Intelligent Medicine 3 (2023) 213–227

Contents lists available at ScienceDirect

Intelligent Medicine

journal homepage: www.elsevier.com/locate/imed

Guideline & Standard

The standardized design and application guidelines: A primary-oriented artificial intelligence screening system of the lesion sign in the macular region based on fundus color photography☆

Ocular Fundus Diseases Group of Chinese Ophthalmological Society; Expert Group for Artificial Intelligence Research, Development, and Application

ARTICLE INFO

Keywords:
Artificial intelligence
Screening of the lesion sign in the macular region
Primary screening
Fundus color photography
Lesion sign analysis

ABSTRACT

With the popularity and development of artificial intelligence (AI), disease screening systems based on AI algorithms are gradually emerging in the medical field. Such systems can be used for primary screening of diseases to relieve the pressure on primary health care. In recent years, AI algorithms have demonstrated good performance in the analysis and identification of lesion signs in the macular region of fundus color photography, and a screening system for fundus lesion signs applicable to primary screening is bound to emerge in the future. Therefore, to standardize the design and clinical application of macular region lesion sign screening systems based on AI algorithms, the Ocular Fundus Diseases Group of Chinese Ophthalmological Society, in collaboration with relevant experts, developed this guideline after investigating issues, discussing production evidence, and holding guideline workshops. It aimed to establish uniform standards for the definition of the macular region and lesion signs, AI adoption scenarios, algorithm model construction, dataset establishment and labeling, architecture and function design, and image data acquisition for the screening system to guide the implementation of the screening work.

1. Objective and significance of the development and application of the artificial intelligence screening system of the lesion sign in the macular region

The macular region is primarily associated with visual functions such as fine vision and color vision, and the presence of lesions in the macular region is one of the major factors in vision loss. Because of the sensitivity of the macula, most damage to this area results in loss of central vision, which has a significant impact on the quality of life [1]. Early eye screening is an active mean of reducing the risk of lesions and avoiding vision loss. However, there are approximately 44,800 ophthalmologists nationwide, which means that, on average, one ophthalmologist provides care to approximately 31,250 Chinese people [2]. It can be seen that the extreme imbalance between supply and demand causes ophthalmologists in China facing tremendous work pressure in screening of ocular diseases, while some patients are unable to receive timely screening and treatment. Therefore, it is an inevitable trend to achieve the screening of large base populations for ocular diseases with the help of artificial intelligence (AI), automation, and information technology [3].

Fundus color photography is a convenient, noncontact ophthalmic examination for visualizing lesions of the optic nerve, retina, choroid, and refractive media of the eye and is currently the most cost-effective imaging modality for screening of fundus disease [4]. A wide variety of diseases can occur within the macular region, such as age-related macular degeneration (AMD), central choroidal retinopathy, central exudative chorioretinopathy, and retinal vein occlusion [5]. Because of the limited lesion information provided by the fundus images, it is difficult to provide accurate diagnosis conclusions based on this single imaging examination. However, in the application scenario of primary-oriented screening, the screening system only needs to provide the conclusion as to whether there are suspicious lesion signs in the macular region. If certain lesion signs are present, the subject is considered to be at a risk of eye disease in the macular region and requires further examination at a hospital. Therefore, this guideline focuses on the design and application

☆ The Chinese version of the manuscript "The standardized design and application guidelines: A primary-oriented artificial intelligence screening system of the lesion sign in the macular region based on fundus color photography" has been published in the *Chinese Journal of Ocular Fundus Diseases* (*2022,38(9):711–728 DOI:10.3760/cma.j.cn511434-20220616-00364*).

Corresponding authors: Xiaoxin Li, Fujian Provincial Key Laboratory of Ocular Surface and Corneal Diseases; Eye Institute and Affiliated Xiamen Eye Center of Xiamen University, School of Medicine, Xiamen University, Xiamen, Fujian 361001, China (Email: dr_lixiaoxin@163.com); Ming Zhang, Department of Ophthalmology, West China Hospital, Sichuan University, Chengdu, Sichuan 610044, China (Email: Zhangmingscu@126.com); Yanwu Xu, Beijing Baidu Netcom Science Technology Co. Ltd. Beijing 100085, China (Email: ywxu@ieee.org).

https://doi.org/10.1016/j.imed.2023.05.001
Received 6 May 2023; Accepted 8 May 2023

of a primary screening system for macular lesion signs. Fundus color photos collected under nonmydriatic conditions can satisfy the presentation of the examination region. Therefore, in general, the fundus color photos discussed in this guideline are collected under nonmydriatic conditions.

With the development and application of big data-based AI technologies in the medical field, many high-quality AI algorithms have been successfully applied to assisted diagnosis, such as an AI algorithm being used to discriminate skin cancer from testing images [6] and to analyze lymph node metastasis of breast cancer [7]. In ophthalmology, Gargeya and Leng [8] trained an AI model using color fundus color photographs of 75,137 patients and achieved 94% sensitivity and 98% specificity in detecting diabetic retinopathy. Meanwhile, an AI screening system based on color fundus images demonstrated good sensitivity and specificity in glaucoma screening [9].

At present, there are no unified standards for AI screening of lesion signs in the macular region in the world, including data sources, model construction, and clinical evaluation, thus limiting its clinical application. This guideline was inspired by the "provide Standardized Design and Application Guide of AI-Assisted Screening System for Glaucoma based on Fundus Photography in China (2020)" [9] and "Application Guide of AI-Assisted Screening System for Diabetic Retinopathy Based Fundus Photography" [10], referring to the artificial intelligence medical device industry standard [11–13] drafted by the National Institute for Food and Drug Control, and referring to the group standard issued by the China Association for quality inspection [14]. The guideline group developed standardized design and application guidelines for AI screening of lesion signs in the macular region for engineers, primary ophthalmologists, and other relevant personnel. The content of this guideline will be timely updated according to the new regulations and standards in the field of medical devices, as well as the new guidelines for diagnosing and treating diseases in the macular area. In particular, this guideline is intended to provide reference opinions for the definition of the lesion signs in the macular area, AI application scenarios, algorithm model building, the establishment of a dataset and labeling, screening system architecture and function design, and image data acquisition. Meanwhile, it can standardize the application of AI systems in the basic screening of lesion signs in the macular region and promote the overall improvement of the early screening level of fundus disease in China.

2. Definition of the macular region and lesion signs

2.1. Macular region

According to the Age-Related Eye Disease Study (AREDS) [15], the macular region considered when discussing age-related macular degeneration (AMD) is a circular area with a radius of 2 times diameter of the optic disk (2 disc diameters (2 DD)) centered on the fovea. Meanwhile, it can be seen from the schematic diagram (Figure 1) that the macular region is almost tangential to the edge of the optic disk. In addition, anatomically, the macular region is defined as the area within 5.5 mm of the superior and inferior temporal vascular arches [16].

In conjunction with the above definitions, this guideline defines the macular region as a circular area centered on the fovea with a radius of the smallest of three values: 2DD, the minimum distance from the fovea to the superior and inferior vascular arches, and the minimum distance from the fovea to the edge of the optic disk. As shown in Figure 2, the radius of the green circle is 2 DD; the radius of the blue circle is the minimum distance from the fovea to the edge of the optic disk, and the radius of the red circle is the minimum distance from the fovea to the superior and inferior vascular arches (main venous vessels). The circular area shown with solid lines in Figure 2(a-c) is the macular region defined in this guideline.

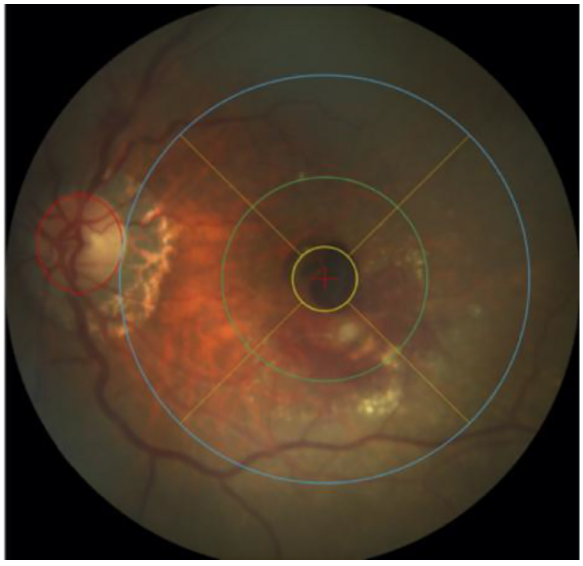

Figure 1. Schematic diagram of the macular region in the Age-related Eye Disease Study [15].

2.2. Definition of lesion signs in the macular region

When screening for lesion signs of macular diseases, not only high-risk disease-related signs but also potential risk signs and disease areas are needed to be detected. The following is the definition of the lesion signs in the macular region in this guideline.

(1) Drusen ≥ 125 μm in diameter [17]: 125 μm is equivalent to the diameter of the vein at the inferior temporal edge of the optic disk. Drusen are gel-like or translucent bodies that result from the deposition of abnormal metabolites of pigment epithelial cells. Drusen can be divided into 4 types [18]: 1) hard drusen refers to yellow dots with well-defined borders; 2) soft drusen are the lesions with less well-defined borders and fusion, accompanied by retinal pigment epithelium (RPE) pigment changes; 3) the mixed type has the manifestations of both hard and soft; 4) degenerative drusen is an irregular map-shaped or halo-shaped atrophy area, which gradually expands and may appear white because of the calcification.

(2) Geographic atrophy (GA) [17]: GA is defined as the loss of RPE, choriocapillaris, and photoreceptor cells. Single or multifocal chorioretinal atrophy areas with clear borders can be seen in the fundus color photography, in which the visibility of the choroid increased, and drusen with various sizes can be observed around it.

(3) New blood vessels with at least one hemorrhage or exudation ≥ 125 μm in diameter [17]: Macular neovascularization (MNV) can be divided into 3 subtypes [19]: 1) Type 1 MNV represents areas of neovascular complexes originating from the choroid and terminating under the RPE; 2) Type 2 MNV: Neovascular originating from the choroid, passing through the RPE and growing under the retinal nerve fiber layer. The neovascular complex is located in the subretinal space, above the level of the RPE. Type 2 MNV can be observed as orange-red lesions in the fundus color photo; 3) Type 3 MNV is the extension of hyperreflectivity from the middle retina toward to the level of the RPE associated with intraretinal edema, hemorrhage, telangiectasis, and retinal detachment. Additionally, the subretinal neovascular membrane is manifested as an irregular circular lesion in the central fovea or a quadrant next to the central fovea, which is gray-white or yellow-white, and hemorrhage or halo can be seen around the lesion. Retinal neovascularization can be observed as small veins originating from

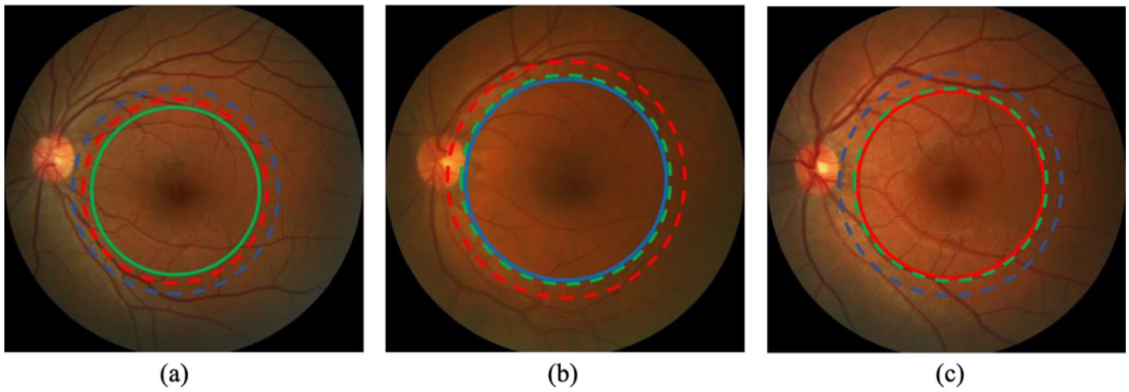

Figure 2. Determination of the macular region. (a): The green circle is the macular region; (b): the blue circle is the macular region; (c): the red circle is the macular region.

the surface of the optic disk and retina, growing along the surface of the retina, and crawling into the vitreous, containing various amounts of fibrous tissue.

(4) Exudation with at least one lesion ≥ 125 μm in diameter [20]. Hard exudation appears as single or multiple yellow-white globular materials in/under the retina. Soft exudation also known as cotton wool spots or leukoplakia appears as grayish-white spots with ill-defined boundaries.

(5) Hemorrhage with at least one lesion ≥ 125 μm in diameter [20]. Hemorrhage can be divided into preretinal hemorrhage, superficial retinal hemorrhage, deep retinal hemorrhage, subretinal pigment epithelium hemorrhage, and simultaneous involvement of multiple layers according to the different retinal layers of the hemorrhage. Preretinal hemorrhage is bright red, boat-shaped, forming a typical liquid surface. Superficial retinal hemorrhages are mostly flames in bright red, and Roth's spots or small white dots in the center of the hemorrhagic lesions can be observed, which are platelet-fibrin thrombi produced when capillaries rupture [21]. Deep retinal hemorrhages are punctate, small, round, dark red patches that spread longitudinally along the course of the nerve. Subretinal hemorrhages are mostly dark red, and subretinal pigment epithelium hemorrhage are mostly brown-red.

(6) Scar [20]: RPE and glial cells proliferate in the retinal space under the stimulation by exudation, resulting in the formation of fibrotic organized scar tissue, which appears as a round, tan-like plaque with a well-defined border [22].

(7) Pigment mottling: Pigment mottling can be divided into hyperpigmentation and depigmentation. Hyperpigmentation [20] appears as black or white spots with protrusions or branches and gradually gathers into black or white spider-like or osteocyte-like spots, which can be irregular strips. Depigmentation manifests as well-defined whitish or yellowish patches.

(8) Macular hole [23] : A round or oval hole appears with a sharp edge in the macular area, occasionally irregular, and the size of the hole varies.

(9) Epiretinal membrane [24]: A layer of translucent, thin or thickened gray avascular proliferative membrane can be seen in the macular area and nearby retinal surface, the retina appears folded, and the nearby retinal small blood vessels are tortuous.

(10) Macular edema [17]: According to the Early Treatment of Diabetic Retinopathy Study (ETDRS) [25], clinically significant macular edema (CSME) is defined as either retinal edema located at or within 500 μm of the center of the macular, hard exudates at or within 500 μm of the foveal center if associated with thickening of the adjacent retina, and extent of retinal thickening large, located in either quadrant of the macula, but with a partial invasion of the central macular region within 1DD. The manifestation of cystoid macular edema (CME) is diffusion or disappearance of foveal light reflection in the center of the macular, accompanied by retinal thickening [26–27]. Edema areas with enhanced or satin-like reflections can also be observed in CME. Advanced macular edema can be observed as a honeycomb or cystic appearance, the retinal thickness of the cyst wall is uneven, and internal separation and vascular shadows can also be observed.

(11) Retinal detachment [28–29]: This includes rhegmatogenous retinal detachment (RRD), exudative retinal detachment, and traction retinal detachment. The retina in RRD with shallow detachment is pale in color and has a clear orange-red boundary. The retina with partial severe detachment appears grayish-white in color and protrudes into the vitreous cavity in a spherical shape. The retina with extensive detachment is undulating and flat along with tortuous and crawling blood vessels. Proliferative membranes can be seen above and below the staleness retinal detachment, which are irregular strips, and full-thickness folds are formed above the retinal surface. Exudative retinal detachment has a smooth surface without traction folds. Traction retinal detachment shows proliferative membranes or adherent organized tissue on the surface.

(12) Intraocular space-occupying lesions have various manifestations, and other auxiliary examinations should be combined in clinical diagnosis to assist the diagnosis.

(13) Others: Abnormalities of the macular region with signs not included in the categories above.

The corresponding relationship between the above lesion signs and common diseases in the macular region is shown in Table 1.

It is worth noting that the guideline provides the definition of the above 13 lesion signs in the macular region. The designers should refer to this guideline and medical expert advice according to the product settings and actual needs when designing the lesion sign screening system.

3. Application of the AI algorithm in the screening system of the lesion sign in the macular region

3.1. Lesion sign classification

The AI algorithm can provide conclusions about the presence or absence of lesion signs (defined in Section 2.2) in the fundus color photographs. Multiple lesion signs can be classified on the same fundus color photograph; therefore, the task is a multilabel classification task, i.e., the labels are not mutually exclusive [30]. As shown in Figure 3, the AI model can discriminate the input images simultaneously for the presence of drusen, geographic atrophy, neovascularization, etc. In clinical applications, the classification results can be evaluated for each lesion sign, and evaluation metrics such as sensitivity, specificity, and Kappa

Table 1 Brief summary of the relationship between the lesion sign in the macular region defined in the guidelines and common retinal diseases

Number	Lesion sign	Retinal diseases
(1)	Drusen	AMD (dry)
(2)	Geographic atrophy	AMD (dry)
(3)	Macular neovascularization	AMD (wet), central serous chorioretinopathy, retinal vessel occlusion, pathological myopia, diabetic retinopathy
(4)	Exudation	AMD (wet), central serous chorioretinopathy, diabetic retinopathy, hypertension-related retinal disease
(5)	Hemorrhage	AMD (wet), central serous chorioretinopathy, diabetic retinopathy, hypertension-related retinal disease, polypoid choroidal vasculopathy, retinal vascular disease
(6)	Scar	AMD (wet), central serous chorioretinopathy
(7)	Pigment mottling	AMD (dry), retinal pigment epithelium tear, central serous chorioretinopathy, retinitis pigmentosa
(8)	Macular hole	Pathological myopia, cystoid macular edema, retinal degeneration, eclipse retinopathy
(9)	Epiretinal membrane	Retinal vascular disease, diabetic retinopathy, retinal vein occlusion, Posterior uveitis, Bechet, Eales, vitreous hemorrhage, Von-Hippel syndrome, intraocular tumor, trauma, intraocular inflammation
(10)	Macular edema	AMD, retinal vein occlusion, retinitis pigmentosa, epiretinal membrane, diabetic retinopathy, central serous chorioretinopathy
(11)	Retinal detachment	Proliferative diabetic retinopathy, periretinal phlebitis, retinal-choroidal tumor, uveitis, choroiditis
(12)	Intraocular space-occupying lesions	Choroidal osteoma, choroidal melanoma, choroidal hemangioma, metastatic cancer, parasites
(13)	Others	Abnormalities of the macular region with signs not listed in the categories above

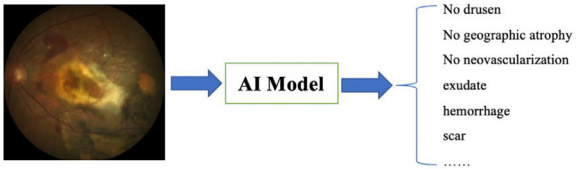

Figure 3. The AI model can classify fundus color photos with multiple labels.

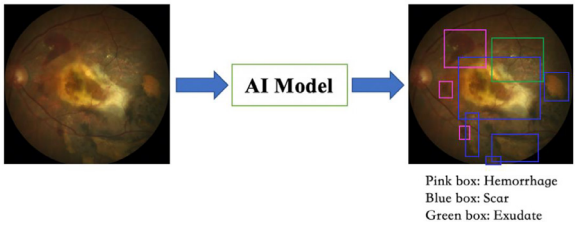

Figure 4. The AI model can detect the lesion area in fundus color photography.

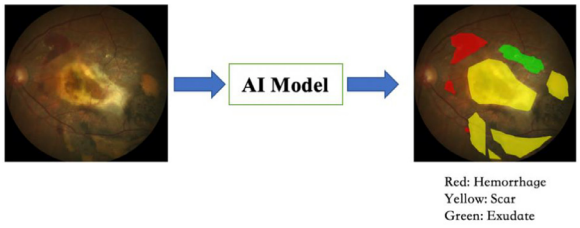

Figure 5. The AI model can segment the lesion contour in fundus images.

can be used. In general, compared with the reference criterion[1] of the presence or absence of the lesion signs in the clinical diagnosis report, the agreement is very good when Kappa is ≥0.80, good when the Kappa is between 0.61 and 0.80, moderate when the Kappa is between 0.41 and 0.60, barely acceptable when the Kappa is between 0.21 and 0.40, and considered almost nonexistent when the Kappa is ≤ 0.20 [31]. The performance goal of the AI algorithm is to achieve a very good agreement between the model prediction results and the professional diagnosis results, i.e., a Kappa of 0.8 or higher.

3.2. Lesion sign detection

The AI algorithm can detect the specified lesion signs in the fundus color images. For example, for the lesion signs defined in Section 2.2, the AI model can locate the specific area from the image (Figure 4). The overlapping boxes in the image indicate the presence of more than one lesion in these boxes.[2] The clinical application of the detection model is usually to visualize target lesions or tissues and assist ophthalmologists in observation. The effect evaluation of such applications is generally consistent with the evaluation indicators of object detection tasks in computer vision [32], which will be described in Section 4.5.2.

3.3. Lesion sign segmentation

The AI algorithm can segment the specified lesion signs in the fundus color images. For example, for the lesion signs defined in Section 2.2, the AI model can segment the specific lesion contour from the image (Figure 5) [33]. The application of the segmentation model in the clinic is usually to visualize the target lesions or tissues and assist ophthalmologists in observation. The effect evaluation of this application is generally consistent with the evaluation index of segmentation tasks in computer vision [34], and the specific index will be introduced in Section 4.5.3.

3.4. Other applications of AI algorithms

AI algorithms can be extended to disease dimension applications in addition to the analysis of lesion signs. This guideline provides two application references for disease classification and grading, which can be used as extended functions of the screening system. The lesion sign screening system does not need to include screening functions for these disease-based functions.

3.4.1. Disease classification

The AI algorithm can discriminate whether the input image has a certain disease in the macular region, and the conclusion of this function is whether the subject has a disease or not, which is a binary classification task. It can be extended to a multiclassification task if it needs to be subdivided into specific disease types. Kappa between the results of the AI algorithm and the reference standard can measure the effectiveness of the AI algorithm in clinical application.

[1] Reference criterion [11]: diagnostic and therapeutic procedures, or benchmarks based on labeling procedures. Reference criterion could contain labels such as disease, physiological state or abnormality, and location as well as degree.

[2] Please note that the target detection boxes are rectangular boxes, the actual lesions are in the rectangular box, but may not completely fill the whole rectangular box.

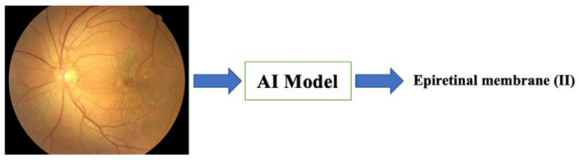

Figure 6. AI model can identify the epiretinal membrane and grade it in fundus images.

3.4.2. Disease grading

AI algorithms can identify and grade diseases based on the input images (Figure 6). Disease grading is a multiclassification task. The evaluation in clinical applications is similar to disease classification, and evaluation measures such as sensitivity, specificity, accuracy, and Kappa can be used to measure the effectiveness of the application.

4. Algorithm construction and accuracy requirements for the AI screening system of the lesion sign in the macular region

4.1. Datasets and quality control

As the basis for AI technology "thinking" and "deciding", data is the cornerstone of AI technology implementation. In the process of construction, verification, and production of the AI screening system, it is inseparable from the support of a large number of datasets. The fundus image collection process requires the shooting ophthalmologist/technician to strictly follow the standard image shooting requirements to avoid uploading unqualified images. This guideline classifies the quality of color fundus images into three categories: qualified, acceptable, and unqualified [35].

(1) Qualified: There is no quality problem in the image, and the shooting position meets the requirements of the color fundus image acquisition standard in this guide (Figure 7);
(2) Acceptable: There are slight exposure problems (light leakage at the edges), small smears that do not affect interpretation, slight out-of-focus or blurred images, etc. (Figure 8);
(3) Unqualified: Severe exposure abnormality, severe refractive interstitial opacity, large-scale contamination, missing information and irrelevant images, etc. (Figure 9)

The guideline requires at least one image of each eye that can be analyzed by the AI system, and read by the doctor, and saved. The image must be in the correct position, with clear focus, suitable exposure, and no interstitial opacity. It should clearly display the retina, macular area, and optic disk structure. The quality requirements are as follows [36]:

(1) Except for drusen, abnormal pigment or new blood vessels and other signs related to fundus macular disease, 90% of the blood vessels in the picture can be identified.
(2) The main fundus structure is in the correct position. When using single-field image screening, the image field of view is no less than 45° horizontally and vertically, the macular fovea and the optic disk are more than 2 optic disk diameters (DD) away from the image edge. When using double visual field images, the horizontal and vertical directions of each visual field are no less than 45°; the macular area image requires that the fovea is less than 1.5 DD away from the image center. The optic disk area image requires that the optic disk center is less than 1.5 DD away from the image center. The angle between the optic disk and the macular center line and horizontal line is not greater than 24°.
(3) There are no dark and/or bright reflective areas that affect interpretation within the imaging.
(4) Moderate exposure, no overexposure or underexposure.
(5) No lens stains, eyelids and/or eyelashes and other occlusion shadows, no motion artifacts.
(6) There are no other image errors such as the diseased eye is not displayed in the image or the color photo of the fundus is taken outside the fundus range.

When the image does not meet the above quality requirements, the following adjustments are required.

(1) The position of the main fundus structure is incorrect: Adjust the sitting posture of the subject, adjust the fixation point, confirm whether the patient has strabismus or other ocular abnormalities, and retake the image;
(2) Overexposure, underexposure, and wrong focus: Adjust the exposure and focus settings of the fundus camera, and retake the image. If the image is too dark, the patient's pupil size should be confirmed, and the exposure time of the affected eye to bright light should be shortened accordingly, to reduce the inspection room brightness;
(3) The pupil is too small to obtain satisfactory fundus images: Retake the fundus images after excluding the contraindications of mydriasis;
(4) Obstruction of eyelids and eyelashes: Prompt the patient to open the eyelids during the photo-taking process, assist the patient to raise the eyelids if necessary, and retake the image;
(5) Reflection of the iris: Remind the patient to stare at the fixed point of view and to look away, and retake the image;
(6) Lens stains: Check and clean the lens;
(7) Missing information: Confirm that the image information is stored completely, re-shoot and store the image;
(8) External eye and nonfundus images: Prevent uploading of irrelevant images such as images with the lens cover not opened, environmental images shot by mistake, and external eye inspection images.

This guideline recommends that the collection of images in the dataset should consider the following inclusion and exclusion criteria:

Inclusion criteria:

(1) Color photos of fundus taken with a single/double field of view.
(2) The image quality is qualified or acceptable.

Exclusion criteria:

(1) The image quality is unqualified.
(2) Treatment traces can be observed in the image.

4.2. Dataset establishment

This guideline classifies the datasets involved in the design and application of screening systems into three major categories (as shown in Figure 10): datasets for the model establishment, validation, and for the clinical trials. In the model construction dataset, the tuning set can be called the validation set, the test set can be called the internal test set, and the datasets for the model validation, and the clinical trials can be called the external test set. Duplicate samples or duplicate examinees should not appear in the different subdatasets. In addition, the inclusion of medical data requires ethical approval and desensitization in advance. Guide users can selectively refer to the relevant contents of this guide according to the actual application scenarios and conditions.

4.2.1. Source of the dataset for model establishment

The purpose of the model establishment dataset is to provide labeled data for model training, tuning, and self-testing so that the model can learn the optimal parameters and achieve good screening results. Because of the differences in shooting angles, imaging colors, and resolution of different cameras, it is necessary to collect data involving multiple camera equipment and under multiple shooting conditions to

improve the generalization ability[3] of the AI model. With reference to the guidelines for the application of AI systems for diabetic retinopathy [10] and glaucoma [9], this guideline recommends collecting data from no less than three medical institutions (to satisfy the diversity of acquisition environments and personnel operations) and no less than three devices (to satisfy the diversity of acquisition devices). Since this guideline targets the Chinese population eye screening application scenario, the data source for model establishment should belong to the Chinese population. Meanwhile, the samples in the dataset should cover multiple ethnic groups, regions, and ages. In the process of AI model establishment, the source of the dataset should include various public datasets (such as hospitals at all levels, ophthalmology centers, epidemiological survey data, public data, etc.), but the data should comply with the labeling method of Section 4.3, and must be relabeled if needed. All data acquisition must be approved by the ethical review committee of the relevant units.

Note that when collecting data in the different medical institutions, samples from the previous cohort should be randomly selected to construct the dataset, and high-quality images should not be selected by manual intervention.

4.2.2. Partition for the model establishment dataset

This guideline suggests dividing the model establishment dataset using the leave-one-out method [37], which directly divided the dataset into three mutually exclusive sets, including training set, tuning (validation) set, and test set. Referring to the data division ratio used in the paper on diabetic retinopathy identification published by Google in the journal of *JAMA* [38]: the total ratio of training and tuning sets was 91.56% and the ratio of test sets was 8.35%. This guideline recommends that the proportion of training and tuning sets should be around 90% and the proportion of test sets be at least 10% in the model establishment dataset. The division of training and tuning sets can be divided into an 8:1 or 7:2 ratio according to the commonly used in AI technology applications. It is worth noting that the test set here is only used for self-testing in the model establishment session. In addition to the division of the training, tuning, and test sets, the screening system can divide the training (including tuning) and test sets based on the model training characteristics.

It should be noted that the dataset partitioning should be consistent with data distribution among the training, validation, and test sets [39], i.e., the proportion of samples in each category should be consistent, and the categories here should be considered as lesion signs, subject age and gender category, image acquisition institution, and camera type. Meanwhile, the data distribution of the test set should be in line with the expected application scenarios, and multiple fundus images of the same subject should be assigned to one of the training set, validation set, or test set, i.e., multiple fundus images of the same subject should not be divided into different subdatasets.

4.2.3. Quantity of the samples in the model establishment dataset

From the division of the model establishment datasets in the previous section, it is clear that this guide recommends that the test set should account for 10% of the total number of datasets. In the following, this guideline will derive the number of images needed for each lesion sign category in the test set based on the expected effect of the AI model and thus provide suggestions for the total number of model establishment datasets.

The calculation of the required sample size for each category is based on the confidence interval method [40] of the assessment index:

$$N = \frac{\left[Z_{1-a}\right]^2 P(1-P)}{\Delta^2} \tag{1}$$

where Z is the Z statistic at the confidence level, Δ is the allowable error, P is the expected assessment index (sensitivity or specificity), and N is the required sample size. Setting the significance level $\alpha = 0.05$ (two-sided), $Z_{1-\alpha} = 1.96$, and the allowable error Δ for the expected assessment indicator is set to 5%.

(1) When the expected sensitivity of the AI model is 90% and the specificity is 90%, the minimum positive or negative sample size corresponding to the samples belonging to each lesion sign is

$$N = \frac{1.96^2 \times 0.9 \times (1-0.9)}{0.05^2} \approx 138 \tag{2}$$

that is, in the test set, at least 138 samples are required for each of the lesion sign and no-target sign categories defined in Section 2.2.

(2) Similarly, when the sensitivity and specificity of the AI model is expected to be 85%, a minimum of 196 samples are required for a single category in the test set.

Referring to the above calculations, the guideline recommends that the number of samples in the test set of the model establishment dataset involving each type of lesion sign defined in Section 2.2 be at least 200. Referring to the guidelines for the standardized design and application of glaucoma-assisted screening systems [9] (screening positive samples: normal fundus samples: other eye disease samples = 2:7:1), it is recommended[4] that the test set should contain at least 700 clinically normal fundus images and at least 100 fundus images with other eye diseases.[5] On the basis of the proportion of the test set in the total dataset, it is deduced that the samples of the total dataset of the model establishment that meet the various types of lesion signs defined in Section 2.2 are at least 2,000. It is recommended that the total dataset contains at least 7,000 clinically normal fundus images and at least 1,000 fundus images with other fundus diseases. Please note that multiple lesion signs appearing on the same image are acceptable.

In addition, this guideline recommends that the number of samples in each type of subdataset be evenly distributed according to the collection institution and machine model. It is recommended to refer to the incidence rates of macular regional diseases in different sexes and different age groups when considering the gender and age distribution of each type of sample in the dataset.

4.2.4. Source and quantity of the samples in the model validation dataset

To verify the effect of AI models, except for using test sets for validation during model establishment, collecting additional datasets is recommended to test the effect after the model establishment. The data collected in this session are required to be consistent with the model establishment dataset and meet the requirements of the diversity of collection environments, collector operations, the diversity of collection devices as well as ethnicity, region, age, etc. This dataset can be acquired from the public dataset but should be confirmed using the labeling method of Section 4.3, and needs to be relabeled if needed. All data acquisition is subject to the approval of the ethical review committee of the relevant unit. To meet the validation of the

[3] Generalization ability: It refers to the ability of artificial intelligence algorithms to adapt to fresh samples. The purpose of machine learning in artificial intelligence techniques is to learn the features implied behind the data, and the ability of trained models to give suitable outputs for datasets other than those with the learned features is called generalization ability.

[4] On the basis of ensuring at least 200 positive samples for each category, it is very difficult to collect samples from normal fundus and other eye diseases that meet the epidemiological distribution. Therefore, suggestions are made on the number of samples from normal fundus and other eye diseases in this guide. On this premise, the guidelines also suggest that the sample size of each type in the collected dataset should meet the epidemiological distribution as much as possible.

[5] If the designed screening system can screen multiple lesion signs, samples with different signs can be "samples with other fundus diseases".

generalization ability of the screening system, the acquisition environment, acquisition personnel, and acquisition equipment involved in this dataset should be as different as possible from the model establishment dataset above. As shown in Figure 10, the model validation datasets can consist of multiple compositions, and these datasets can be used for validation of the screening system by multiple vendors or third-party organizations.

The quantity of the validation datasets should be established by the vendor and third-party organizations based on the testing requirements. The quantity requirement can also be referred in Section 4.2.3. The number of samples in the prepared dataset involving each type of lesion sign as defined in Section 2.2 is at least 200, and the dataset should also contain at least 700 clinically normal fundus images and at least 100 fundus images containing other fundus diseases. To validate the screening effect of images with various qualities, samples with two image qualities of qualified and acceptable are required in the model validation dataset. The image quality is judged by the image quality criteria; please refer to Section 4.1 in this guideline. According to the confidence interval sample size calculation formula (1), there should be at least 200 images of each quality in the validation dataset. The image quality label and the lesion sign label are present simultaneously on each image.

4.2.5. Source and quantity of samples in datasets for clinical trials

Clinical trials are designed to validate the performance of AI models; therefore, this guideline recommends that clinical trials should be designed using a prospective, paired, multicenter, and target value approach. The main bases for the design are the Quality Management Standards for Medical Device Clinical Trials [41], the Guidelines for the Registration Review of Artificial Intelligence Medical Devices [39], and the Approval Points for Deep Learning-Assisted Decision of Medical Device Software [42].

(1) Multicenter: Clinical trials of medical devices should be conducted in 3 or more clinical trial institutions, i.e., multi-center clinical trials. The clinical trial centers are different from the sources of the datasets for model construction, and it is recommended that the sample number of each category in each center be balanced.
(2) Paired: Fundus color photographs of the same subject in the trial should be read using both the AI model and the control method (central reading), i.e., paired design.
(3) Target value method: The amount of the data collected in the clinical trials needs to meet the conditions to verify whether the sensitivity and specificity of AI models applied in clinical scenarios meet the expected targets.
(4) The clinical trial dataset used to test the clinical effect of the AI model should reflect the quality and diversified distribution of the data in the real world and ensure that the sample distribution conforms to the clinical reality and has no intersection with the dataset used in the model establishment.

In clinical trials, the size of samples required should be consistent with the amount of data in the test set calculated in Section 4.2.3, i.e., this guideline recommends that clinical trials provide a dataset involving at least 200 images of each lesion sign as defined in Section 2.2, at least 700 images of clinically normal fundus images, and at least 100 fundus images with other fundus diseases. Similar to the model validation dataset introduced in Section 4.2.4, in order to evaluate the screening effects of images of different qualities, clinical trial datasets should contain qualified images and acceptable images. According to the formula (1) for calculating the sample size of the confidence interval of evaluation indicators, there should be at least 200 images of each quality in clinical trial datasets.

It is worth noting that this guideline recommends that the datasets used for model testing and validation conform to the data distribution of the intended application scenario. According to the epidemiological investigation of macular lesions [43–45], among the lesion signs mentioned in Section 2.2, the lowest prevalence was pigment proliferation (0.10%), and the highest prevalence was drusen and geographic atrophy (7.21%). If the number of images of pigment proliferation signs in the clinical trial dataset is 200, the number of images corresponding to the two signs of drusen and geographic atrophy should reach 14,420, and the number of images without macular lesions should be at least 160,000, which is hard to achieve. Therefore, this guideline suggests that on the premise that "the dataset provided by clinical trials involves at least 200 images of each pathological sign defined in Section 2.2, at least 700 images of clinical normal fundus, and at least 100 fundus images of other fundus diseases", the number of samples corresponding to the categories of lesion signs with high prevalence and cases without macular lesions should be increased as far as possible.

4.3. Annotation for datasets

The annotation of the above datasets and the reading of the clinical trial datasets can be done in the way described in this section, and the annotated results are used as the "reference standard" for each task.

4.3.1. Annotation model

In accordance with the annotation and quality control specifications for fundus color photographs [14], this guideline recommends the use of a labeling model of 3 + 2 + 1 or 3 + 2 + 2, i.e., 3 labeling physicians + 2 senior physicians + 1 or 2 arbitration experts. The labeling physicians are medical professionals who are licensed to practice medicine or have passed the relevant professional training. The senior physicians are professionals who have worked in the field of ocular disease for 5 years or more, and the arbitration experts are higher-level experts with the title of an associate chief physician or above. If two of the three labeling physicians do not have identical labeling results, two senior physicians should be introduced. If the two senior doctors have the same labeling, this labeling result shall be regarded as the final annotation; if disagreement still exists, one or two arbitration experts shall be introduced. The arbitration experts provide the final annotation after the review.

4.3.2. Annotation staff

Annotation staff should receive data labeling training including labeling software operating procedures, labeling protocols, etc. The specific contents of the labeling training and assessment are as follows.

(1) Training content:
   a) To standardize the labeling process, the labeling staff should be trained by the system engineer before using the labeling system.
   b) To standardize the understanding of the rules of fundus image labeling, training in labeling lesion signs should be conducted by an ocular disease specialist (chief physician of a tertiary care hospital for ocular disease).
(2) Assessment mechanism:
   a) The evaluation was carried out by a practical operation method: A certain number of evaluation images will be randomly selected from each lesion sign case, and the experts and the labeling physicians are asked to label the evaluation images with the specified forms (Section 4.3.3 for details), and the consistency rate of the annotations obtained by the labeling physicians and experts will be calculated.
   b) Assessment index: The labeling consistency rate is required to be no less than a certain percentage. According to the ocular specialists, this guideline recommends that the labeling consistency rate between qualified labeling physicians and ocular specialists should not be less than 80%.

Table 2 The corresponding relationship between the functions of the screening system and the forms of the data annotation

Screening system functions	Data annotation forms
Referral advice	Mark one of three labels: No referral, recommended referral, or confirmed referral
Lesion sign	Provide the names of the lesions (0, 1, or more labels)
Lesion sign region	Mark the boxes (0, 1, or more) covering the lesions
Lesion sign profile	Delineate the lesions region (0, 1, or more)
Disease	Provide macular disease name (0, 1, or more)
Disease severity	Mark the severity of the arisen macular disease, e.g., grade/staging labels (0, 1, or more)
Image quality	Refer to Section 4.1 to mark one of the qualified, acceptable, or unqualified labels

4.3.3. Annotation procedure

The labeling object is a fundus color image, and the labeling form should be set according to the function of the screening system, as shown in Table 2. The screening system can contain multiple functions at the same time; therefore, various labeling forms should be viable.

The following three important elements are included in the labeling processing: treatment traces determination, image readability determination, and specific annotation procedure. Before labeling, the annotation system administrator should assign the user's name and password to the staff. Then, the staff can log in to the annotation system and start labeling.

(1) Treatment traces determination: If treatment traces are found, such as laser point, silicone oil filling, gas filling, external pressure/ring ligation, etc., the image does not need to be labeled and directly removed.
(2) Image readability determination: Image shooting quality will affect readability. If more than 1/3 of the macular area is unreadable due to underexposure or overexposure, it is suggested that the image should not be labeled and directly removed.
(3) The complete labeling process is as follows (Figure 11).

4.3.4. Annotation evaluation

During the labeling process, this guide recommends assessing the quality of the annotation. The assessment can be performed from both correctness and repeatability perspectives. The evaluation process is similar to that of the annotation staff assessment, in which a certain number of images are randomly selected from the annotated images for evaluation (it is recommended to contain 200 samples), and the ocular specialist is asked to label the selected images.

Correctness assessment: The annotation results of the selected images are compared between the annotation staff and the ocular specialist. If the consistency rate reaches the preset proportion, the annotation is considered acceptable.

Repeatability assessment: The selected images are repeatedly put into the annotation data pool and annotated by the annotator for a second time. Then, the two labeling results of each selected image are compared and evaluated. If the consistency rate reaches the preset ratio, the labeling repeatability of the annotation is considered to be qualified.

For unqualified images, the existing labels should be erased and put back into the annotation data pool for re-annotation, and the unqualified labeling physicians should be retrained.

4.4. Algorithm model construction

In this guideline, the AI screening system of the lesion sign in the macular region is responsible for providing the conclusion of the presence or absence of lesion signs, as well as the visual presentation for segmentation or localization of specific lesion signs. During the model construction process, the training set images and the corresponding annotations are input into the model to learn the mapping relationship between the image features and the target task and update the model parameters [37]. Before the training process, a designed algorithm is required to perform preliminary image preprocessing, such as image foreground extraction, image scaling, and image tone enhancement, to reduce the sample differences between models by normalizing the size and tone of the input images. At the same time, the algorithm can flip, rotate, or crop the original image randomly, as well as adjust the image color and shade level to enhance the diversity of training samples. In addition, the algorithm can also introduce random noise to the images during the training process to enhance the robustness of the model against noise and sample attacks [46]. After the training, the model is optimized on the tuning set. After tuning, the model can be tested and the performance can be evaluated on standard test sets.

4.5. Evaluation metric for AI models

Section 3 introduces the application scenarios of AI algorithms in the screening of lesion signs in the macular region, including classification of lesion signs, detection of lesion signs, segmentation of lesion signs, and extended disease grading and diagnosis. This section describes the evaluation metrics for the models according to different application scenarios.

4.5.1. Lesion sign classification

Lesion sign classification belongs to the multilabel classification task, and the mainly used model evaluation metrics including sensitivity, specificity, miss rate, accuracy, F_1 score, area under the curve (AUC), Kappa, etc., and the specific evaluation metrics are defined and calculated as shown below [11-12].

Sensitivity (Sen), also known as recall (R), is the proportion of true positive samples to all positive samples.

$$Sen = \frac{TP}{TP + FN}$$

Specificity (Spe) is the proportion of true negative cases to all negative cases.

$$Spe = \frac{TN}{TN + FP}$$

Miss rate (MR) is the proportion of positive samples not found in the test to all positive samples.

$$MR = 1 - \frac{TP}{TP + FN}$$

Precision (Pre), also known as a positive predictive value, is the proportion of true-positive samples to those predicted positively by the model.

$$Pre = \frac{TP}{TP + FP}$$

Negative prediction value (NPV), is the proportion of true negative samples to those predicted as negative by the model.

$$NPV = \frac{TN}{TN + FN}$$

Accuracy (ACC), is the proportion of samples correctly predicted by the model to the total sample.

$$ACC = \frac{TP + TN}{N}$$

F_1 score is a summed average of recall and precision.

$$F_1 = \frac{2 \times P \times R}{P + R}$$

The Jorden index, also known as the correct index, assumes that false negatives and false positives are equally harmful. The Jorden index is the sum of sensitivity and specificity minus one, and a larger index indicates better screening.

$$Jorden\ index = Sen + Spe - 1$$

AUC: The area under the receiver operating characteristic (ROC) curve. Where ROC is the curve formed by calculating the sensitivity and specificity of the screening system on the test set at a set of predefined thresholds, resulting in a set of (1-specificity, sensitivity) operating points that are connected in sequence.

The Kappa coefficient is an indicator used to evaluate the agreement of the screening system with the reference criterion.

$$Kappa = \frac{N(TP + TN) - \left(R_1C_1 + R_2C_2\right)}{N^2 - \left(R_1C_1 + R_2C_2\right)}$$

where R_1 is the sum of true positive and false positive cases, R_2 is the sum of false negative and true negative cases, C_1 is the sum of true positive and false negative cases, C_2 is the sum of false positive and true negative cases, and N is the sum of sample cases.

4.5.2. Lesion sign detection

The results of the localization task can be presented in two ways: one is to output the coordinate information of the target centroid, and the other is to output the information of the box of the region where the target is located. For the model that outputs only a one-point coordinate, the evaluation metric is usually chosen as the average Euclidean distance; for the model that outputs a target box, the evaluation metric is usually the intersection over union and the mean average precision.

Mean Euclidean distance (MED):

$$MED = \frac{1}{N}\sum_{i=1}^{N}\left(\sqrt{\left(x_i - x_i^0\right)^2 + \left(y_i - y_i^0\right)^2}\right)$$

where $(x_i,\ y_i)$ is the model prediction coordinate, and $(x_i^0,\ y_i^0)$ is the reference criterion.

Intersection over union (IoU) measures the extent to which the prediction box overlaps with the reference criterion.

$$IoU(X,Y) = \frac{|X \cap Y|}{|XY|} = \frac{TP}{TP + FN + FP}$$

Mean average precision (mAP) is used to calculate the average accuracy of all kinds of object detection in multiple object detection tasks [47]. The calculation process is as follows.

In the object detection task, the model will provide the probability value corresponding to a certain object class in the target box, which is called confidence. When the confidence is higher than threshold $th_{confidence}$, the current prediction results are retained. Then, the IoU of the reference criterion of this target box and the predicted box will be calculated. When the IoU is greater than threshold th_{IoU} and the predicted category in the predicted box is consistent with that in the reference criterion, the prediction result is considered a true positive. The rest are considered false positives. By changing the confidence threshold $th_{confidence}$ from large to small, the corresponding precisionp and recallr of the predicted results can be calculated, and the precision-recall curve can be drawn. The area under the curve is denoted as average precision (AP) [48]. After obtaining the AP of each category, the mAP of the multiple object detection task can be obtained:

$$mAP = \frac{\sum_{i=1}^{K} AP_i}{K}$$

where K is the number of the object categories contained in the detection task and AP_i is the AP value of the prediction result of the *ith* category object.

4.5.3. Lesion sign segmentation

The evaluation indices of the segmentation model mainly include the DICE coefficient, Jaccard coefficient, IoU, sensitivity, specificity, and so on. Since the segmentation task can be regarded as a binary classification task that divides image pixels into foreground and background, sensitivity and specificity can be used to evaluate the discrimination effect of the segmentation model for each pixel.

The DICE coefficient refers to the proportion of the intersection of the segmented contour and the reference criterion contour to the average value of the segmented contour and the reference standard contour:

$$DICE(X,Y) = \frac{|X \cap Y|}{\frac{1}{2}(|X| + |Y|)} = \frac{2 \times TP}{(TP + FN) + (TP + FP)}$$

where $|X \cap Y|$ represents the intersection between X and Y, and $|X|$ and $|Y|$ represent the number of pixels of region X and region Y.

Jaccard coefficient: the proportion of the intersection of the segmented contour and the reference criterion to the union of the segmented contour and the reference criterion, also known as the intersection ratio (IoU):

$$Jaccard(X,Y) = IoU(X,Y) = \frac{|X \cap Y|}{|XY|} = \frac{TP}{TP + FN + FP}$$

4.5.4. Extended function - disease grading and diagnosis

For AI models, both disease grading and disease diagnosis are classification tasks, so the evaluation metrics can refer to those listed in Section 4.5.1.

4.5.5. Comprehensive evaluation metric

Screening compliance rate: The ratio of positive cases predicted by the model to those determined by the ophthalmologist reading.

Repeatability requirements: The same ophthalmologist/technician deploying and running the AI-assisted screening system on different servers that meet the requirements of the configuration environment should be able to reproduce the predicted results for the same images.

Reproducibility requirements: Different ophthalmologists/technicians deploy and run AI-assisted screening systems in the same configuration environment at different time periods. For the same image, it should be able to reproduce its prediction results.

Robustness requirements: The AI-assisted screening system should have stable performance for multi-center, multimodel, and multiparameter clinical data, and the statistical performance changes of the main metrics are recommended to be controlled within a certain range. The specific range value depends on the function of the screening system and the corresponding medical device standards and regulations.

5. Application standard of the AI screening system of the lesion sign in the macular region

5.1. Type of the AI screening system

The offline version of the AI system can be installed on a computer or mobile electronic device and can perform intelligent screening of the macular lesions based on the input fundus photographs without a network, as well as generate screening reports. Referring to the guidelines for the application of AI screening systems for diabetic retinopathy [10], the time of the offline version of the AI system from the input of color fundus photos to the output of the screening results should be controlled within 1 min. It is recommended that the application organizations should retain the offline version of the screening system.

The online version of the AI system is suitable for application organizations with a nice network. The desensitized fundus color photos need to be uploaded to the cloud via a network. Then the cloud server should perform intelligent screening of macular area lesion signs and send results back. Afterward, the report combining the information of the examinees is generated for users to download. The online version

of the system has high requirements for hardware equipment, network transmission speed, and transmission security. Referring to the guidelines for the application of AI-assisted screening systems for diabetic retinopathy [10], the time of the online version of the AI system from the input of color fundus photos to the output of screening results should be controlled within 5 min.

5.2. Architecture and functionality of the AI screening system

The framework of the AI screening system should include patient management, image display, image quality assessment, AI screening result generation, the input of specialist's diagnostic opinion, and report generation. The function of the screening system depends on the actual application requirements, including the identification of the presence and grade of lesion signs, and the provision of the location and specific contour of lesion signs. The deployment of the system should be adapted to the environment and network conditions of the corresponding organization.

5.2.1. Patient management

The system allows the entry, modification, and viewing of patients' information. The examinee's information includes basic personal information (such as name, age, gender, medical ID, etc.) and relevant examination information (such as visual acuity, intraocular pressure, blood pressure, etc.). Detailed patient information is provided for doctors to verify that image information matches patient information without errors.

5.2.2. Image display

The fundus images of the right and left eyes are displayed. The images cannot be edited but can be zoomed in and translated to facilitate physician analysis.

5.2.3. Image quality assessment

This guideline recommends a quality scoring scheme to cover the four components of shot location, brightness, clarity, and whether the information is missing (see Section 4.1 for details) to determine whether the image quality can be used to observe or detect lesion signs in the macular region and also to provide an assessment of image quality (qualified, acceptable, unqualified).

5.2.4. AI screening

The most important module of the screening system is AI screening. According to the designed screening function, the system developer needs to train the corresponding AI model to complete the screening task. The output results of the AI model will be presented in the final report provided by the screening system and should be marked "the results are from the AI screening system". The form of screening results presented in the report depends on the preset function of the system, including printing whether the subject is a referral case, printing what specific lesion signs appear in the image, showing the location or profile of lesion signs, etc.

5.2.5. Specialist's diagnostic opinion

The specialist reviews the AI screening results and provides a diagnostic opinion. Referring to the existing guidelines for the application of AI-assisted screening systems at home and abroad [9], the specialist's diagnosis opinion is recommended to be provided within 24 h. This guideline recommends that the input module for expert-recommended diagnostic opinions be retained in the screening system.

5.2.6. Report generation

The screening report is provided to the examinees in an electronic or a paper form, and the content should cover the screening scope of this guideline and comply with the standards and specifications for writing the main health examination report [49].

5.2.7. Screening system deployment

The offline version of the AI screening system is deployed on a local device, and the online version is deployed remotely using the cloud.

5.3. Hardware requirements for the color fundus image acquisition

The AI screening system of the lesion sign in the macular region recommends the application of semiautomatic or fully automatic nonmydriatic flat color fundus cameras. The parameters of each index should refer to the following requirements:

(1) Focusing: the machine supports automatic focusing;
(2) Exposure: the machine supports automatic exposure;
(3) Field angle: when using a single field of view, the horizontal direction and the vertical direction ≥45° [35]. When using double fields of view, the horizontal direction and the vertical direction of each field of view ≥45°;
(4) The minimum pupil diameter that can be photographed: the minimum pupil diameter of the standard mode is about 4 mm, and the minimum pupil size of the small pupil mode is about 3.3 mm;
(5) Resolution: The minimum resolution within the field of view is not less than the industry standard requirements of fundus cameras [50];
(6) Refractive compensation range: ≥ ±15 D;
(7) Internal fixation point: The built-in fixation point of the fundus camera can be adjusted to at least 3 positions corresponding to the center of the field of view, including the center of the optic disk, the center of the macula, and the midpoint of the line connecting the optic disk and the macula, and the fixation point can be specified as needed;
(8) Image storage format: Support lossless compressed TIFF format or lossless compressed PNG format, support compressed JPEG format (compression ratio not greater than 10:1), and support DICOM format;
(9) Image shooting pixels: No less than 18 million pixels.

5.4. Standard of the image acquisition field of view

(1) Single-field imaging method: The midpoint of the line connecting the macula and the optic disk is the center of the field of view, and the imaging covers at least 45° of the retinal area [35] (Figure 12).
(2) Double-field imaging method: Field 1 takes the fovea as the center of the field of view, and the imaging covers at least 45° of the retinal area; Field 2 takes the optic disk as the center of the field of view, and the image covers at least 45° of the retinal area and can cover the area of 2 DD around the optic disk. The optic disk, the macular area, and the first branch of the upper and lower retinal vascular arches can be clearly observed (Figure 13).

The quality of the captured images needs to be judged according to the data quality control standards in Section 4.1, and the unqualified images need to be reshot.

5.5. AI screening scheme

5.5.1. Quality assessment for the color fundus images

The AI system could evaluate the image quality before analysis. This part is about evaluating the quality of the captured image from the three perspectives of shooting position, sharpness, and readability according to the image quality control standards.

(1) Shooting position: Check whether the deflection or displacement in each area of the image.
(2) Sharpness: Check the focusing and whether the structures such as the optic disk surface, retinal main vessels, retinal branch vessels, macular, and lesion are clearly discernible.

Figure 7. Schematic diagram of 'qualified' fundus color photo.

(3) Readable range: Check whether the optic disk, macular area, and upper and lower vascular arch areas are completely included in the field of view and whether there are eyelashes or other foreign objects blocked.

Image quality can be divided into three categories: qualified, acceptable, and unqualified. If the quality of the images during the test is qualified, the follow-up AI screening analysis will be carried out normally; if the image quality is acceptable, the operator can choose to continue the AI screening analysis, but it should be marked "Image quality deviation, the results are only for reference" in the "Screening Results". In addition, the operator can also choose to retake the image and if the image quality is unqualified, the operator must recollect the fundus image, if the image quality is not up to standard after three shots, the shooting fails, and subsequent AI screening and analysis cannot be performed.

5.5.2. Guidelines and principles for screening protocols

5.5.2.1. Guidance criteria for segmentation/localization of lesion signs in the macular region. For the lesion signs listed in Section 2.2 (drusen, geographic atrophy, neovascularization, exudation, hemorrhage, scarring, pigment mottling, etc.), it is recommended that the AI screening system should provide a visual display of the lesion segmentation results. To facilitate the doctor to observe the lesions, this guideline recommends that the visualization method provided by the screening system is the edge display of the lesion segmentation or the detection box display of the lesion localization.

5.5.2.2. Guidelines for the detection of lesion signs in the macular region. According to the definition of lesion signs listed in Section 2.2, it is recommended that the AI screening system provide the judgment results of whether the above lesions exist in the input fundus color photos.

5.5.2.3. Guidelines for screening the lesion signs in the macular region. The doctor first refers to the AI results, corrects the parts that are not approved, and then explains the content and significance of the AI screening results to the patient in detail. At the same time, it is necessary to explain that the AI-assisted diagnosis is not equivalent to the diagnosis of a professional doctor. Doctors can provide professional diagnosis and treatment recommendations for the next step based on the results of AI screening, combined with the subject's vision [51]and clinical manifestations:

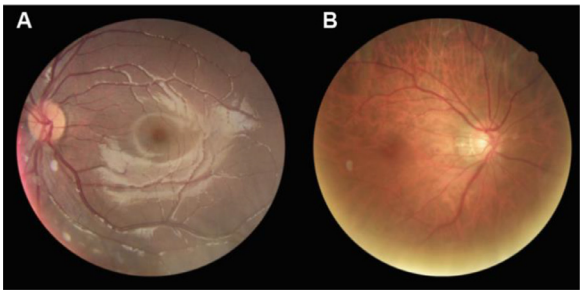

Figure 8. Schematic representation of 'acceptable' macular images. A: Visible reflections in the periphery, and the macular area is not affected; B: The marginal light leaks slightly, which does not affect the interpretation of the results.

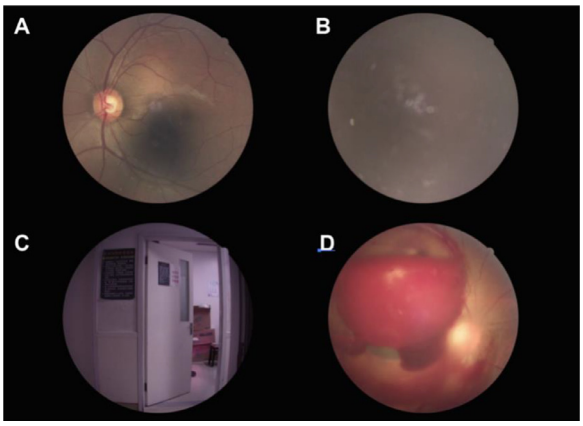

Figure 9. Schematic representation of 'unqualified' in color fundus images. A: Underexposure; B: Severe opacity of the refractive interstitium; C: Irrelevant images; D: Large coverage of the macular area.

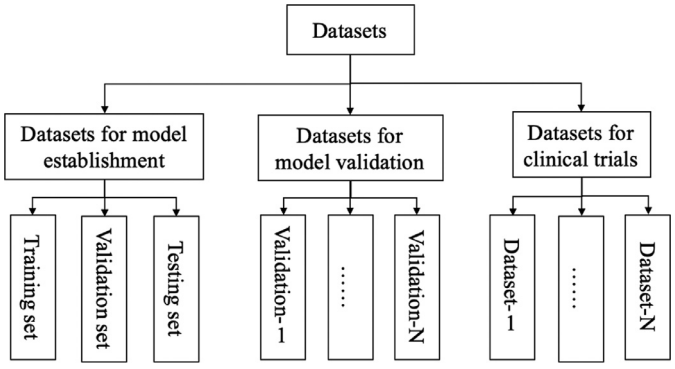

Figure 10. Dataset classification involved in the design and application of the screening system.

(1) For subjects without lesion signs in the macular region (low risk, no referral required), it is recommended for them to have a fundus health examination once a year;
(2) For subjects who are suspected to have 1 lesion sign in the macular region (medium risk, referral recommended) or who cannot be clearly determined to contain the lesion sign mentioned in Section 2.2 only based on the fundus color image, it is recommended to conduct a further examination;
(3) For subjects with at least one lesion sign in the macular region (high risk, confirmed referral), it is recommended to go to the hospital for a thorough examination as soon as possible.

5.6. Report requirements

(1) Basic information of the examinee: medical record number/medical ID, name, age, gender, and eye of the examinee;
(2) Ophthalmology-related information: vision, intraocular pressure, history of ophthalmic diseases, history of ophthalmic treatment (surgery, medication, etc.);
(3) Image acquisition information: date, equipment model, imaging range, and image storage format;
(4) Information about the screening system: the type and version of the algorithm, the model applicable to the algorithm, and the date of image analysis;
(5) Image evaluation: image quality;
(6) AI screening results: the diagnosis of the inspected eye, the probability of being sick or healthy, and the description of the pathological changes in the macular region. Specifically, it is recommended that the results should include the presence or absence of the lesion signs listed in Section 2.2 and the segmentation/localization display of the lesion signs;
(7) The opinion of the specialist on the proposed diagnosis;
(8) The description of rights and responsibilities.

5.7. Data storage

The storage of the fundus color images needs to meet the data storage security, the efficiency of management, and the convenience of HIS [52], PACS [53], and AI-assisted screening systems in use. The specific requirements are as follows.

(1) The model construction dataset can be stored by the development organization, the model validation dataset can be stored by the third-party testing organization, and the clinical trial dataset can be stored by each trial center; when the model is tested by the third-party testing organization and the clinical trial center, blind testing is recommended, i.e., the dataset does not flow into the model development organization;
(2) PNG, JPG, TIFF, or DICOM format for storage;
(3) The results of AI model calculations need to be saved and be able to be associated with the subject, examination information, and fundus color photographs;
(4) The data in the locally deployed and the cloud-deployed AI-assisted screening systems both need to be backed up. When the storage capacity reaches a certain threshold, the system administrator needs to update the storage;
(5) The docking of an AI-assisted screening system and the PACS and HIS in the hospital need to meet the medical-related protocol specifications.

5.8. Data safety

With reference to “Data Security Law of the People’s Republic of China”, “Personal Information Protection Law”, “Network Security Law”, and “Measures for the Management of Population Health Information”, the following things should be done.

(1) Data desensitization: Subjects’ sensitive information that does not affect disease diagnosis, such as name, phone number, ID number, home address, etc., should be desensitized. It should be noted that the original data characteristics should be maintained after desensitization. That is, development, testing, etc., should be not be affected by desensitization.
(2) User management: This includes user identity management, authentication management, and authorization management. User management systems should give each user a unique account and identify the user to ensure that the account can be traced back. Data access should have a unified identity authentication mechanism, and for sensitive data, multistep authentication technology should be used to prevent misuse. Access authorization for various types of data needs to be realized based on the identity of the user and the confidentiality level of the accessed data.
(3) Data usage log management: The system should capture a complete and unalterable record of activity, ensuring that every disruptive operation is recorded and audited, ensuring that operations are traceable, and thus aiding in the rapid location of malicious operations and attacks on the system.
(4) Transmission encryption: It is required to meet the needs of the platform as well as the transmission of sensitive data through secure transmission methods and standard encryption protocols to avoid illegal access, eavesdropping, or bypass sniffing of data. Setting up sensitive data transmission monitoring, data transmission correlation analysis, and other mechanisms to ensure that the transmission is carried out safely.
(5) Storage security: The data storage process needs to be protected with encryption measures to reduce the risk of data leakage. Encryption algorithms should use commercial cryptographic algorithms with security strength that meet national security requirements. In addition to data access control, the storage stage also needs to consider backup and disaster recovery, which can be achieved through local storage, network storage, and other ways to synchronize and backup data in multiple copies and data centers to achieve off-site disaster recovery.
(6) Clarify the responsible person: The institution where the AI screening system is applied should clarify the responsible person for data security and develop a management system and emergency mechanism for data security.

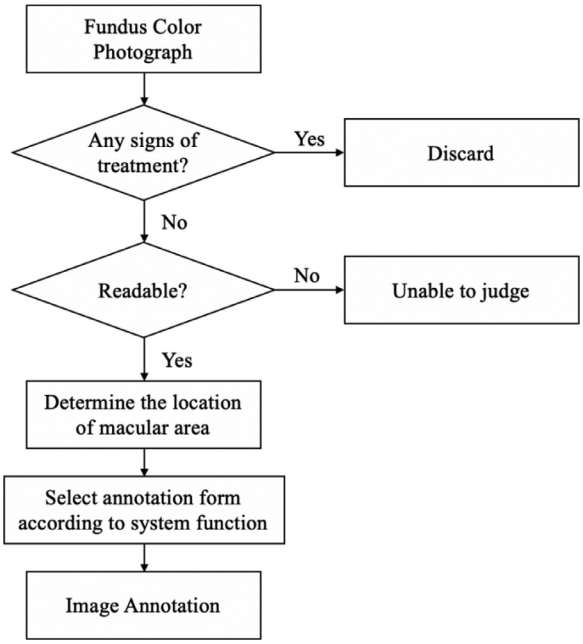

Figure 11. Labeling process for the AI screening system.

6. Limitations and future development of the AI screening system

Currently, the stability of AI screening systems dealing with the images captured by different types of cameras and with different qualities needs to be improved. Therefore, the existing AI-assisted screening systems need to be strictly constrained to shooting devices and shooting environments. It is worthwhile to look forward to the ap-

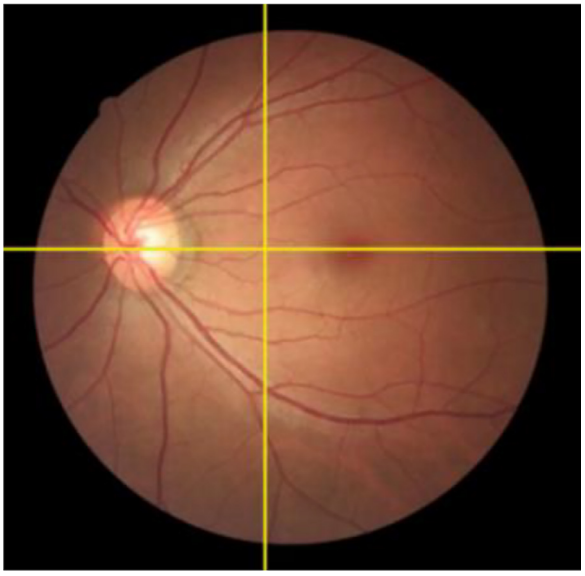

Figure 12. Schematic diagram of single-field imaging.

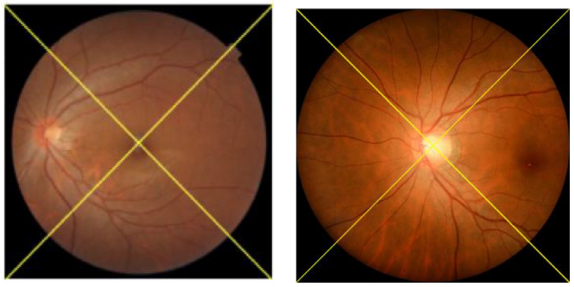

Figure 13. Schematic diagram of double-field photography method. A: Taking the fovea as the shooting center; B: Taking the optic disk as the shooting center.

pearance of the domain adaptation learning[6] in AI technology [54–55], which can alleviate the problem of poor robustness of AI models for screening the images of different quality or acquired by different camera types. In addition, the AI system based on deep learning techniques is often a "black box", and it is difficult to interpret for disease discrimination. This guideline takes the lesion sign as the detection unit, and in Section 3, the detection, localization, segmentation, and other tasks of lesion signs are proposed. Among them, the localization of lesion signs can be used as an indication of the area of concern when the subject is reviewed, prompting the physician to pay attention to the increase or decrease of lesion signs in the area over a period of time. Therefore, it is recommended to cover several subtasks in the disease diagnosis process when applying AI systems in medical scenarios.

This guideline focuses on the screening scene of lesion signs with single-modality data of fundus color images. Due to the complexity of macular diseases, single-modality data cannot provide comprehensive information for the ocular structures and lesions. Therefore, an AI-assisted screening system for macular diseases considering multiple fundus image modalities can be developed in the future. Notably, such assisted screening systems urgently require the development of fundus imaging hardware, such as the emergence of easily photographed, low-priced OCT acquisition devices. At present, it has been agreed that the application of the AI system in the medical needs to be reviewed by professional doctors. Therefore, AI can be regarded as a tool to assist doctors, and AI screening is not the same as a diagnosis by professional doctors. AI screening systems have good applications in the early screening of certain diseases/signs, which can alleviate the problem of a large screening population and an insufficient number of professional doctors. It is believed that in the future, more diseases/signs screening systems will appear with the help of AI technology (Figure 7).

[6] Domain adaptation learning: An artificial intelligence technology, a kind of transfer learning. It can effectively solve the problem of the inconsistent probability distribution of training samples and test samples. It is a hot topic in machine learning and has been widely used in natural language processing, text analysis, image analysis, bioinformatics, cross-language analysis, video analysis, emotion analysis, and handwriting recognition.

7. Abbreviation

ACC	accuracy
AI	artificial intelligence
AMD	age-related macular degeneration
AP	average precision
AUC	area under curve
CME	cystoid macular edema
CSME	clinically significant macular edema
DICE	dice coefficient
ETDRS	early treatment of diabetic retinopathy study
FN	false negative
FP	false positive
GA	geographic atrophy
HIS	hospital information system
ID	identity document
IoU	intersection over union
mAP	mean average precision
MED	mean euclidean distance
MNV	macular neovascularization
MR	miss rate
NPV	negative prediction value
PACS	picture archiving communication system
Pre	precision
ROC	receiver operating characteristic
RRD	rhegmatogenous retinal detachment
Sen	sensitivity
Spe	specificity
TN	true negative
TP	true positive

Expert Group on Guidance Development
Medicine group:
Ocular Fundus Diseases Group of Chinese Ophthalmological Society

Ming Zhang, Department of Ophthalmology, West China Hospital, Sichuan University (Director, Writing)

Xun Xu, Department of Ophthalmology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine (Group leader)

Yusheng Wang, Department of Ophthalmology, Xijing Hospital, the Fourth Military Medical University (Vice group leader)

Wenbin Wei, Beijing Tongren Eye Centre, Beijing Key Laboratory of Intraocular Tumour Diagnosis and Treatment, Beijing Tongren Hospital, Capital Medical University (Vice group leader)

Mingwei Zhao, Department of Ophthalmology, Ophthalmology and Optometry Centre, Peking University People's Hospital (Vice group leader)

Youxin Chen, Department of Ophthalmology, Peking Union Medical College Hospital

Xiaorong Li, Tianjin Key Laboratory of Retinal Functions and Diseases, Tianjin International Joint Research and Development Centre of Ophthalmology and Vision Science, Eye Institute and School of Optometry, Tianjin Medical University Eye Hospital

Fang Lu, Department of Ophthalmology, West China Hospital, Sichuan University

Qinghuai Liu, Department of Ophthalmology, the First Affiliated Hospital of Nanjing Medical University

Bo Lei, Department of Ophthalmology, Henan Eye Institute, Henan Eye Hospital, Henan Provincial People's Hospital, People's Hospital of Zhengzhou University

Jinfeng Qu, Department of Ophthalmology, Ophthalmology and Optometry Centre, Peking University People's Hospital

Feng Wen, Department of Ophthalmology, State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Centre, Sun Yat-sen University

Suqin Yu, Department of Ophthalmology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine

Weihua Yang, The Affiliated Eye Hospital of Nanjing Medical University, Jiangsu Province Red Cross Eve Hospital

Shaochong Zhang, Shenzhen Eye Hospital

Liang Du, Department of Clinical Epidemiology and Evidence-Based Medicine, West China Hospital, Sichuan University

Dongyue Wang, Department of Ophthalmology, West China Hospital, Sichuan University (Secretary)

AI group:

Yanwu Xu, Beijing Baidu Netcom Science Technology Co. Ltd. (Writing)

Jiang Liu, Department of Computer Science and Engineering, Southern University of Science and Technology

Xiangmin Xu, School of Future Technology, South China University of Technology

Jie Chen, Pengcheng Laboratory

Mingkui Tan, School of Software, South China University of Technology

Yong Xia, School of Computer Science, Northwestern Polytechnical University

Baiying Lei, Health Science Center School of Biomedical Engineering, Shenzhen University

Yitian Zhao, Cixi Institute of BioMedical Engineering

Pengzhi Zhu, Guangdong Medical Devices Quality Surveillance and Test Institute

Huihui Fang, Beijing Kangfuzi Health Technology Co. Ltd (Secretary)

Dalu Yang, Beijing Baidu Netcom Science Technology Co. Ltd.

Qinpei Sun, Beijing Kangfuzi Health Technology Co. Ltd

Siqi Zhang, Beijing Baidu Netcom Science Technology Co. Ltd.

Expert Group on Guidance Review

Medicine group:

Ke Yao, Eye Center, Second Affiliated Hospital, School of Medicine, Zhejiang University

Xiaoxin Li, Fujian Provincial Key Laboratory of Ocular Surface and Corneal Diseases; Eye Institute and Affiliated Xiamen Eye Center of Xiamen University, School of Medicine, Xiamen University

Ningli Wang, Beijing Tongren Eye Centre, Beijing Key Laboratory of Intraocular Tumour Diagnosis and Treatment, Beijing Tongren Hospital, Capital Medical University

Xinghuai Sun, Department of Ophthalmology and Vision Science, Eye & ENT Hospital, Shanghai Medical College, Fudan University

Yannian Hui, Department of Ophthalmology, Xijing Hospital, the Fourth Military Medical University

Changzheng Cheng, Department of Ophthalmology, People's Hospital of Wuhan University

Shiying Li, Ophthalmology Department, Southwest Hospital, Army Medical University (Third Military Medical University); Department of Ophthalmology, Xiang'An Hospital of Xiamen University

Xiaolin Liu, Eye Hospital, WMU Zhejiang Eye Hospital

Hai Lu, Beijing Tongren Eye Centre, Beijing Key Laboratory of Intraocular Tumour Diagnosis and Treatment, Beijing Tongren Hospital, Capital Medical University

Lin Lv, Department of Ophthalmology, State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Centre, Sun Yat-sen University

Xiaoyan Peng, Beijing Institute of Ophthalmology, Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University, Beijing Ophthalmology and Visual Science Key Lab

Yin Shen, Department of Ophthalmology, People's Hospital of Wuhan University

Xiaodong Sun, Shanghai Key Laboratory of Ocular Fundus Diseases, Shanghai General Hospital, Shanghai JiaoTong University

Min Wang, Department of Ophthalmology and Vision Science, Eye & ENT Hospital, Shanghai Medical College, Fudan University

Gezhi Xu, Department of Ophthalmology and Vision Science, Eye & ENT Hospital, Shanghai Medical College, Fudan University

Hong Yan, Department of Ophthalmology, Tangdu Hospital, Fourth Military Medical University

Hua Yan, Department of Ophthalmology, Tianjin Medical University General Hospital

Weihong Yu, Department of Ophthalmology, Peking Union Medical College Hospital

Han Zhang, Department of Ophthalmology, the First Hospital of China Medical University

Peiquan Zhao, Department of Ophthalmology, Xinhua Hospital affiliated to Shanghai Jiao Tong University School of Medicine

AI group:

Dong Min, Institute of Cloud Computing and Big Data, China Academy of Information and Communication Technology

Hao Wang, National Institutes for Food and Drug Control

Huiguang He, Institute of Automation, Chinese Academy of Sciences

Lixin Duan, School of Computer Science and Engineering, University of Electronic Science and Technology of China

Qingyao Wu, School of Software, South China University of Technology

Declaration

This guideline is developed in strict accordance with the guidelines of the World Health Organization and the Chinese Medical Association and references the research and evaluation tools of guidelines and the standard of international practice guidelines reports. Its purpose is to provide guidance for primary fundus screening services and is not a medical standard that must be followed in all situations, nor is it a health measure for individual cases. The accuracy, completeness, legality, reliability, and operability of the information or data included in this guideline are not liable for any legal responsibility. During the development of this guideline, the regulations on conflicts of interest in WHO guideline development and the ethical standards for guideline development are strictly followed, and all participating members have filled the conflict-of-interest declaration form. After evaluation, there is no direct conflict of interest in the development of this guideline.

Conflicts of interest statement

The authors declare that there are no conflicts of interest.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.imed.2023.05.001.

References

[1] Chen H. Diagnostic and Management of Diabetic Retinopathy Screening Manual -. 2nd Ed. Beijing: People's Medical Publishing House; 2016.

[2] Sha WR. China's White Papers on Eye Health: Blinding eye diseases can be effectively curbed. China Medicine and Pharmacy 2020;10(13):3–5. doi:10.3969/j.issn.2095-0616.2020.

[3] Balyen L, Peto T. Promising Artificial Intelligence-Machine Learning-Deep Learning Algorithms in Ophthalmology. Asia Pac J Ophthalmol (Phila) 2019;8(3):264–72. doi:10.22608/APO.2018479.

[4] Orlando JI, Fu H, Breda JB, et al. Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. Med Image Anal 2020;59:101570. doi:10.1016/j.media.2019.101570.

[5] Wu LZH, Lin SCH. Clinical Macular Disease. Beijing: Beijing Science and Technology Press; 2007.

[6] Esteva A, Kuprel B, Novoa RA, et al. Corrigendum: Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;546(7660):686. doi:10.1038/nature22985.

[7] Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. JAMA 2017;318(22):2199–210. doi:10.1001/jama.2017.14585.

[8] Gargeya R, Leng T. Automated Identification of Diabetic Retinopathy Using Deep Learning. Ophthalmology 2017;124(7):962–9. doi:10.1016/j.ophtha.2017.02.008.

[9] Glaucomatology GroupOphthalmology Society of Chinese Medical Association, Chinese Medical Equipment Association Ophthalmic artificial Intelligence Group. Standardized design and application guide of AI-assisted screening system for glaucoma based on fundus photography in China. Chinese Journal of Ophthalmology 2020;56(6):423–32. doi:10.3760/cma.j.cn112142-20200102-00003.

[10] Yuan J. Intelligent Medicine Special Committee of China Medicine Education Association, China "Dev. and Appl. of Ophthalmic Multimodal Imaging and AI Diagnosis and Treat Sys"Team. Guidelines for artificial intelligent diabetic retinopathy screening system based on fundus photography. Chinese Journal of Experimental Ophthalmology 2019;37(8):593–8. doi:10.3760/cma.j.issn.2095-0160.2019.08.001.

[11] National Medical Products Administration. Artificial intelligence Medical Device Quality Requirements and Evaluation Part 1: Terminology. Available from http://app.nifdc.org.cn/biaogzx/qxqwk.do?formAction=view&id=2c9048d881c8ca520181d69c35d44998 (Accessed on May 29, 2023).

[12] National Medical Products Administration. Artificial intelligence Medical Device Quality Requirements and Evaluation Part 2: General Requirements for Datasets. Available from http://app.nifdc.org.cn/biaogzx/qxqwk.do?formAction=view&id=2c9048d881c8ca520181d69cd9e149ae (Accessed on May 29, 2023).

[13] National Medical Products Administration. Artificial intelligence Medical Device Quality Requirements and Evaluation Part 3: General Requirements for Data Annotation. Available from http://app.nifdc.org.cn/biaogzx/qxqwk.do?formAction=view&id=2c9048d882c4f3180182c936a8b124f9 (Accessed on May 29, 2023).

[14] China Association for Quality InspectionAnnotation and quality control specifications for fundus color photograph. Intell Med 2021;1(02):80–7. doi:10.1016/j.imed.2021.05.006.

[15] Davis MD, Gangnon RE, Lee LY, et al. The Age-Related Eye Disease Study severity scale for age-related macular degeneration: AREDS Report No. 17. Arch Ophthalmol 2005;123(11):1484–98. doi:10.1001/archopht.123.11.1484.

[16] Britton G, Liaaenjensen S, Hanspeter P. Carotenoids Volume 5: Nutrition and Health. 2009. doi:10.1007/978-3-7643-7501-0.

[17] American Academy of Ophthalmology Preferred Practice Pattern. Beijing: People's Medical Publishing House(PMPH); 2013.

[18] Zhang CHF. Disease of Ocular Fundus. Beijing: People's Medical Publishing House(PMPH); 2010.

[19] Yang YL, Yang TT, Lu F, et al. Brief interpretation of the consensus nomenclature for reporting neovascular age-related macular degeneration data. Chinese Journal of Ocular Fundus Diseases 2022;38(2):99–107. doi:10.3760/cma.j.cn511434-20220128-00055.

[20] Chinese Association of Ophthalmology Ophthalmology Group senile macular degeneration in China. Clinical pathway of age-related macular degeneration in China. Chinese Journal of Ocular Fundus Diseases 2013;29(004):343–55. doi:10.3760/cma.j.issn.1005-1015.2013.04.002.

[21] Wen F, Zhang XZ. Classification and clinical significance of retina hemorrhage. Ophthalmol CHN 2009;18(4):221–4.

[22] Sarks J, Tang K, Killingsworth M, et al. Development of atrophy of the retinal pigment epithelium around disciform scars. Br J Ophthalmol 2006;90(4):442–6. doi:10.1136/bjo.2005.083022.

[23] Gass JD. Reappraisal of biomicroscopic classification of stages of development of a macular hole. Am J Ophthalmol 1995;119(6):752–9. doi:10.1016/s0002-9394(14)72781-3.

[24] Zhang HR, Wang W. Idiopathic preretinal macular membrane. Chinese Journal of Practical ophthalmology 1997;15(10):4.

[25] Early Treatment Diabetic Retinopathy Study design and baseline patient characteristics. ETDRS report number 7. Ophthalmology 1991;98(5 Suppl):741–56. doi:10.1016/s0161-6420(13)38009-9.

[26] Sigler EJ, Randolph JC, Kiernan DF. Longitudinal analysis of the structural pattern of pseudophakic cystoid macular edema using multimodal imaging. Graefes Arch Clin Exp Ophthalmol 2016;254(1):43–51. doi:10.1007/s00417-015-3000-8.

[27] Yan M. Cystoid macular edema. Chinese Journal of Ocular Fundus Diseases 2002;18(3):234–5.

[28] Zayit-Soudry S, Moroz I, Loewenstein A. Retinal pigment epithelial detachment. Surv Ophthalmol 2007;52(3):227–43. doi:10.1016/j.survophthal.2007.02.008.

[29] Daniel AB, Wilkinson CP. Retinal Detachment: Principles and Practice. 3rd edition. Beijing: People's Medical Publishing House(PMPH); 2011.

[30] Madjarov G, Kocev D, Gjorgjevikj D, et al. An extensive experimental comparison of methods for multi-label learning. Pattern Recognit 2012;45(9):3084–104. doi:10.1016/j.patcog.2012.03.004.

[31] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159–74.

[32] Liu L, Ouyang W, Wang X, et al. Deep learning for generic object detection: A survey. Int J Comput Vis 2020;128(2):261–318. doi:10.1007/s11263-019-01247-4.

[33] Li T, Bo W, Hu C, et al. Applications of deep learning in fundus images: A review. Med Image Anal 2021;69:101971. doi:10.1016/j.media.2021.101971.

[34] Zhang YJ. A survey on evaluation methods for image segmentation. Pattern Recognit 1996;29(8):1335–46. doi:10.1016/0031-3203(95)00169-7.

[35] Li J, Xu L, Peng XY, et al. Quality Standard of Single Fundus in Remote Ophthalmology (Draft for comments). Ophthalmology 2015;24(1):11–12.

[36] Ophthalmology Branch of Chinese Medical Association, Ophthalmologist Branch of Chinese Medical Doctor Association fundus disease Professional Committee. Guidelines for image collection and film reading for screening diabetic retinopathy in China (2017). Chinese Journal of Ophthalmology 2017;53(012):890–6. doi:10.3760/cma.j.issn.0412-4081.2017.12.003.

[37] Zhou Z. Machine Learning. Beijing: Tsinghua University Press; 2016.

[38] Gulshan V, Peng L, Coram M, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA 2016;316(22):2402–10. doi:10.1001/jama.2016.17216.

[39] National Medical Products Administration. Guidelines for Registration Review of Artificial Intelligence Medical Devices. Available from https://www.cmde.org.cn/flfg/zdyz/zdyzwbk/20220309091014461 (Accessed on May 29, 2023).

[40] Thompson SK. *Sampling*. 2nd ed. New Jersey: John Wiley & Sons, Inc; 2012.

[41] National Medical Products Administration. Standard for Quality Management of Clinical Trials on Medical Devices. Available from https://www.nmpa.gov.cn/xxgk/fgwj/bmgzh/20160323141701747.html?type=pc&m= (Accessed on May 29, 2023).

[42] National Medical Products Administration. Key points and related description of deep learning assisted decision making for medical devices. Availabe from https://www.cmde.org.cn//xwdt/zxyw/20190628151300923.html (Accessed on May 29, 2023).

[43] Jonas JB, Xu L, Wang YX. The Beijing Eye Study. Acta Ophthalmol 2009;87(3):247–61. doi:10.1111/j.1755-3768.2008.01385.x.

[44] Yang Q, Han C, Liu N, et al. Application of non-mydriatic digital fundus photography in fundus macular disease screening. Recent Advances in Ophthalmology 2012;32(05):473–5.

[45] Cao X, Yao Y, Fu D. Prevalence study of age-related macular degeneration over the age of 50's in Wuxi. Chinese Journal of Practical Ophthalmology 2013(04):494–8.

[46] Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. J Big Data 2019;6(1):1–48.

[47] Padilla R, Netto SL, Silva Da, et al. A survey on performance metrics for object-detection algorithms. In: Proceedings of 2020 international conference on systems, signals and image processing (IWSSIP); 2020.

[48] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks. Advances in Neural Information Processing Systems 2015:28.

[49] Health Management Branch of Chinese Medical Association, Editorial Board of Chinese Journal of Health Management. Expert consensus on the chief physician report for health checkup. Chinese Journal of Health Management 2020;14(1):8–11. doi:10.3760/cma.j.issn.1674-0815.2020.01.003.

[50] State Food and Drug Administration Ophthalmic instruments-Fundus cameras. Beijing: Standards Press of China; 2008.

[51] Song LL, Li ZQ, Ma WJ, et al. Establishment of a hierarchical diagnosis and treatment model for diabetic retinopathy by cloud medical remote platform. Modern Hospital Management 2020;18(3):17–20. doi:10.3969/j.issn.1672-4232.2020.03.005.

[52] Tan DH, Wan LF, Zou YH. Hospital Information system HIS and its application. China Medical Device Information 2007;13(9):39–40.

[53] Li G, Liu G, Li G. The design and implementation of PACS system in large-scale comprehensive hospitals. Chinese Hospital Management 2005;25(1):46–8.

[54] Long M, Zhu H, Wang J, et al. Unsupervised domain adaptation with residual transfer networks. In: Proceedings of 30th Annual Conference on Neural Information Processing Systems, NIPS 2016; Barcelona; Spain; 2016.

[55] Ganin Y, Lempitsky V. Unsupervised domain adaptation by backpropagation. In: Proceedings of International conference on machine learning; 2015.