-50    | 0.91 (0.89–0.93) | ×1.0          | 25.6 M     |
| ResNet-101   | 0.92 (0.90–0.94) | ×1.5          | 44.5 M     |
| InceptionV3  | 0.89 (0.87–0.92) | ×1.8          | 27.2 M     |

AUROC, area under the receiver operating characteristic curve; CI, confidence interval.
This study has some limitations. First, the system has a low diagnostic accuracy for signs of TAO that require auxiliary modalities to aid evaluation, such as chemosis. Second, although the modules were trained using a large-scale dataset, the system may be affected by the imaging environment, resulting in a decline in diagnostic accuracy. The combination of facial images and other data, such as the patient's chief complaint, should be the focus of future studies.

**CONCLUSION**

The deep learning-based automatic system for the detection of signs of TAO including eyelid retraction, eyelid edema, eyelid congestion, conjunctival congestion, and eye movement disorders based on facial images presented in this study is a cost-effective, accurate method for the auxiliary diagnosis of TAO, especially for ophthalmologists or general practitioners with limited experience in TAO diagnosis.

**DATA AVAILABILITY STATEMENT**

The original contributions presented in this study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

**ETHICS STATEMENT**

The studies involving human participants were reviewed and approved by the Medical Ethics Committee of Naval Medical University. The patients/participants provided their written informed consent to participate in this study.

**AUTHOR CONTRIBUTIONS**

XH designed the research study. LJ performed the research and analyzed the data. XW, ZY, JX, LW, and XZ provided help and advice on methodology. XH, JL, LH, FT, SL, and YZ assisted with validation. XH and LJ wrote the manuscript. All authors contributed to editorial changes and read and approved the final manuscript.

**FUNDING**

This research was funded by a grant from the Young Science Foundation of the National Natural Science Foundation of China (project number: 81800865).

**ACKNOWLEDGMENTS**

We would like to express their gratitude to the patients of the Department of Ophthalmology at Changzheng Hospital for their help with this study. We would like to thank Editage (www.editage.cn) for English language editing.
REFERENCES

1. Bahn RS. Graves’ ophthalmopathy. N Engl J Med. (2010) 362:726–38.
2. Kahaly GJ, Petrak E, Harde J, Pitz S, Egle UT. Psychosocial morbidity of Graves’ orbitopathy. Clin Endocrinol (Oxf). (2005) 63:395–402. doi: 10.1111/j.1365-2265.2005.02352.x
3. Lee H, Roh HS, Yoon JS, Lee SY. Assessment of quality of life and depression in Korean patients with Graves’ ophthalmopathy. Korean J Ophthalmol. (2010) 24:65–72. doi: 10.3341/kjo.2010.24.2.65
4. Bruscolini A, Sacchetti M, La Cava M, Nibbio M, Iannitti A, Quartini A, et al. Quality of life and neuropsychiatric disorders in patients with Graves’ orbitopathy: current concepts. Autoimmun Rev. (2018) 17:639–43. doi: 10.1016/j.autrev.2017.12.012
5. Paunkovic N, Paunkovic J. The diagnostic criteria of Graves’ disease and especially the thyrotropin receptor antibody; our own experience. Hell J Nucl Med. (2007) 10:89–94.
6. Bartalena L, Kahaly GJ, Baldeschi L, Dayan CM, Eckstein A, Marcocci C, et al. The 2021 European Group on Graves’ orbitopathy (EUGOGO) clinical practice guidelines for the medical management of Graves’ orbitopathy. Eur Thyroid J. (2021) 185:643–67. doi: 10.1530/EJE-21-0479
7. Estcourt S, Hickey J, Perros P, Dayan C, Vaidya B. The patient experience of services for thyroid eye disease in the United Kingdom: results of a nationwide survey. Eur J Endocrinol. (2009) 161:483–7. doi: 10.1530/EJE-09-0383
8. Zouvelou V, Potagas C, Karandreas N, Rentzos M, Papadopoulou M, Zis VP, et al. Concurrent presentation of ocular myasthenia and euthyroid Graves orbitopathy: a diagnostic challenge. J Clin Neurosci. (2008) 15:719–20. doi: 10.1016/j.jocn.2007.09.028
9. Jang SY, Lee SY, Lee EJ, Yoon JS. Clinical features of thyroid-associated orbitopathy in clinically euthyroid Korean patients. Eye (Lond). (2012) 26:1263–9. doi: 10.1038/eye.2012.132
10. Kahaly GJ, Grebe SK, Lupo MA, McDonald N, Sipos JA. Graves’ disease: diagnostic and therapeutic challenges (multimedia activity). Am J Med. (2011) 124:S2–3. doi: 10.1016/j.amjmed.2011.03.001
11. Andris C. A simple red eye? Or the thyroid ophthalmopathy pitfall. Rev Med Liege. (2002) 57:334–9.
12. Tanwani LK, Lohano V, Ewart R, Broadstone VL, Makshagundam SP. Myasthenia gravis in conjunction with Graves’ disease: a diagnostic challenge. Endocr Pract. (2001) 7:275–8. doi: 10.4158/EP.7.4.275
13. De Roceed Y, Phillippe E, Twickler TB, Van Gaal L. Misdiagnosis of Graves’ hyperthyroidism due to therapeutic biotin intervention. Acta Clin Belg. (2018) 73:372–6. doi: 10.1080/17843286.2017.1396676
14. European Group of Graves’ Orbitopathy, Perros P, Baldeschi L, Boboridis K, Dickinson AJ, Hullo A, et al. A questionnaire survey on the management of Graves’ orbitopathy in Europe. Eur J Endocrinol. (2006) 155:207–11.
15. Russell BG, Torralba A, Murphy KP, Freeman WT. LabelMe: a database and web-based tool for image annotation. Int J Comput Vis. (2008) 77:157–73.
16. Asbury T, Fredrick DR. Strabismus. 15th ed. In: Vaughan D, Asbury T, Riordan-Eva P editors. General Ophthalmology, New York, NY: McGraw-Hill, Medical Publishing Division (1999). p. 216–8.
17. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, et al. SSD: single shot MultiBox detector. Proceedings of the European Conference on Computer Vision: Lecture Notes in Computer Science. Cham: Springer (2016). p. 21–37.
18. He K, Zhang X, Sun RSJ. Deep residual learning for image recognition. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV: (2016).
19. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention: Lecture Notes in Computer Science. Cham: Springer (2015). p. 234–41.
20. Paszke A, Gross S, Massa F, Lerer A, Chintala S, PyTorch: an imperative style, high-performance deep learning library. arXiv [Preprint]. (2019):
21. Bottou L. Stochastic gradient descent tricks. In: Montavon G, Orr GB, Müller KR editors. Neural Networks: Tricks of the Trade. Berlin: Springer (2012). p. 421–36.
22. Kingma DP, Ba J. Adam: a method for stochastic optimization. arXiv [Preprint]. (2014):
23. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem. (1993) 39:561–77.
24. Ren S, He K, Girshick R, Sun J. Faster r-cnn: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intel. (2017) 39:1137–49. doi: 10.1109/TPAMI.2016.2577031
25. He K, Gkioxari G, Dollar P, Girshick R. Mask r-cnn. IEEE Trans Pattern Anal Mach Intel. (2020) 42:386–97.
26. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV: (2016).

Conflict of Interest: LJJ, XW, ZY, JX, LW, XZ, WH, YH, ZG, and XY were employed by Airdoc LLC.

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

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Huang, Ju, Li, He, Tong, Liu, Li, Zhang, Wang, Yang, Xiong, Wang, Zhao, He, Huang, Ge, Yao, and Wei. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.