Heliyon 9 (2023) e16149

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

A grading method for Kayser Fleischer ring images based on ResNet

Wei Song [a], Ling Xin [a,*], Jiemei Wang [b]

[a] The First Affiliated Hospital of Anhui University of Chinese Medicine, Hefei, 230031, China
[b] Department of Otolaryngology, The First Affiliated Hospital of Anhui University of Chinese Medicine, Hefei, 230031, China

ARTICLE INFO

Keywords:
HLD
Wilson's disease
YOLO algorithm
K–F
ResNet

ABSTRACT

The corneal K–F ring is the most common ophthalmic manifestation of WD patients. Early diagnosis and treatment have an important impact on the patient's condition. K–F ring is one of the gold standards for the diagnosis of WD disease. Therefore, this paper mainly focused on the detection and grading of the K–F ring. The aim of this study is three-fold. Firstly, to create a meaningful database, the K–F ring images are collected which contains 1850 images with 399 different WD patients, and then this paper uses the chi-square test and Friedman test to analyze the statistical significance. Subsequently, the all collected images were graded and labeled with an appropriate treatment approach, as a result, these images could be used to detect the corneal through the YOLO. After the detection of corneal, image segmentation was realized in batches. Finally, in this paper, different deep convolutional neural networks (VGG, ResNet, and DenseNet) were used to realize the grading of the K–F ring images in the KFID. Experimental results reveal that the entire pre-trained models obtain excellent performance. The global accuracies achieved by the six models i.e., VGG-16, VGG-19, ResNet18, ResNet34, ResNet50, and DenseNet are 89.88%, 91.89%, 94.18%, 95.31%, 93.59%, and 94.58% respectively. ResNet34 displayed the highest recall, specificity, and F1-score of 95.23%, 96.99%, and 95.23%. DenseNet showed the best precision of 95.66%. As such, the findings are encouraging, demonstrating the effectiveness of ResNet in the automatic grading of the K–F ring. Moreover, it provides effective help for the clinical diagnosis of HLD.

1. Introduction

Hepatolenticular degeneration (HLD) is also known as Wilson's Disease (WD) [1]. It is characterized by copper accumulation in different tissues and organs such as the liver, nervous system, kidney, and eyes [2]. It is an autosomal recessive genetic disease of copper metabolism disorder caused by a pathogenic mutation of the ATP7B gene [3–7]. The age of onset of WD is broad, mainly in children and adolescents, and the age of onset is concentrated between 5 and 35 years old. The minimum age of onset can be 8 months old children, and the maximum age of onset can be 80 years old [8,9]. In childhood, most of the patients were asymptomatic and showed slight liver damage. In adults, most of the patients had typical corneal K–F ring and neurological symptoms. The worldwide incidence rate of WD is 1/30 000–1/100,000, and the carrier of the disease causing gene is about 1/90 [7]. This disease is common in China, with an incidence rate of about 5.87/100,000 and an incidence rate of 5.87/100,000. There are a few more males than females

* Corresponding author.
E-mail addresses: weisong_1988@163.com (W. Song), xinling_ahtcm@foxmail.com (L. Xin), 1378076538@qq.com (J. Wang).

https://doi.org/10.1016/j.heliyon.2023.e16149
Received 12 December 2022; Received in revised form 4 May 2023; Accepted 7 May 2023
Available online 13 May 2023

Fig. 1. The performance of K–F ring. (a) shows the K–F ring does not exist. (b) shows the K–F ring exists.

Table 1
Literature reviews summary.

	Model/Method	Disease	Best accuracy
Alam Talha Mahboob et al. [26] & Li Ziyi et al. [27]	Resnet	skin damage	91.0%,98%
Devnath Liton et al. [28]	CheXNet	pneumoconiosis	91.5%
Srinivas Chetana et al. [29]	VGG	brain tumor	95%
Kumar M. Rupesh et al. [30]	SVM, REP tree, PRCNN	dementia	87.6%
Wang Yi et al. [31]	U-Net, ResNet	Placenta segmentation	98.6%
Sahli Hanene et al. [32]	ResNet-SVM	tumor segmentation	89.36%
Alghamdi Hanan Saleh [33]	ResNet; DenseNet	diabetic retinopathy	91.1%

in patients with WD, and inappropriate treatment will lead to disability and even death. WD is also one of the few neurogenetic diseases that can be treated. The key is early detection, early diagnosis, and early treatment. In addition, the level of screening, diagnosis, and treatment and the level of professional knowledge of doctors in some remote areas are also of great significance to the determination of diagnosis.

The corneal K–F (Kayser Fleischer) ring is the most common ophthalmic manifestation of WD patients. K–F ring usually exists in 90.4%–100% of WD patients with neuropsychiatric symptoms, and 50–60% of WD patients without neurological symptoms, but only in 10%–40% of asymptomatic WD patients [10–13]. The K–F ring is named after Kayser (1902) and Fleischer (1903) respectively. It is a brown pigment ring (also yellow-green, ruby red or dark blue) located at the edge of the corneal posterior elastic layer. It first occurs at the top of the cornea and then inferiorly and finally appears circular, like a ring. It may be disappeared or faded after professional treatment [14,15]. Therefore, the reflection of the K–F ring can also be used as a measure of the success of treatment. Fig. 1(a) and (b) shows different images of the K–F ring, in which Fig. 1(a) shows the normal corneal without the K–F ring, and Fig. 1(b) shows the existence of the K–F ring. At present, the corneal K–F ring, as a characteristic clinical manifestation of WD patients, is one of the gold standards for the diagnosis of WD disease. It is of great significance to diagnose WD by examining the presence of the K–F ring in patients through a slit lamp.

As mentioned above, the K–F ring is not used as a pathological diagnostic indicator of WD in general, but it can be used as a diagnostic standard and monitor treatment response. Therefore, clinicians in the neurology department need to consult with ophthalmologists to diagnose patients suspected of having WD due to clinical and biochemical findings or evaluate the diagnosis and treatment effect. Ophthalmologists usually use slit lamps for K–F ring examinations. Unfortunately, for non-professional neurologists or ophthalmologists, the existence of the K–F ring may be ignored, especially for inexperienced clinicians. It is difficult to make a diagnosis with this method. Therefore, it is of great significance to construct an applicable, available, and accurate the K–F ring image database for later research.

Deep learning algorithms can automate professional-level diagnostic tasks in many areas, such as industry [16–18], engineering [19–23], agriculture [24], medical [25–29]. In the medical field, many people have studied the application of deep learning in medical examination image classification. Ali Zain et al. [25] proposed a framework for early prediction of Schistosomiasis, and the proposed framework performed with good accuracy of 87.1%. Alam Talha Mahboob et al. [26] and Li Ziyi et al. [27] use the models, which were based on ResNet (Residual Network) as a classification method for multi-class skin damage images. And the experiments show that the proposed methods show excellent performance with the classification accuracy rate, which provide effective help for clinical diagnosis. Devnath Liton et al. [28] proposed ensemble learning techniques for detecting pneumoconiosis disease in chest X-ray radiographs (CXRs) using multiple deep learning models. Srinivas Chetana et al. [29] use the pre-trained VGG16 to realize the brain tumor classification. Kumar M. Rupesh et al. [30] proposed a framework which based on different models to recognize dementia. Wang Yi et al. [31] put forward an improved U-Net placenta segmentation network based on ResNet, which achieves ideal results. Sahli Hanene et al. [32] propose a fusion method (ResNet-SVM) that has increased classification results of accuracy. ResNet's medical application also includes the detection of diabetic retinopathy. Alghamdi Hanan Saleh [33] uses explainable Deep Neural Networks to detect diabetic retinopathy. The literature review can be summarized in Table 1.

In addition, this paper reviewed the YOLO (You only look once) algorithm in the application of detection. In the development of object detection, automatic, efficient, and accurate detection of facial information is a key prerequisite [34–37]. As a non-contact technology, image-based or video analysis technology can automatically monitor and perceive the cornea without pressure. At present, deep learning with automatic feature extraction and strong image representation capabilities makes vision-based object detection and behavior recognition more efficient. Therefore, the combination of depth learning and video technology to detect the corneal region is of great significance for the segmentation of the images and the high accuracy of the grading of K–F rings.

Lee HoWon et al. put forward the post disaster personnel detection based on YOLO, and the average accuracy rate reached more than 89% [38]. Kohei Arai and Akihito Yamashita et al. proposed the pedestrian safety detection based on YOLO target detection and achieved ideal goals [39].

YOLO algorithm, is the most typical representative of a one-stage target detection algorithm [40]. Yolo is based on deep neural networks for object identification and positioning, which is fast to run and can be used in real-time systems. YOLOV7 is the most advanced algorithm of the YOLO series at present, surpassing the previous YOLO series in accuracy and speed [41,42]. Therefore, it is feasible to use YOLO for rapid and precise detection of the corneal region.

As such, deep learning algorithms can provide great help in the medical field, especially the application of ResNet model in the medical field. However, the current research does not involve an available image database that is related to the K–F ring and its application to K–F image libraries. Therefore, there is a lack of research and application of deep learning algorithms on K–F images. In this paper, an available database was constructed. Moreover, automatic grading is implemented by different models.

In this paper, the aim mainly focus on the quality assessment of K–F ring examination images of WD patients. First, this paper collected and sorted out 1850 images of 399 WD patients in the hospital from January 2018 to June 2022 through slit lamp corneal K–F ring examination. At the same time, the images were quantified with different grades according to Wang Gongqiang et al. [34] and Degirmenci, C et al. [11]. Second, this paper statistically analyzed the relevant attribute information of the patient's K–F ring image and established a useable corneal K–F ring images database (KFID) after analysis. Finally, this paper verified the usability of the image database by implementing different models.

The contributions of this paper can be summarized as follows:

(1) The first dedicated large-scale natural KFID was built in this paper, which contains 1850 images with 399 different WD patients.
(2) The images in KFID were been graded in different grades by professional clinic doctors. Meanwhile, the statistical significance of the images was been analyzed in 399 WD patients. As a result, the statistical relationship was been obtained in this paper, which was among corneal K–F ring, different ages and different grades. At the same time, the correlation between different ages and the corneal K–F ring grades was been analyzed.
(3) All images were been labeled for detection. In addition, the YOLO algorithm was used to detect the corneal. After the detection, all images were been resized.
(4) In addition, deep convolutional neural networks (VGG, ResNet, and DenseNet) were used to conduct experiments on K–F grading. The results show that the method used in this paper can achieve ideal results.

The remainder of this paper is organized as follows. Section 2 introduces the materials and methods. Section 3 introduces the experiment and results. Section 4 presents the limitations of the study and the conclusion.

2. Materials and methods

2.1. Database construction

2.1.1. Collection of the images

In order to complete this research, we need to obtain images of WD patients through slit lamp examination. Therefore, we collect the images from January 2018 to June 2022 of the hospital, which contains 1850 images of 399 patients. All of the images are obtained through slit lamp examination, which is operated by professional doctors. After obtaining the images, we concealed the basic information to protect the patients' privacy.

2.1.2. Construction

To investigate the visual quality of K–F ring images, we construct a dedicated large-scale natural K–F ring image database (KFID). The images in the KFID database cover all the patient's eye examination images, which are completed by professional doctors through slit lamps. And all of the resolutions of the images were consistent.

KFID contains 1850 K–F ring images with 399 different patients. All the images, we used, were obtained with the same instrument. In addition, the settings are identical. For each patient, the operator will capture two or more images, which include the patient's eyes (left and right). In our research, we take the direct focus illumination to capture the K–F ring images. The settings are the same for different image contents. The resolutions of the original images are all 3264×2448. We obtain the resized images in Section 2 to keep the properties of the images. To include more real-world research, the image contents of the same patient with different examination times do not overlap. The statistical information about the KFID will be discussed in later chapters.

2.1.3. Subjective assessment grades

Subjective assessment methodologies of K–F ring images have been recommended by Wang Chongqiang etc. [34] and Degirmenci,

Table 2
Grading standard of K–F ring.

Grades	Standard
0	No K–F ring is found
1	K–F ring can be seen above or below the corneal limbus in a dotted distribution, with a range ≤1/2
2	K–F ring can be seen above or below the corneal limbus in crescent-shaped distribution, the range ≤1/2
3	The K–F ring is arc-shaped, accounting for more than 1/2 of the limbus, but not all corneas
4	K–F ring is distributed in a round shape, and all corneal limbus have pigment ring

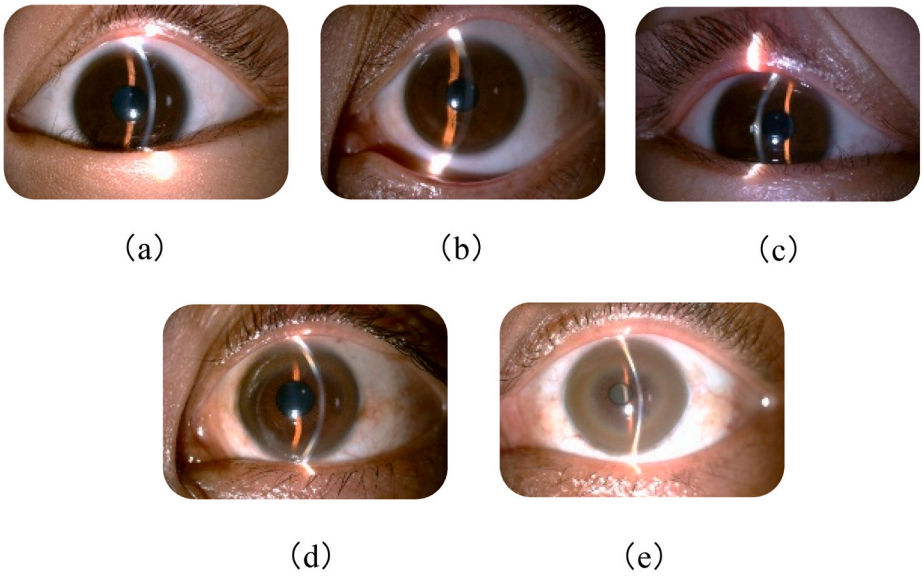

Fig. 2. The performance of different grades of K–F ring. (a) shows grade 0. (b) shows grade 1. (c) shows grade 2. (d) shows grade 3. (e) shows grade 4.

C, etc. [11], including the scope and Scheimpflug imaging. In this research, K–F ring were graded in five grades, of which is grade 0 to 4. Table 2 shows the classification standard of K–F ring images in this paper.

The relation between different grades and K–F ring images is shown in Fig. 2, in which Fig. 2(a)–(e) express the grade 0–4 respectively. Given an image displayed on the screen, a clinician is asked to give a grade on the image based on her/his professional skills. During the grading process, each image is presented in a random order to reduce the memory effect on grades.

At the same time, this paper gives the corresponding color histogram and gray level histogram of K–F ring images with different grades, which is shown in Fig. 3. Fig. 3(a) denotes the gray level histogram and Fig. 3(b) denotes the color histogram.

2.2. Data preprocessing

The images were originally obtained by slit lamp inspection. In those original images, there will be noises in the corners (inner and outer) and eyelids (upper and lower). These noises will seriously affect the accuracy of deep learning in K–F ring grading, which will make the trained model unusable.

In order to verify the impact of noise, this paper completed 10 experiments in total, which were based on ResNet18. Through the experiment, it was found that the average accuracy rate in the grading of the train is 81.62%, and the average accuracy rate of the verification is 75.63%.

After feature extraction, this paper found that the pre-trained ResNet18 model takes the corners of the canthus (inside and outside) and parts of the eyelids (top and bottom) as features, which is inconsistent with the symptoms of hepatolenticular degeneration in the eyes. Therefore, In order to establish an effective database of KFID, The following pre-processing measures are proposed.

2.2.1. Label images

In the process of image labeling, three professional doctors were invited to label 1850 images, including ophthalmologists and neurologists. In this research, LabelImg was used for image labeling, which was commonly used for image labeling. After labeling images, three doctors checked the labeling results of the other two doctors. In addition, if there were any objections to the labeling result, it shall be labeled again after discussion. An example of the labeled image is shown in Fig. 4.

In this section, the image labels were saved as a text document for later corneal detection. In this text document, there are five

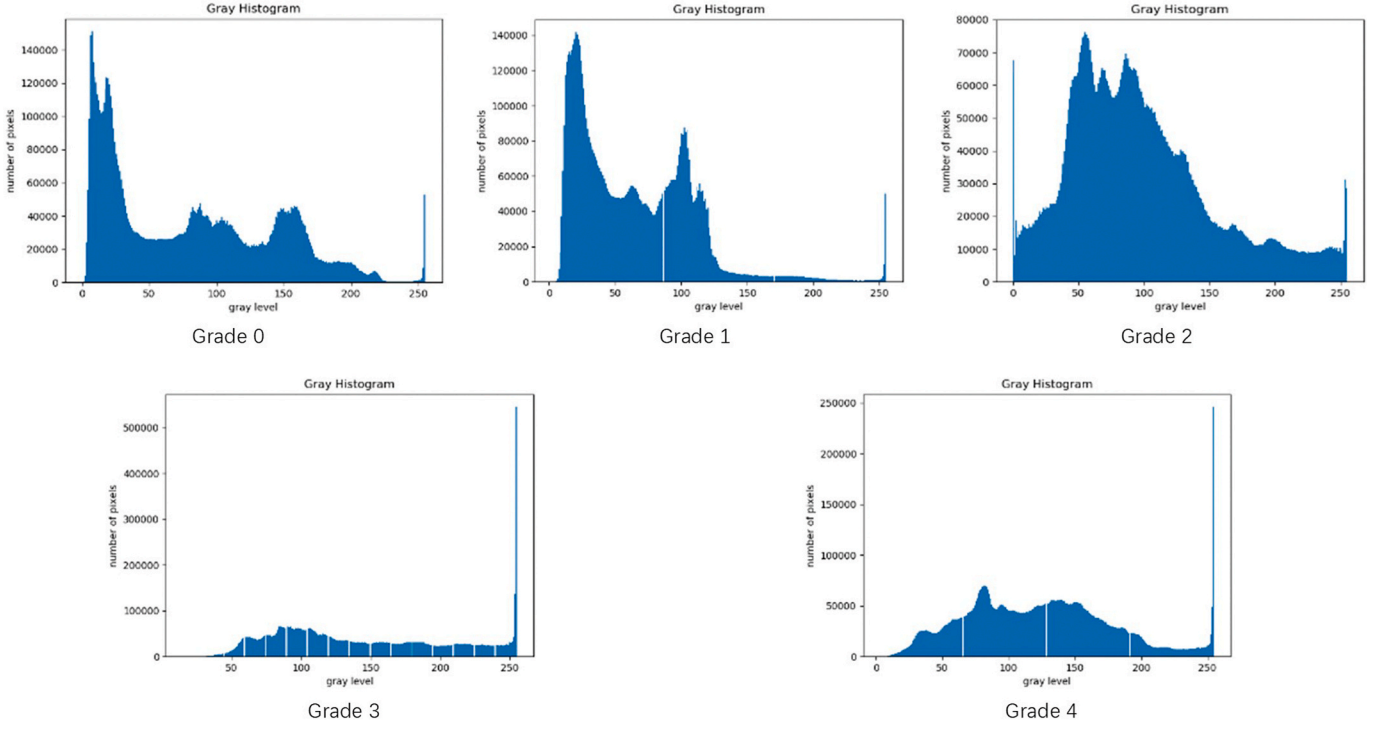

Fig. 3a. Gray level histogram.

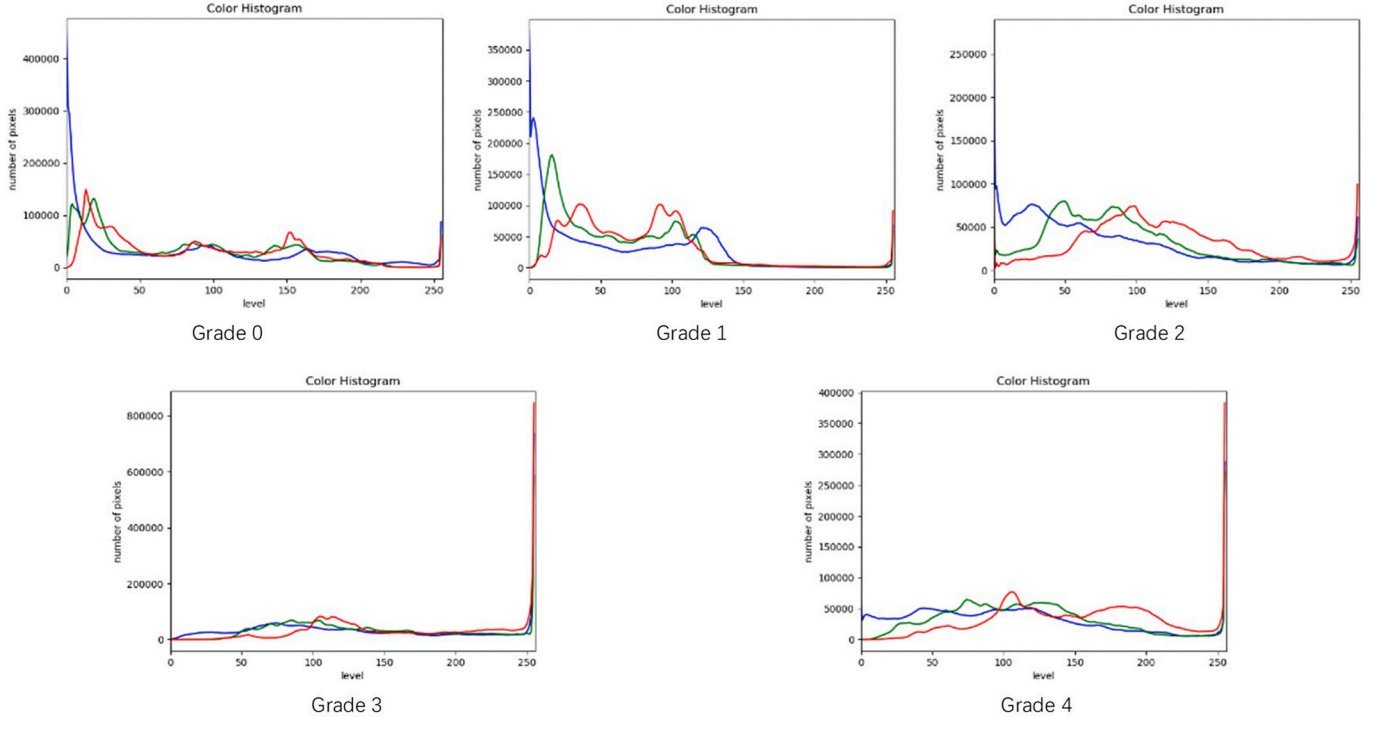

Fig. 3b. Color histogram of K–F ring images with different grades. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

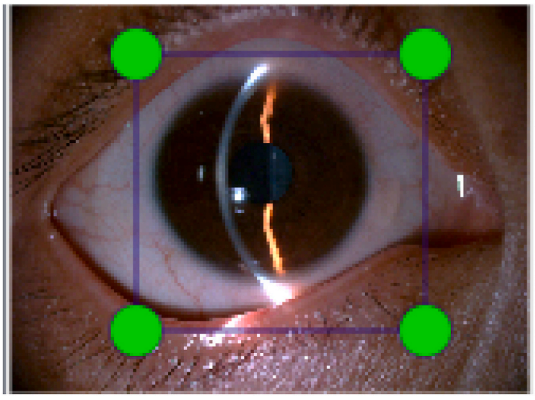

Fig. 4. Example of labeled image.

columns of numbers in the document. The first column of numbers represents the label of the target, and the following four columns of numbers represent the central coordinates of the labeled frame and the relative width and height of the labeled frame respectively.

2.2.2. Image detection

In this study, the classification of the K–F ring is determined by the deposition of copper in the cornea. Therefore, in order to achieve the classification of artificial intelligence, it is necessary to detect the cornea in the images. As mentioned above, YOLO algorithm has an ideal performance in the detection of the object. Therefore, we use the YOLOV7 to detect the cornea in this research.

YOLOv7 was implemented to detect the cornea through Python. In addition, the experimental environment was the same to the experiment that indicated in Section 3. In the experiment, the labeled images were divided into two groups in a random ratio, one is the training group, and the other is the verification group. The distribution ratio is 7:3. That is to say, there are 1295 images in the training group and 555 images in the verification group. Then, the metrics were presented tag as outputs in Figs. 5a and 5b.

Figs. 5a and 5b shows the corneal detection execution screen of the K–F ring image. Fig. 5(a) shows an example of the detection of YOLOv7, and Fig. 5(b) shows various indices of the detection. Through Fig. 5(b), we found that the accuracy of YOLOV7 in corneal detection can reach 99%).

2.2.3. Image segmentation

After labeling and detecting, the images were resized. As mentioned above, the original image cannot be used for grading directly, in which the noises in the image will decrease the accuracy of the grading. Therefore, the purpose of image resizing is to reduce noise and improve the accuracy of the grading. In the image resizing, the labeled images were resized in batches, which is based on Python. An example of the resized image is shown in Fig. 6. The resized images reduce the image noise as much as possible. It can be seen from Fig. 8 that compared with the original image, the resized image only contains or even does not contain the noise of the canthus and eyelids.

2.3. Statistical analysis

2.3.1. Correlation between age and K–F ring

The average age of patients was 20.56 ± 12.8 years, of which the youngest was 2 years old and the oldest was 67 years old. The patients were grouped as follows: 15 patients were ≤5 years old (3.76%), 88 patients were 6–15 years old (22.06%), 109 patients were 16–25 years old (27.32%), 67 patients were 25–30 years old (16.79%), 120 patients was ≥31 years old (30.08%). Table 3 shows the results of the K–F ring examination at different ages. After the chi-square test, the positive rate of K–F ring in patients aged 6–15, 16–25, 25–30, and ≥31 years is significantly higher than that in patients aged ≤5 years ($X^2 = 4.88, 23.34, 38.77, 25.46$, $p = 0.000 < 0.005$). In addition, the positive rate of K–F ring in patients aged 16–25, 25–30, and ≥31 years is significantly higher than that in patients aged 6–15 years ($X^2 = 24.66, 41.01, 28.73$, $p = 0.000 < 0.005$).

2.3.2. Correlation between gender and K–F ring

In the KFID database, 256 of 399 patients were male, accounting for 64.2%, 154 were positive, 143 were female, accounting for 35.8%, and 103 were positive. The chi-square test showed that there was no significant difference in corneal K–F ring between genders. See Table 4 for details ($X^2 = 5.64$, $p = 0.12 > 0.005$).

2.3.3. Correlation between age and K–F ring grade

Of 399 patients, 142 were negative for K–F ring and 257 were positive; Among them, 142 patients (35.6%) were of grade 0, 58 patients (14.5%) were of grade 1, 112 patients (28.1%) were of grade 2, 74 patients (18.5%) were of grade 3, and 13 patients (3.3%) were of grade 4. Table 5 shows the details of the different ages and the K–F ring grade. The correlation coefficients of K–F ring grades

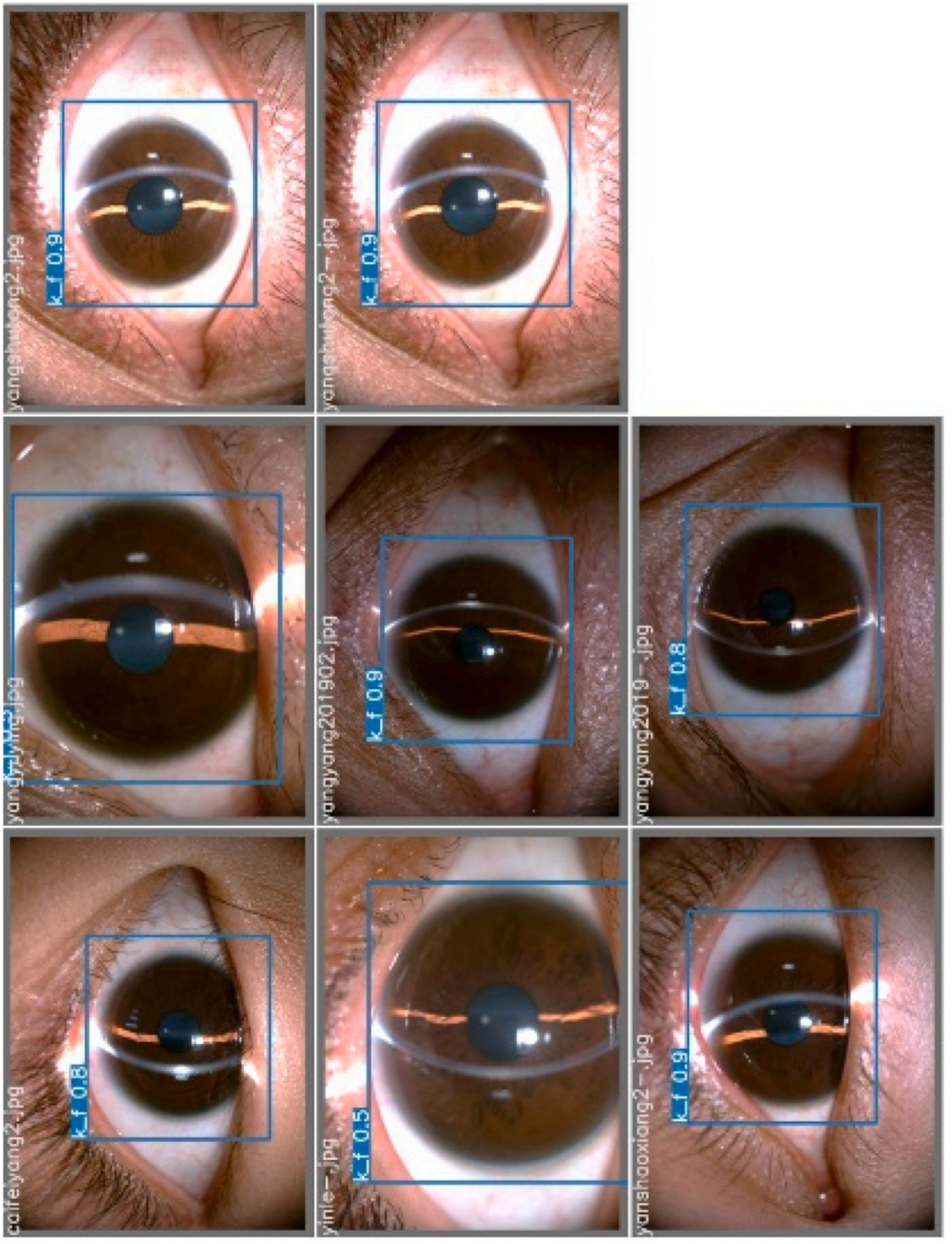

Fig. 5a. Detection of YOLOv7.

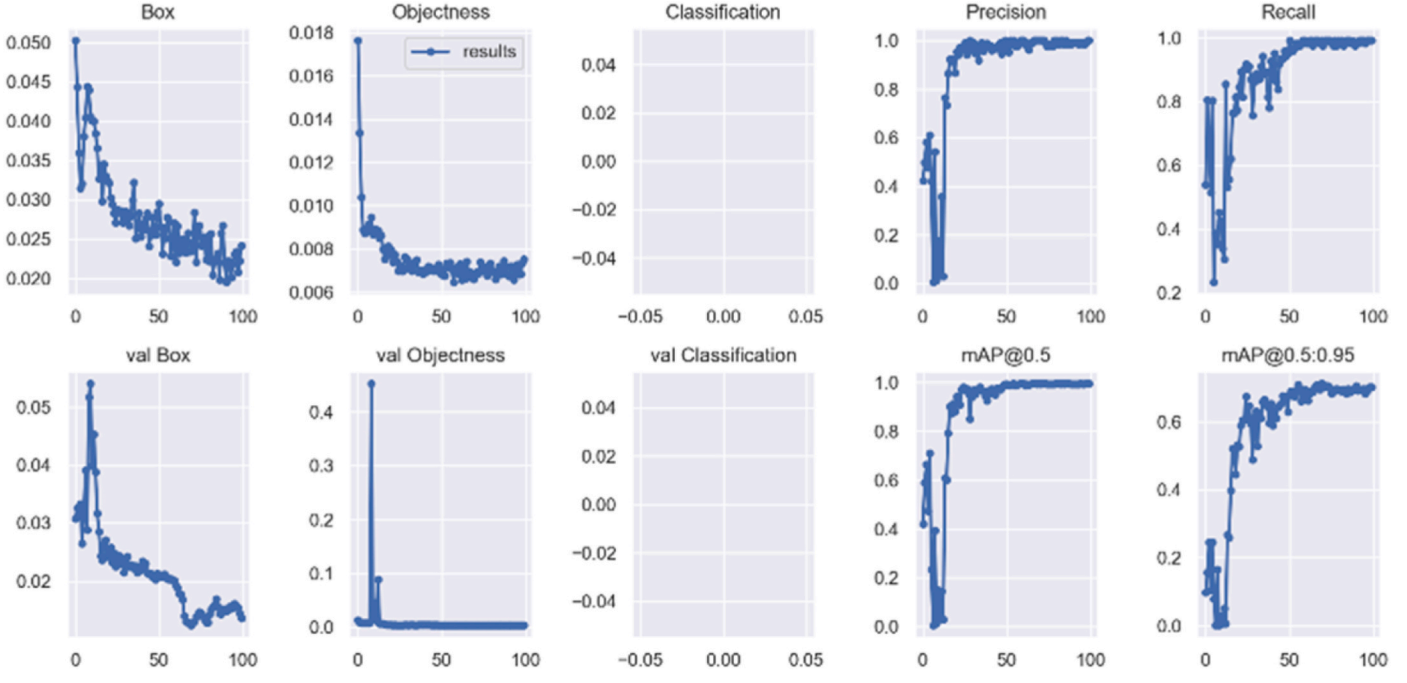

Fig. 5b. Index of the detection.

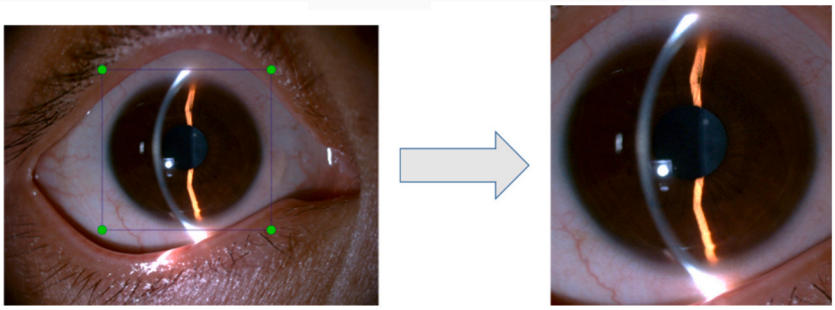

Fig. 6. The example of the resized image.

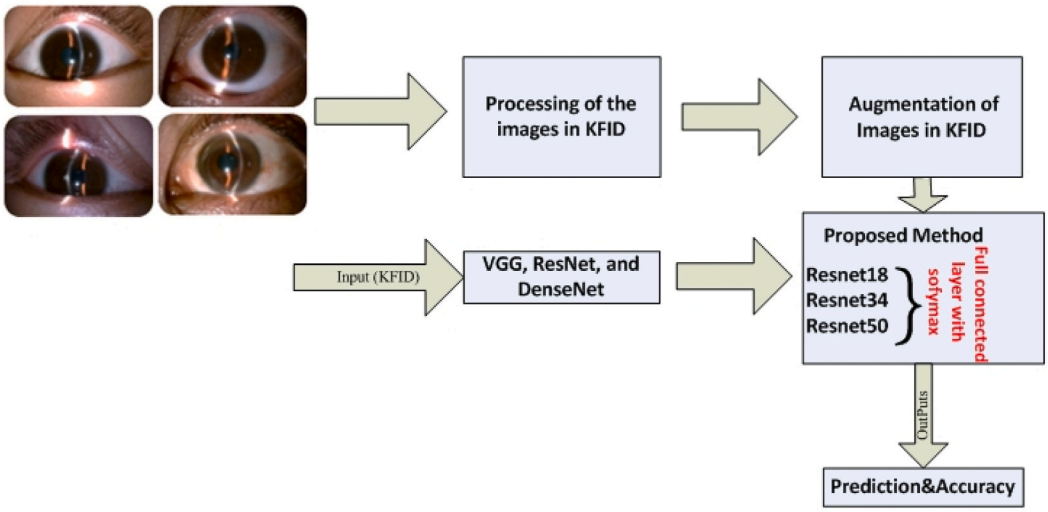

Fig. 7. The Process of the research.

and ages were $r_s = 0.232$, $P = 0.000 < 0.01$, indicating that there was a significant correlation between ages and K–F ring grades.

2.3.4. Friedman test

According to Jamaludin Siti Zulaikha Mohd [43] and Kasihmuddin Mohd Shareduwan Mohd [44], Friedman test rank has been conducted for all the data sets with $\alpha = 0.05$ and degree of freedom, $df = 4$. The P-value is <0.05. Hence, the null hypothesis of equal performance for all the logic mining models was rejected.

2.4. Data enhancement

The KFID was employed in this study, which is a highly imbalanced database. Imbalanced data are a challenging problem while training a deep learning model for a complex task, Therefore, data enhancement has been carried out in this paper.

In this paper, all images with grade 1 were inverted, half of the images with grade 3 were inverted, and all images with grade 4 were inverted and rotated. After data enhancement, the number of different grades was balanced.

3. Methods

As we know, in the process of training a convolutional neural network, when the training effect of a shallow neural network is poor, we can appropriately deepen the number of network layers to obtain a model with a better optimization effect. This is because as the depth of the network increases, the information that can be extracted by the network will be more abundant. However, in the actual experiment process, we will find that: with the deepening of the network depth, the loss of the training set will first gradually decline, and then tend to be flat; when we continue to deepen the depth of the network; the loss of the training set starts to rise instead. In other words, the network has been degraded.

In order to make the model trained by the deep network better than the shallow network, Kaiming He et al. proposed a new network structure, which is ResNet [41].

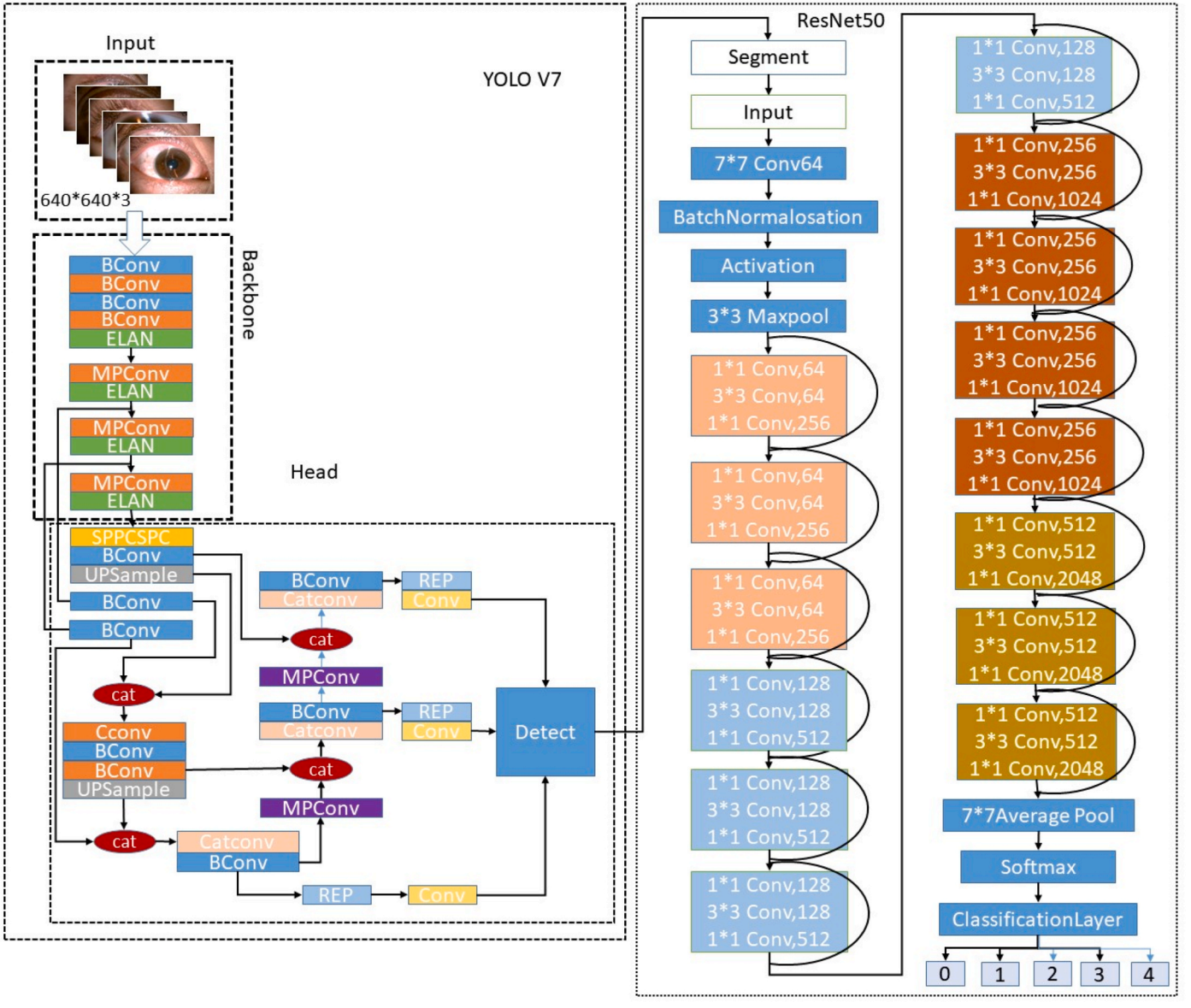

Fig. 8. The architecture of the proposed method.

Table 3
The results of K–F ring examination at different ages.

Ages	Num of Patients	Num of K–F ring positive	Positive rate (%)
≤5	15	1	6.67
6–15	88	31*	35.23
16–25	109	80*★	73.39
25–30	67	58*★	86.57
≥31	120	87*★	72.50

Table 4
Different gender of the K–F ring.

Gender	Sum	Positive	Ratio of Gender	Ratio of Positive
Male	256	154	64.2%	59.9%
Female	143	103	35.8%	40.1%

Table 5
The different ages and grades of K–F ring.

Ages	Grades				
	0	1	2	3	4
≤5	14	0	1	0	0
6–15	57	3	16	9	3
16–25	29	14	34	27	5
25–30	9	17	24	15	2
≥31	33	24	37	23	3

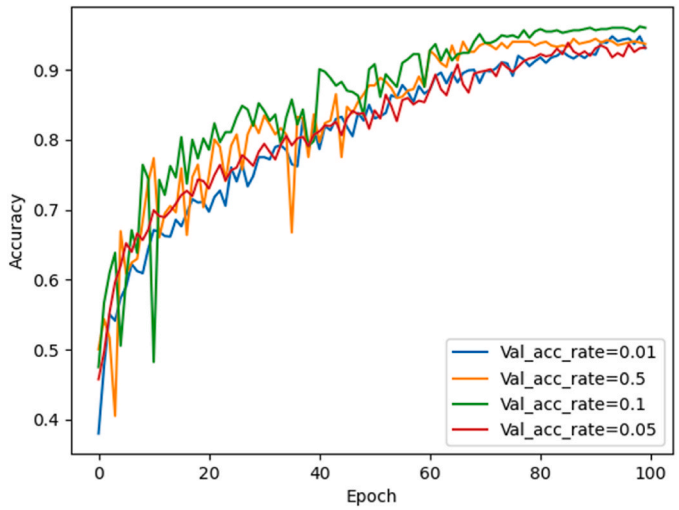

Fig. 9a. The accuracy of the different learning rate.

In this research, the process is as follows: First, labeled the images by professional doctors; then this paper completed the detection and resizing of the labeled images; finally, automatic grading of the processed images was implemented by using VGG (Visual Geometry Group), ResNet, and DenseNet (Dense Convolutional Network). Fig. 7 indicates the process.

The architecture of our proposed method is shown in Fig. 8. The images as the input of the fine-tuned CNNs(Convolutional Neural Networks), and the output is the prediction and accuracy.

4. Experiment and results

After completing the construction of KFID, how to better apply it is our key research task. As mentioned earlier, it is of great significance to improve the diagnosis of low-age patients with hepatolenticular degeneration. Therefore, machine learning or deep learning methods will help clinicians find the symptoms of HLD patients earlier in the eyes. To this end, the following work was done in

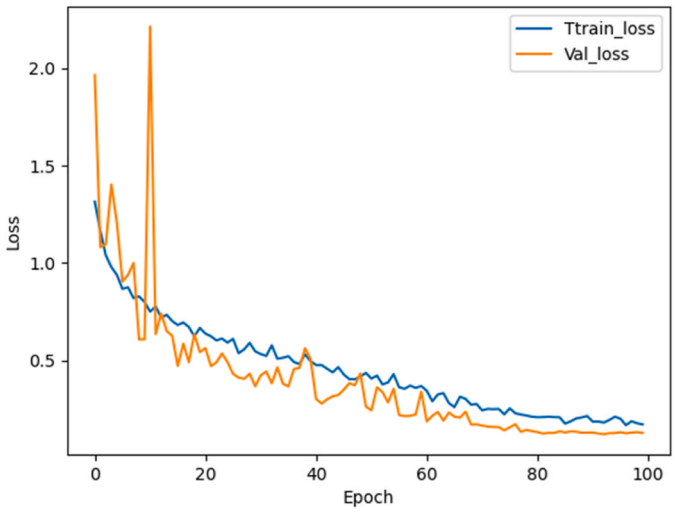

Fig. 9b. The loss of proposed method.

this section:

4.1. Experiment setup

In this study, VGG, ResNet, and DenseNet were used for grading the images in the KFID. Some fine-tuned CNN architectures were developed to classify K–F ring images. These architectures are VGG-16, VGG-19, ResNet18, 34, 50, and DenseNet. Each CNN architecture uses fully connected (FC) layers with softmax layers. This paper transfer the input of the last full connection layer of CNN to a linear layer with 256 output units and then connect the ReLU layer and Dropout layer. This means the fine-tuned is that 256×6 linear layers are used and 5-channel softmax layers are used for output.

The parameters were been stetted though these works. The settings of different learning rate and epoch would affect the accuracy. The experiments with the learning rate of 0.01, 0.05, 0.1, and 0.5 were completed. Fig. 9(a) shows the accuracy of different learning rate, in which the learning rate with 0.1 have the best accuracy. And Fig. 9(b) shows that the loss function was convergent near 100 epochs.

Furthermore, the experiment were implemented under the following environment.

- OS: Windows 10 (64-bit)
  •CPU: Intel i5-6500
  •Memory: 32 GB
  •Storage: 512G SSD
  •Graphics card:GTX1050Ti (the video memory: 4G)
  •Experimental distribution: anaconda3
  •Python version 3.9.13
  •VGG-16, VGG-19, ResNet18, 34, 50, and DenseNet

  The parameter settings were:

  •batch_size = 32
  •num_classes = 5
  •learning_rate = 0.1
  •num_epochs = 100

This is how the data sets. The processed images were been allocated according to the ratio of 0.7 and 0.3 randomly. That is, there are 1923 images in the training set and 826 images in the valid set. The number of each grade is 436, 368, 352, 389, and 378, which correspond to grade 0, grade 1, grade 2, grade 3, and grade 4 in the training set. In addition, the number of each grade in a valid set is 187, 160, 151, 166, and 162.

4.2. Metrics

A confusion matrix is used to gauge the performance of the classification model formally. Table 6 presents a confusion matrix that

Table 6
Confusion matrix.

		Predicted grade	
		Positive	Negative
Actual grade	Positive	TP[1]	FN[3]
	Negative	FP[2]	TN[4]

[1] True positive (TP) = Number of correctly graded positive samples.
[2] False Positive (FP) = Number of incorrectly graded negative samples.
[3] False Negative (FN) = Number of incorrectly graded positive samples.
[4] True Negative (TN) = Number of correctly graded negative samples.

Table 7
The metrics of different models.

Model	Precision	Recall	Accuracy	Specificity	F1-Score	Time/epcoh	Params
VGG-16	88.97 ± 3.22%	90.05 ± 1.58%	89.88 ± 3.05%	91.25 ± 2.69%	89.91 ± 3.06%	409.65s	138.65 M
VGG-19	92.35 ± 1.55%	91.93 ± 1.99%	91.89 ± 2.01%	92.33 ± 2.31%	91.55 ± 1.94%	512.99s	145.63 M
DenseNet	95.66 ± 1.83%	95.02 ± 1.33%	94.58 ± 1.15%	95.78 ± 1.89%	95.03 ± 1.33%	211.52	28.68 M
ResNet18	93.65 ± 2.06%	94.05 ± 1.96%	94.18 ± 1.05%	95.25 ± 2.11%	93.55 ± 2.16%	203.49s	12.63 M
ResNet34	95.27 ± 0.29%	95.23 ± 0.3%	95.31 ± 0.29%	96.99 ± 1.78%	95.23 ± 0.29%	210.74s	21.83 M
ResNet50	94.05 ± 1.12%	93.98 ± 1.02%	93.59 ± 0.81%	95.21 ± 2.03%	94.02 ± 1.63%	248.16s	25.56 M

depicts the grading result into four categories, namely TP, FP, TN, and FN. Other evaluation metrics are formed based on these four categories. Such as precision, recall, accuracy, specificity, and F1-score [45,46].

4.2.1. Precision

The ratio of the total number of positive correctly graded samples to the total count of all positive graded samples is called precision, as shown in Eq. (1).

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

4.2.2. Recall (sensitivity)

The percentage of the total number of positive correctly graded samples to the total count of all positive graded samples is called recall, as shown in Eq. (2).

$$Recall(Sensitivity) = \frac{TP}{TP + FN} \tag{2}$$

4.2.3. Accuracy

The ratio of the total number of positive correctly graded samples to all the samples in the dataset is called accuracy, as shown in Eq. (3).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

4.2.4. Specificity

The ratio of the total number of positive correctly graded negative samples to the total count of all negative graded samples is called specificity, as shown in Eq. (4).

$$Specificity = \frac{TN}{FP + TN} \tag{4}$$

4.2.5. F1-score

F1-score is defined as the harmonic mean of precision and recall, as shown in Eq. (5).

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{5}$$

4.3. Results and analysis

A total of 10 experiments were conducted in this paper with all models mentioned above respectively. In each time of the dataset was redistributed in 7:3. The metrics were recorded in each experiment, at the same time, the execution time and the time complexities

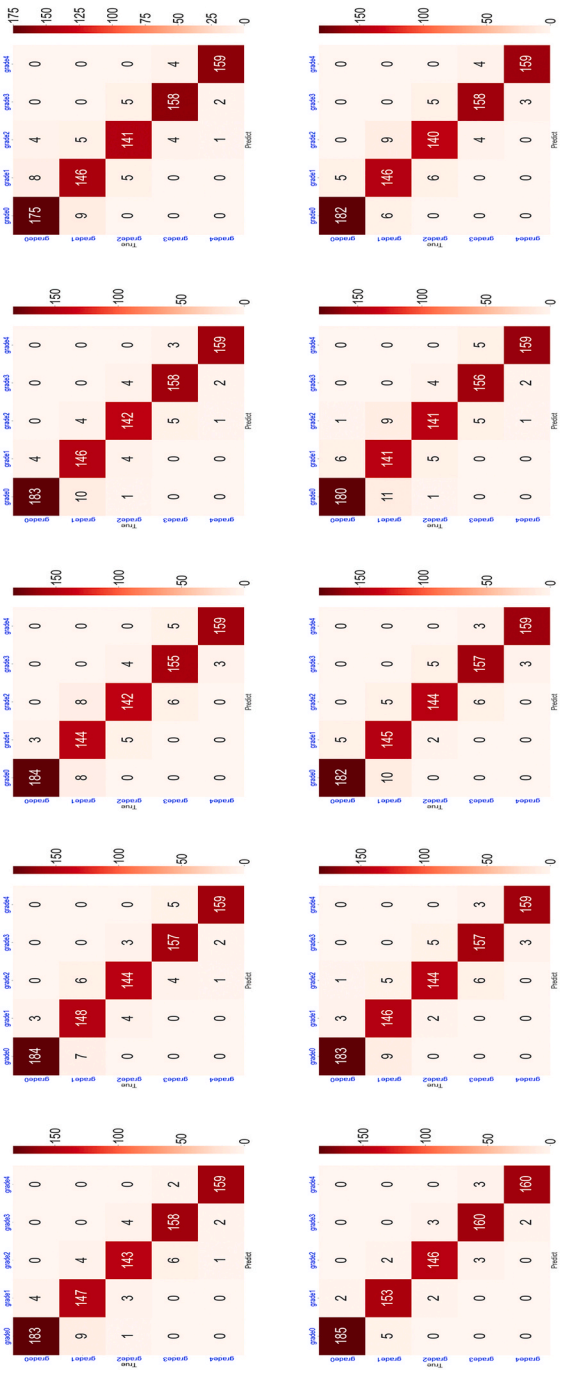

Fig. 10. The confusion matrices of ResNet34.

were counted. Table 7 shows the metrics of each model in this paper.

Table 7 shows the results of the models in this paper. The models have an excellent outcome of grading. Compared to the previous experiment, the processing measures in this paper increases the accuracy. The best accuracy of the models reaches by over 95%, indicating that this paper's work plays an important role in model performance. The execution time of these models were similar except the VGG. The metric of the time shows that the ResNet34 runs faster than others. However, the size of params of these models were vary widely. The size of VGG was much larger than other models. In addition, Table 7 gives the statistical information of the metrics, which indicate that the models are stable in grading.

In order to analyze the causes of grading errors, the confusion matrixes were carried out in each epoch. Taking ResNet34 as an example, the confusion matrixes of 10 experiments were shown in Fig. 10. Through the confusion matrix, it can be observed that the main errors in the experiment were between grade 0 and 1.

5. Conclusion & discussion

In this study, the KFID was constructed which contains 1850 images with 399 different WD patients. Meanwhile, some work elaborated on the statistical significance of KFID. To increase the accuracy of automatic grading, the YOLOv7 was used to detect the corneal and the different models were implemented to grade the images, which reach ideal results. The results indicate that ResNet34 has the highest accuracy of 95.31%.

Although the accuracy has achieved ideal results, this paper still has some work to improve in the later period. The later works mainly include as follows: first, the accuracy of grade 0, 1, and 2 should be improve; second, more detail work should be done to realize the measurable grading. Third, multimodal data should be collected to support early diagnosis, and the explained artificial intelligence can be proposed.

The limitation of this paper is how to apply the model in real-time work. Due to the particularity of medical devices, some interfaces are not allowed to be accessed if they are not licensed. That will be a fetter of the application in the grading.

Author contribution statement

Wei Song: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Ling Xin: Conceived and designed the experiments; Analyzed and interpreted the data.

Jiemei Wang: Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.

Data availability statement

Data will be made available on request.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

References

[1] Juan Zhang, et al., Differentially expressed lncRNAs in liver tissues of TX mice with hepatolenticular degeneration, Sci. Rep. 11 (1) (2021) 1377.
[2] Hussein Mohammed Salah, et al., Epidemiology, evaluation and management of Wilson disease: review article, J. Pharm. Res. Int. (2021) 94–103.
[3] K. Petrukhin, et al., Mapping, cloning and genetic characterization of the region containing the Wilson disease gene, Nat. Genet. 5 (4) (1993) 338–343.
[4] P.C. Bull, et al., The Wilson disease gene is a putative copper transporting P-type ATPase similar to the Menkes gene, Nat. Genet. 5 (4) (1993) 327–337.
[5] Corrigendum: the Wilson disease gene is a putative copper transporting P–type ATPase similar to the Menkes gene, Nat. Genet. 6 (2) (1994) 214.
[6] Prat Davies Lisa, et al., New mutations in the Wilson disease gene, ATP7B: implications for molecular testing, Genet. Test. 12 (1) (2008) 139–145.
[7] Shijie Zhang, et al., Clinical and genetic characterization of a large cohort of patients with Wilson's disease in China, Transl. Neurodegener. 11 (1) (2022) 13.
[8] Abuduxikuer Kuerbanjiang, et al., Wilson disease with hepatic presentation in an eight-month-old boy, World J. Gastroenterol. 21 (29) (2015) 8981–8984.
[9] Kim Joo Whee, et al., Genetically confirmed Wilson disease in a 9-month old boy with elevations of aminotransferases, World J. Hepatol. 5 (3) (2013) 156–159.
[10] J. Suvarna, Kayser-Fleischer ring, J. Postgrad. Med. 54 (3) (2008) 238.
[11] Degirmenci Cumali, Palamar Melis, Evaluation and grading of Kayser-Fleischer ring in Wilson disease by Scheimpflug camera, Eur. J. Ophthalmol. 31 (4) (2020), 1120672120931025.
[12] N. Arora, et al., Kayser-Fleischer ring, QJM: Mon. J. Assoc. Phys. 113 (5) (2020) 361.
[13] G. Joshi, et al., Kayser-Fleischer ring in Wilson's disease, QJM: Mon. J. Assoc. Phys. 112 (8) (2019) 629.
[14] A. Ala, et al., Wilson's disease, Lancet 369 (9559) (2007) 397–408.
[15] J. Seniomicronw, et al., Cognitive functioning in neurologically symptomatic and asymptomatic forms of Wilson's disease, Mov. Disord. 17 (5) (2002) 1077–1083.
[16] Alharbi Fahad, et al., A brief review of acoustic and vibration signal-based fault detection for belt conveyor idlers using machine learning models, Sensors 23 (4) (2023) 1902.
[17] A. Sezer, et al., Optimization of Deep Learning Model Parameters in Classification of Solder Paste Defects, 2021.
[18] Sezer Ali, Aytaç Altan, Detection of solder paste defects with an optimization-based deep learning model using image processing techniques, Solder. Surf. Mt. Technol. 33 (5) (2021) 291–298.
[19] K.S. Dar, et al., Cyber threat detection using machine learning techniques: a performance evaluation perspective, in: IEEE International Conference on Cyber Warfare and Security, IEEE, 2020.

[20] Shaukat Kamran, Suhuai Luo, Varadharajan Vijay, A novel method for improving the robustness of deep learning-based malware detectors against adversarial attacks, Eng. Appl. Artif. Intell. (2022) 116.
[21] Kamran Shaukat, et al., Performance comparison and current challenges of using machine learning techniques in cybersecurity, Energies 13 (10) (2020).
[22] K.S. Dar, S. Luo, V. Varadharajan, et al., A survey on machine learning techniques for cyber security in the last decade, IEEE Access (2020) 8.
[23] A.T. Mahboob, S. Kamran, K. Adel, et al., A Fuzzy inference-based decision support system for disease diagnosis, Comput. J. 1 (2022) bxac068.
[24] Arunangshu Pal, Vinay Kumar, AgriDet: plant Leaf Disease severity classification using agriculture detection framework, Eng. Appl. Artif. Intell. (2023) 119.
[25] Zain Ali, et al., A proposed framework for early prediction of Schistosomiasis, Diagnostics 12 (12) (2022) 3138.
[26] Talha Mahboob Alam, et al., An efficient deep learning-based skin cancer classifier for an imbalanced dataset, Diagnostics 12 (9) (2022) 2115.
[27] Ziyi Li, et al., A classification method for multi-class skin damage images combining quantum computing and Inception-ResNet-V1, Front. Phys. (2022) 10.
[28] Devnath Liton, et al., Deep ensemble learning for the automatic detection of pneumoconiosis in coal worker's chest X-ray radiography, J. Clin. Med. 11 (18) (2022) 5342.
[29] Chetana Srinivas, et al., Deep transfer learning approaches in performance analysis of brain tumor classification using MRI images, J. Healthc. Eng. 2022 (2022) 3264367.
[30] Kumar M. Rupesh, et al., Dementia detection from speech using machine learning and deep learning architectures, Sensors 22 (23) (2022) 9311.
[31] Yi Wang, et al., RU-Net: an improved U-Net placenta segmentation network based on ResNet, Comput. Methods Progr. Biomed. 227 (2022) 107206.
[32] Sahli Hanene, et al., ResNet-SVM: fusion based glioblastoma tumor segmentation and classification, J. X Ray Sci. Technol. 31 (1) (2022) 27–48.
[33] Saleh Alghamdi Hanan, Towards explainable deep neural networks for the automatic detection of diabetic retinopathy, Appl. Sci. 12 (19) (2022) 9435.
[34] Gongqiang Wang, et al., Analysis on related Factors of Kayser-Fleischer ring in patients with Wilson's disease, J. Med. Res. 42 (7) (2013) 98–101.
[35] Lin Jiang, et al., Application of a fast RCNN based on upper and lower layers in face recognition, Comput. Intell. Neurosci. 2021 (2021) 9945934.
[36] Yuanzhang Zhao, Shengling Geng, Face occlusion detection algorithm based on yolov5, J. Phys. Conf. 2031 (1) (2021).
[37] Aralikatti Anish, et al., Real-time object detection and face recognition system to assist the visually impaired, J. Phys. Conf. 1706 (1) (2020), 012149.
[38] HoWon Lee, et al., Using hybrid algorithms of human detection technique for detecting indoor disaster victims, Computation 10 (11) (2022) 197.
[39] Kohei Arai, et al., Pedestrian safety with eye-contact between autonomous car and pedestrian, Int. J. Adv. Comput. Sci. Appl. 10 (5) (2019).
[40] Yu Fan, et al., Research on face recognition technology based on improved YOLO deep convolution neural network, J. Phys. Conf. (1) (2021) 1982.
[41] Kaiming He, et al., Deep Residual Learning for Image Recognition, CoRR, 2015 abs/1512.03385.
[42] Hussain Muhammad, et al., Domain feature mapping with YOLOv7 for automated edge-based pallet racking inspections, Sensors 22 (18) (2022) 6927.
[43] Mohd Jamaludin Siti Zulaikha, et al., Novel logic mining incorporating log linear approach, J. King Saud Univ. - Comput. Inf. Sci. 34 (10PB) (2022) 9011–9027.
[44] Mohd Kasihmuddin Mohd Shareduwan, et al., Supervised learning perspective in logic mining, Mathematics 10 (6) (2022) 915.
[45] Yueling Guo, et al., YRAN2SAT: a novel flexible random satisfiability logical rule in discrete hopfield neural network, Adv. Eng. Soft. (2022) 171.
[46] Zamri Nur Ezlin, et al., Multi-discrete genetic algorithm in hopfield neural network with weighted random k satisfiability, Neural Comput. Appl. 34 (21) (2022) 19283–19311.