Skin lesion image classification method based on extension theory and deep learning

Xiaofei Bian[1] · Haiwei Pan[1] · Kejia Zhang[1] · Pengyuan Li[2] · Jinbao Li[3] · Chunling Chen[1]

Abstract

A skin lesion is a part of the skin that has abnormal growth on body parts. Early detection of the lesion is necessary, especially malignant melanoma, which is the deadliest form of skin cancer. It can be more readily treated successfully if detected and classified accurately in its early stages. At present, most of the existing skin lesion image classification methods only use deep learning. However, medical domain features are not well integrated into deep learning methods. In this paper, for skin diseases in Asians, a two-phase classification method for skin lesion images is proposed to solve the above problems. First, a classification framework integrated with medical domain knowledge, deep learning, and a refined strategy is proposed. Then, a skin-dependent feature is introduced to efficiently distinguish malignant melanoma. An extension theory-based method is presented to detect the existence of this feature. Finally, a classification method based on deep learning (YoDyCK: YOLOv3 optimized by Dynamic Convolution Kernel) is proposed to classify them into three classes: pigmented nevi, nail matrix nevi and malignant melanomas. We conducted a variety of experiments to evaluate the performance of the proposed method in skin lesion images. Compared with three state-of-the-art methods, our method significantly improves the classification accuracy of skin diseases.

✉ Haiwei Pan
    panhaiwei@hrbeu.edu.cn; panhaiwei2006@hotmail.com

[1]   Computer Science and Technology, Harbin Engineering University, NO.145 Nantong Street, Harbin 150001 Heilongjiang, China

[2]   Department of Computer and Information Sciences, University of Delaware, Newark, USA

[3]   Shandong Artificial Intelligence Institute, Qilu University of Technology (Shandong Academy of Science), Jinan, China

1 Introduction

Malignant melanoma is one of the most dangerous skin cancers [39]. The signs and symptoms of malignant melanoma are hard to identify and easily misdiagnosed in the early stage based on visual observations from doctors [8, 12, 23, 44]. As malignant melanoma is untreatable at an advanced stage after the spread of the disease, the precise detection of malignant melanoma in its early stages is thus important for saving lives [27, 40]. For dermatologists, classifying skin lesion images is a very heavy job for doctors, which will take considerable manpower and time, and it will lead to an increase in the misdiagnosis rate [14, 21, 22, 28, 32]. This motivates us to automatically classify skin lesion images quickly and accurately. Our research can assist doctors in improving the effect of diagnosis. Several questions need to be addressed. First, while many machine learning methods have achieved good results, most existing methods only use deep learning features to classify images. Few researchers have integrated medical domain features with deep learning features well [17]. To improve the classification accuracy, a classification framework integrated with medical domain knowledge, deep learning, and a refined strategy is proposed, and a two-phase classification method for skin lesions is proposed in this paper. To the best of our knowledge, our work represents the first to combine medical domain features (extracted by using Extenics) with deep learning features (extracted by using YoDyCK) for classification.

Second, although many color and texture features have been proposed and tested, these traditional methods still have some limitations, that is, the detection of Blue White Veil (BWV) in melanoma dermoscopic images has not been well studied or addressed. BWV is a critical feature that is summarized based on expert experience for the diagnosis of malignant melanoma [4, 6, 15]. Detecting this skin-dependent feature in a rapid and accurate way is essential for the preliminary classification of malignant melanoma. Currently, most automatic skin lesion classification methods use deep learning methods [5]. However, deep learning methods not only take a long time to train but also require considerable data. In this paper, the extension-dependent function in extension theory is used to rapidly and accurately detect the existence of BWVs. The extension-dependent function can better deal with the complexity and uncertainty of medical images.

Third, how to classify suspected malignant skin lesion images rapidly and accurately. For skin lesion images, the lesion area is mainly located in the middle of the image, and BWV is not an edge feature. Therefore, a YoDyCK (YOLOv3 optimized by Dynamic Convolution Kernel) model is proposed in this paper. The model can quickly eliminate the interference of the image background and highlight the features in the middle of the image, and it is used to extract the useful features from suspected malignant skin lesion images. This model uses a dynamic convolution kernel to optimize the convolution process of YOLOv3, and the size of the convolution kernel is gradually decreasing.

Overall, the main contributions of our work are:

- In this paper, a classification framework integrated with medical domain knowledge, deep learning, and a refined strategy is proposed. Medical domain features and deep learning features are combined in the framework. A two-phase classification method is proposed to rapidly and accurately classify skin lesion images.

- BWV is a critical feature for the diagnosis of malignant melanoma. In this paper, the extension-dependent function in extension theory is introduced to detect the existence of BWVs. The skin lesion images are classified preliminarily into benign skin lesion images and suspected malignant skin lesion images.
- In this paper, to improve the classification accuracy of suspected malignant melanoma, a YoDyCK model is proposed to extract the deep learning feature rapidly and accurately on suspected malignant skin lesion images.

The rest of the paper is organized as follows: Section 2 discusses related works about skin lesion classification methods, deep learning, and extension theory. Section 3 provides an overview of the proposed two-phase classification method and discusses the details of the model. Section 4 presents the performance evaluation of the method for skin lesion image classification. Section 5 concludes the paper and discusses possible future works.

2 Related Works

Skin lesion classification method For malignant melanoma detection and classification. A variety of diagnostic criteria have been introduced by doctors. Among them, the ABCD principle and seven-point examination, which are mostly based on visual observations from skin images, are widely used, including the color, texture, and shape information obtained from images [2, 9, 16, 43]. Some patterns, such as atypical pigment networks, blue–white veils (BWVs), and irregular stripes, have been identified as effective for melanoma detection [1, 24, 37]. Zarit et al. presented a comparative evaluation of the pixel classification performance of two skin detection methods in five color spaces. Skin detection methods are color-histogram-based approaches that are intended to work with a wide variety of individuals, lighting conditions, and skin tones [55]. Setiawan A W et al. used the combination of R, G, and B pixel matrixes under different weights and proved that the coupling results of pixel matrixes under different channels can enhance the discrimination of color features [41]. Chandy et al. proposed a gray level statistical matrix from which four statistical texture features were estimated for the retrieval of mammograms from the mammographic image analysis society (MIAS) database [7]. Choi, YH. et al. proposed a new scheme for a self-diagnostic application that can estimate the actual age of the skin on the basis of the features on a skin image. In accordance with dermatologists' suggestions, the length, width, depth, and other cell features of skin wrinkles are examined to evaluate skin age [13].

Against the limitations of traditional methods of automatic detection and classification accuracy, machine learning technologies have attracted extensive attention in automatic malignant melanoma detection and classification. F. Warsi et al. proposed a method to extract color-texture features and detect malignant melanoma by using a back-propagation multilayer neural network classifier [47]. Mhaske et al. used a two-dimensional wavelet to extract features for the classification of skin cancer using neural networks and support vector machines (SVMs). It has been proven that color features can effectively diagnose skin cancer [30]. Melbin, K. et al. introduced an integrated method for detecting skin lesions from dermoscopic images. The integrated cumulative level difference mean (CLDM)-based modified ABCD features and support vector machine (SVM) are proposed to detect and classify skin lesion images [29]. Roslin, S. et al. introduced a comparative study for classifying melanoma using supervised machine learning algorithms. Classification of melanoma from dermoscopic data is

proposed to help the clinical utilization of dermatoscopy imaging methods for skin sore classification. More fusion methods have been proposed to improve classification accuracy [38]. Zhang-China et al. proposed a two-stage multiple classifier system to improve classification reliability by introducing a rejection option [57]. Yun, Y. et al. proposed a novel method for multiview object pose classification through sequential learning and sensor fusion. The basic idea is using images observed in visual and infrared bands, with the same sampling weight under a multiclass boosting framework [54]. Nver H M et al. proposed a novel and effective pipeline for skin lesion classification and segmentation in dermoscopic images combining a deep convolutional neural network named the third version of You Only Look Once (YOLOv3) and the GrabCut algorithm [31].

Deep learning With advances in deep learning, particularly convolutional neural networks (CNNs), in computer vision applications, the accuracy of classification has reached an impressive improvement. Since 2012, several CNN algorithms and architectures have been proposed, such as Faster-RCNN, YOLO, YOLOv2, and YOLOv3. They combined the region proposal algorithm with CNN [34–36]. In the literature [42], the role of the number of dimensions and colors of skin cancer images in the development of CNN models is comprehensively discussed. Machine learning with limited resources can be achieved by minimizing the size and number of colors to reduce the amount of computation without losing the key information of the image. YOLOv3 is a state-of-the-art version of the You Only Look Once (YOLO) model, which unifies target classification and localization into a regression problem [45]. As YOLOv3 uses multiscale predictions to detect the final targets, it can more effectively detect small targets than previous versions of YOLO. The technology has been widely used in the fields of image detection, video detection, and real-time camera detection, and it has a good effect. An efficient and accurate vehicle detection algorithm in aerial infrared images is proposed [56] via an improved YOLOv3 network, which increases the detection efficiency and the classification accuracy by using the improved YOLOv3. Pang, S et al. built the first medical image dataset of cholelithiasis by collecting 223,846 CT images with gallstones from 1369 patients. With these CT images, a neural network is trained to "pick up" CT images of high quality as the training set, and then a novel YOLO neural network, named the YOLOv3-arch neural network, is proposed to identify cholelithiasis and classify gallstones on CT images [33]. As far as we know, it has a good effect on natural images but is rarely used in the classification of skin lesions.

Extension theory Extenics has been used since 1983 by Cai Wen and has been successfully used in various applications. Extension set theory [48–52] is a mathematical formalism for representing uncertainty that can be considered an extension of classical set theory. It has been used in many different research areas [25]. Different from Euclidean distance, Cai Wen believes that the value of the distance between a point and an interval is not 0. Therefore, Cai Wen proposed the extension distance formula, which is the distance from the point to the midpoint of both ends of the interval minus the interval radius. Cai Wen further proposed the point-to-two interval place value formula based on the extension distance and constructed a certainty elementary dependent function formula with two nested intervals. The elementary dependent function formula constructed above can express the properties of things in a multidimensional and quantitative way and can also be used to calculate the value of the dependent function when the point falls into any position under two nested intervals [18, 19, 58]. The above mentioned is only the case of a point-to-two interval. However, it is difficult to

collect complete and accurate data in practice, and there are usually some errors. Therefore, the primitive value is usually a range value, not an exact value. In this paper, to more accurately reflect the complexity of medical image features, an uncertainty extension-dependent function with primitive values and interval values is proposed to detect the existence of BWV. Then, the skin lesion images are preliminarily classified into benign skin lesion images and suspected malignant skin lesion images.

3 Methods

3.1 Overview

A classification framework integrated with medical domain knowledge, deep learning, and a refined strategy is proposed in this paper for automatically classifying skin lesions. As described in Fig. 1, given an image that is converted to three channels R, G, and B, B-G is used to enhance the features of BWV. Then, these features are fed into average pooling to reduce the noise and dimensionality and summarize the features contained in the input subregions. As of now, the preprocessing process is completed.

Next, starting from the extension theory-based method, a two-phase classification process classifies the skin lesion images. Our model uses the extension theory-based method to classify the skin lesion images in the first phase, and the input images are divided into benign skin disease images and suspected malignant skin disease images. Then, the second phase classification is started. Finally, the benign skin disease images are input into YOLOv3, and the suspected malignant skin disease images are input into YoDyCK. The final result is obtained.

3.2 BWV augment method

The color of BWV (a critical feature for distinguishing benign and malignant melanoma) is characterized by large changes and a complex color distribution. The difficulty of extracting the BWV is that it is difficult to accurately describe using only the three color components of R, G, and B, and the description requires a large amount of computation. To retain as much effective information as possible in malignant melanoma images, the color model of BWV is established in this paper based on the difference of R, G, and B channel components.

The M × N melanoma image in RGB color space is denoted as I, and the channel component number of per pixel is 3. Therefore, the melanoma image has M × N pixels.
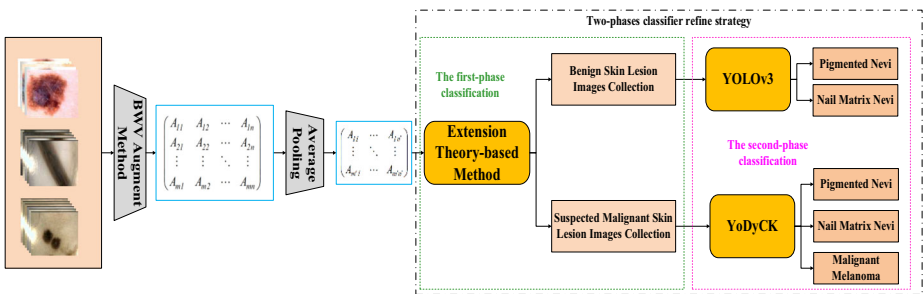

Fig. 1 Illustration of the proposed model. YoDyCK denotes YOLOv3 optimized by a dynamic convolution kernel

First, the three color components of the lesion area are extracted and denoted by R, G, and B. The extraction formula is as follows:

$$R = I(:,:,1)$$

$$G = I(:,:,2)$$

$$B = I(:,:,3)$$

where R is the red component, G is the green component, B is the blue component, and the size is M × N. Then, according to the color features of the blue–white veil, the blue region features of the malignant melanoma images are extracted and denoted as B_g. The calculation formula is as follows:

$$B_g = B-G \tag{1}$$

To obtain more accurate extraction results of BWV features, this paper uses the B_g component instead of R, G, and B to describe malignant melanoma images, and the image pixel matrix of B_g is $I_{M \times N}$. The visualization of the result is shown in Fig. 2. Obviously, the processed image reveals the features of BWV more clearly.

3.3 Average pooling

As the BWV is a structureless zone with an overlying white "ground-glass" haze, much noise exists among the BWV regions. Average pooling is typically used to downsample the input image to reduce the noise and dimensionality and to summarize the features contained in subregions of the input. Thus, average pooling is applied to B_g images to reduce the amount of calculation and retain more features of the original image. The B_g image is divided according to the fixed size grid, and the pixel value in the grid is the average value of all the pixels in the grid. The average value v of all the pixels in the grid is computed as:

$$v = \frac{1}{|n|} * \sum_{w=1}^{n} v_w \tag{2}$$

where n is the total number of pixels in each grid and v_w is the pixel in the grid.

(a)   (b)   (c)
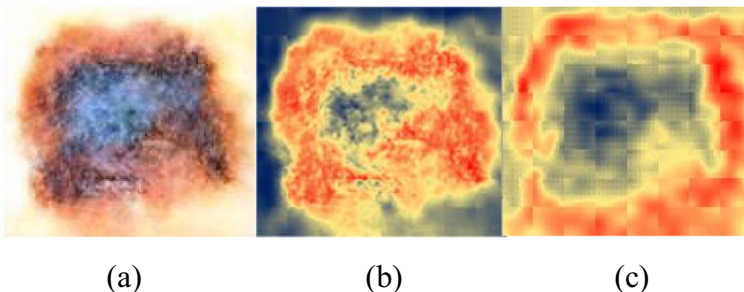
Fig. 2 Visualization of the results. a Raw malignant melanoma image. b Heatmap of the extracted blue channel of the raw image. c Heatmap of the processed B_g-image

In this paper, the size of the convolution kernel is set as 3*3, and the step size is 1. After the average pooling step, an image is obtained, and it is denoted as $B_{g\text{-avg}}$. The pixel matrix after average pooling is denoted as $I'_{i \times j}$.

3.4 Extension theory-based classification method

Extension theory is a new theory that aims to describe the change of the nature of matters, thus taking both qualitative and quantitative aspects into account. Extension theory has introduced the notion of extension distance. New concepts of "distance" and "side distance", which describe distance, are established to break the classical mathematics rule that the distance between points and intervals is zero if the point is within the interval. The dependent function established on the basis of this can quantitatively describe the objective reality of "differentiation among the same classification" and further describe the process of qualitative change and quantitative change [48].

Suppose x is any point in the real axis, and $U' = \langle a_1, a_2 \rangle$ is an interval in the real field, then

$$\rho\left(x, U'\right) = \left| x - \frac{a_1 + a_2}{2} \right| - \frac{a_2 - a_1}{2} \tag{3}$$

is the extension distance between point x and interval $\langle a_1, a_2 \rangle$, where $\langle a_1, a_2 \rangle$ can be an open interval, a closed interval, or a half-open and a half-closed interval U'. In effect, this is the distance between the point considered and the closest border of the interval. In this case, the values of x and v are the same. When the point is on the border of the interval (i.e. $x = a_1$ or $x = a_2$), the extension distance will be null, while the minimum possible value for the extension distance is the negative of half of the interval length.

With the aid of this new take on the distance between a point and an interval, a new concept can be introduced. The place value is an indicator of the relative position of a point in relation to two nested intervals. Suppose $U' = \langle a_1, a_2 \rangle$ and $U = \langle b_1, b_2 \rangle$, then the specified place value of point x about the nest of intervals composed of intervals U' and U is

$$D\left(x, U', U\right) = \begin{cases} \rho(x, U) - \rho\left(x, U'\right), \rho(x, U) \neq \rho\left(x, U'\right) \text{ and } x \notin U', \\ \rho(x, U) - \rho\left(x, U'\right) + a_1 - a_2, \rho(x, U) \neq \rho\left(x, U'\right) \text{ and } x \in U', \\ a_1 - a_2, \rho(x, U) = \rho\left(x, U'\right). \end{cases} \tag{4}$$

This describes the locational relation between point x and the nest of intervals composed of U' and U [48].

This new definition of distance is further used to define a new indicator for the measurement of compatibility within an extension set. This indicator is called a dependent function and is defined as follows.

Then for any x, the elementary dependent function K(x) of the optimal point at the midpoint of interval U' is

$$K(x) = \begin{cases} \dfrac{\rho\left(x, U'\right)}{D\left(x, U', U\right)} - 1, \rho\left(x, U'\right) = \rho(x, U) \text{ and } x \notin U', \\ \dfrac{\rho\left(x, U'\right)}{D\left(x, U', U\right)}, \quad \text{others}, \end{cases} \tag{5}$$

This provides an indicator for the degree of compatibility of a given problem that has been expressed numerically, much in the same way that a membership function determines the degree of membership in a fuzzy set. In extenics, however, the dependent function is generalized to the entire real domain so that it also takes into account qualitative changes, as well as quantitative changes. It should be noted that this is the simplest, earliest definition of the dependent function. Since the color of BWV can considerably vary and the corresponding pixel positions are typically random, a dependent function with a nest of two intervals is employed to calculate the correlation coefficient between each pixel and three regions to identify the existence of BWV. Each region has one of the following tags: BWV, healthy skin, and other lesions.

In the dependent function, the classical field of each region is defined as a range of pixel values, e.g., for the region with BWV tag, the values of its pixels are generally in the range of $\langle p_{bw1}, p_{bw2} \rangle$. Therefore, the classical field of the BWV tag is determined as $U'_{bw} = \langle a_{bw1}, a_{bw2} \rangle = \langle p_{bw1}-\delta, p_{bw2} + \delta \rangle$ and the controlled field as $U_{bw} = \langle b_{bw1}, b_{bw2} \rangle = \langle p_{bw1}-2\delta, p_{bw2} + 2\delta \rangle$, where δ is a parameter used to adjust the deviation in the preprocessing step. Similarly, the classical fields of the healthy skin tag and other lesion tags are U'_{hs} and U'_{ol}, and the controlled fields are U_{hs} and U_{ol}, respectively. Taking the region with a BWV tag as an example, the classical field and the controlled field of the region with a BWV tag are described in the image as shown in Fig. 3.

The dependent function with a nest of two intervals is calculated by the following steps [10, 11].

Step 1: The extension distances between pixels and the classical field and controlled field of BWV, healthy skin, and other lesions are calculated. I'_{ij} is transformed into interval form and then denoted as X_{ij}, and $I'_{ij} \in I'_{i \times j}$. The formula is as follows.
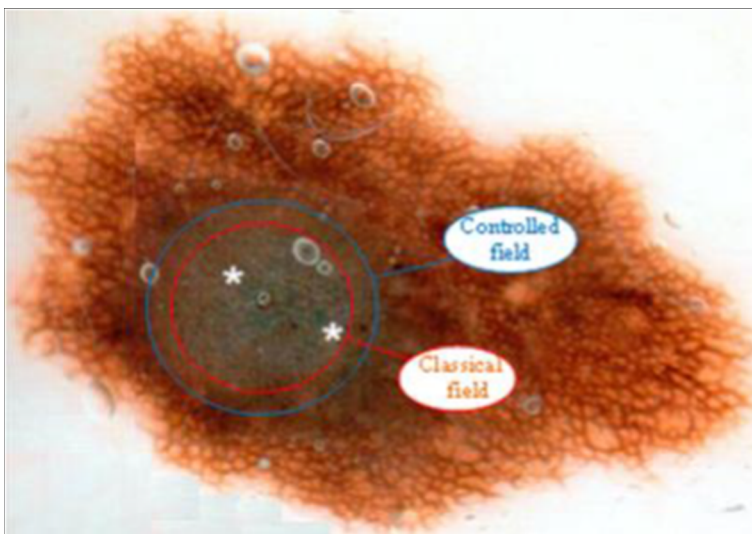

Fig. 3 The classical field and controlled field of the region with a BWV tag, the asterisk part is the BWV area marked by the doctor

$X_{ij} = \langle x_1, x_2 \rangle$, $U'_m = \langle a_1, a_2 \rangle$ and $U_m = \langle b_1, b_2 \rangle$, where $X_{ij} \subset U'_m \subset U_m$; then, the distance between X_i and U'_m or U_m is

$$\rho_m\left(X_{ij}, U'_m\right) = \left| \frac{x_1 + x_2}{2} - \frac{a_1 + a_2}{2} \right| - \left(\frac{a_2 - a_1}{2} - \frac{x_2 - x_1}{2} \right) \tag{6}$$

$$\rho_m\left(X_{ij}, U_m\right) = \left| \frac{x_1 + x_2}{2} - \frac{b_1 + b_2}{2} \right| - \left(\frac{b_2 - b_1}{2} - \frac{x_2 - x_1}{2} \right) \tag{7}$$

where X_{ij} is the interval of a pixel at position (i,j) in the image pixel matrix, U'_m is the classical field, U_m is the controlled field, and $\rho_m(X_{ij}, U'_m)$ represents the distance from BWV, healthy skin or other lesions, as denoted by ρ_{bw}, ρ_{hs}, and ρ_o, respectively.

Step 2: The place values between pixels and the classical field and controlled field of BWV, healthy skin, and other lesions are calculated. The formula is as follows.

$$D_m\left(X_{ij}, U'_m, U_m\right) \tag{8}$$
$$= \begin{cases} a_1 - a_2, & \rho\left(X_{ij}, U_m\right) = \rho\left(X_{ij}, U'_m\right), \\ \rho\left(X_{ij}, U_m\right) - \rho\left(X_{ij}, U'_m\right) + a_1 - a_2, & \rho\left(X_{ij}, U_m\right) \neq \rho\left(X_{ij}, U'_m\right) \text{ and } X_{ij} \subset U'_m, \\ \rho\left(X_{ij}, U_m\right) - \rho\left(X_{ij}, U'_m\right), & \rho\left(X_{ij}, U_m\right) \neq \rho\left(X_{ij}, U'_m\right) \text{ and } \exists x \in X_{ij} - U'_m. \end{cases}$$

$D(x_{ij}, U'_m, U_m)$ describes the locational relation of the nested interval, that is, the locational relation between each pixel in the image and BWV, healthy skin, or other lesions, denoted as D_{bw}, D_{hs}, and D_o, respectively.

Step 3: The dependent function between pixels and the classical field and controlled field of BWV, healthy skin, and other lesions are calculated. The formula is as follows.

$$D_m\left(X_{ij}\right) = \begin{cases} \rho_m\left(X_{ij}, U'_m\right) / D_m\left(X_{ij}, U'_m, U_m\right) - 1, & \rho_m\left(X_{ij}, U'_m\right) = \rho_m\left(X_{ij}, U_m\right) \text{ and } \exists x \in X_{ij} - U'_m, \\ \rho_m\left(X_{ij}, U'_m\right) / D_m\left(X_{ij}, U'_m, U_m\right), & \text{others.} \end{cases} \tag{9}$$

This function represents the correlation degree between each pixel in an image and a BWV, healthy skin or a lesion, denoted as K_{bw}, K_{hs}, and K_o, respectively.

Based on the calculation results of the above three steps, the extension-dependent function matrix $K(X) = (K_{bw}(X), K_{hs}(X), K_o(X))$ of the image is obtained. Images are classified according to the classification principle of extension-dependent function: $K(X) = \max K_m(X)$, m = bw, hs, o. If the dependent function value of the BWV tag is the largest, the corresponding images are classified as malignant melanoma.

3.5 Deep learning-based refined classification method

The model of the YOLOv3 network in our work is shown in Fig. 4. The network first divides the input image into S*S grids, and the grid that contains the center of the object plays a major role in classification. Then, each grid predicts bounding boxes and the corresponding confidence scores, as well as class conditional probabilities. The confidence score is calculated as

$$\textbf{\textit{Confidence}} = \text{P}_r(\textbf{\textit{Object}}) * \text{IoU}_{pred}^{truth}, \text{P}_r(\textbf{\textit{Object}}) \in \{\textbf{0}, \textbf{1}\} \tag{10}$$

where Pr(Object) denotes whether the target is in the grid and IOU indicates the coincidence between the reference and the predicted bounding box. Finally, based on the confidence scores and class probabilities, the pigmented nevi and nail matrix nevi are detected and marked by a box.

According to the diagnosis results, first, labelImg was used to annotate the locations of lesions in medical images. Three class labels are assigned to those annotated regions based on doctors' decisions. The labels and ground-truth annotations of regions are then used as the inputs of YOLOv3 for training. Throughout the training, a batch size of 64 and a momentum of 0.9 are used. The factor of learning rates is $1/3 \approx 0.33$. The pretraining weight given in YOLOv3 is used to accelerate the training speed. The output results are pigment nevus and nail matrix nevus.

Since the lesions of malignant melanoma images are mostly in the middle of the images, there are many background skin areas. YOLOv3 is suitable only for the classification of benign skin diseases. To improve the classification accuracy of the malignant melanoma images, YoDyCK is proposed to classify the malignant melanoma images output by the first
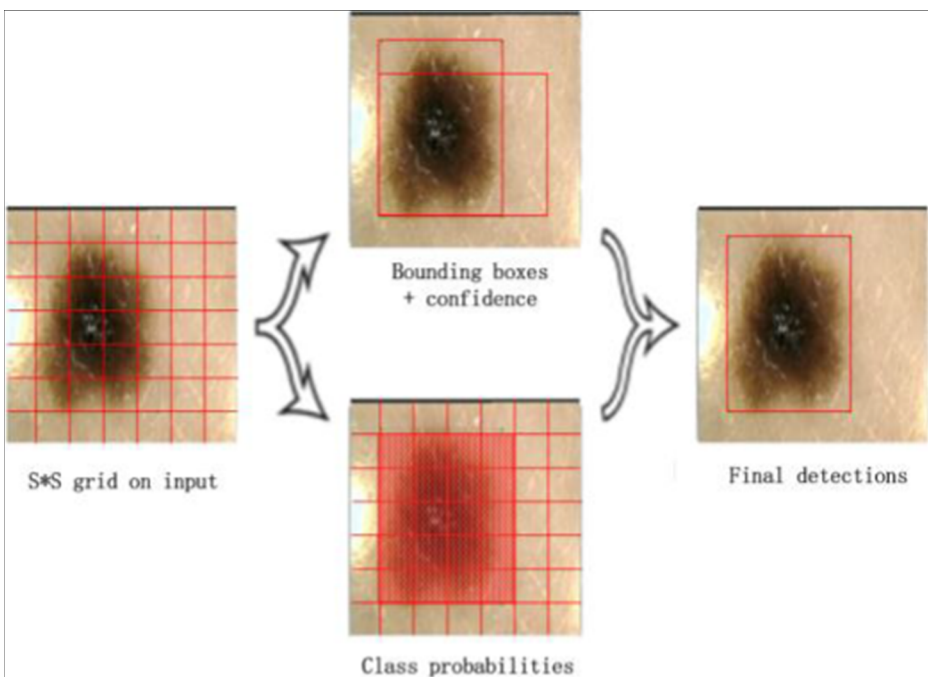

Fig. 4 YOLOv3 detection

phase classification. The structure of YoDyCK is shown in Fig. 5 [53] (DBL:Darknetconv2d_BN_Leaky is the basic component of YOLOv3; resn: n represents a number, res1, res2..., res8, etc., indicating how many res_units are contained in this res_block).

In the YoDyCK model, to ensure that information is not lost in the feature extraction process, we keep a relatively large filter size (11 × 11) for the first convolutional layer and gradually reduce the size of the filter to 3 × 3. In this way, the lesion areas can be quickly located, and the useful features on images can be extracted.

Based on these analyses, our proposed network of YoDyCK is designed. Specifically, Darknet-53 in YOLOv3 is improved in this paper. The improved network is shown in Table 1. Note that d × {}represents that res1, res2, res8, res8, and res4 in Fig. 5 are repeated d times.

Suspected malignant image outputs in the first phase classification are classified by using YoDyCK. The output results are divided into pigment nevus, nail matrix nevus, and malignant melanoma.

4 Evaluations

In this section, the performance of the skin lesion image classification method based on extension theory and deep learning is assessed. First, the experimental environment and the dataset used are described. Then, the results of our experiments are discussed.

4.1 Dataset and experimental environment

Dataset This study was evaluated on two datasets, ISBI 2016 and our dataset. The ISBI 2016 dataset was created for a challenge called Skin Lesion Analysis Toward Melanoma Detection. Our dataset, which contains 1200 dermoscopic images, was collected from a well-known local hospital, and two experienced dermatologists provided image labels in an Excel document. The pixel resolution is 768 × 576. To improve the data quality, some images were excluded if they fell into the following two cases: (1) the lesion was severely blocked by hair or oil products; and (2) the lesion was located in the palms, soles, lips, or private areas. Three hundred dermoscopic images, including 150 pigmented nevus, 100 nail matrix nevus, and 50 malignant melanoma images, were eventually selected, following the original ratio in the complete dataset. The data are also preprocessed by removing noise in the images, such as hair and marks, with an inpainting method.

A lack of images is a typical issue in medical image analysis. Moreover, an imbalanced number of images of various skin diseases will reduce the accuracy of specific disease
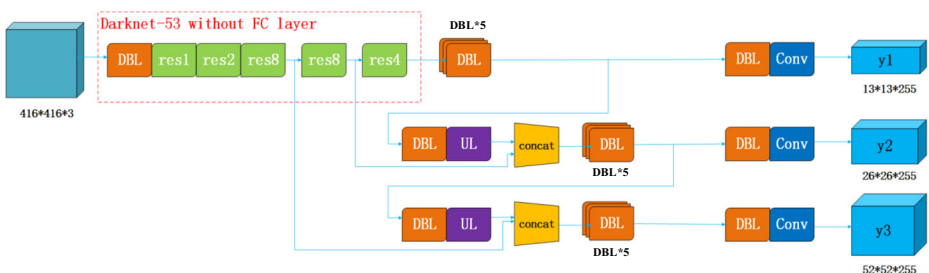

Fig. 5 YoDyCK Structure (DBL:Darknetconv2d_BN_Leaky, is the basic component of YOLOv3)

Table 1 Network parameters

	Type	Filters	Size	Output
	Convolutional	32	3×3	256×256
	Convolutional	64	11×11/2	128×128
1×	**Convolutional**	**32**	**1×1**	
	Convolutional	**64**	**3×3**	
	Residual			**128×128**
	Convolutional	128	9×9/2	64×64
2×	**Convolutional**	**64**	**1×1**	
	Convolutional	**128**	**3×3**	
	Residual			**64×64**
	Convolutional	256	7×7/2	32×32
8×	**Convolutional**	**128**	**1×1**	
	Convolutional	**256**	**3×3**	
	Residual			**32×32**
	Convolutional	512	5×5/2	16×16
8×	**Convolutional**	**256**	**1×1**	
	Convolutional	**512**	**3×3**	
	Residual			**16×16**
	Convolutional	1024	3×3/2	8×8
4×	**Convolutional**	**512**	**1×1**	
	Convolutional	**1024**	**3×3**	
	Residual			**8×8**
	Avgpool		Global	
	Connected		1000	
	Softmax			

identification. To enlarge our dataset, some geometric methods are applied to the original images, such as image scaling (up and down), rotation, and flipping. After geometric data augmentation, there were 750 pigmented nevus, 500 nail matrix nevus, and 250 melanoma images in the dataset.

Although geometric methods can increase the number of images, they do influence the diversity of the data. WGANs are generative models that learn to map samples z from some prior distribution Z to samples q from another distribution Q (e.g., images and audio). The component of the WGAN structure that performs the mapping is called the generator, and its main task is to learn an effective mapping strategy that can imitate the real data distribution and generate novel samples related to those in the training set [3, 20, 26, 46]. This approach can not only increase the diversity of features among image data but also increase the amount of data. To increase the diversity of available medical data collected from a small number of patients, the WGAN is used to balance the dataset. In this study, there were 300 pigmented nevus, 300 nail matrix nevus, and 300 melanoma images in the dataset after data augmentation.

Experimental environment The WGAN, YOLOv3, and YoDyCK models in our method are implemented with the Keras framework. To train the deep learning network, an i7-8700k CPU and two 1080ti GPUs are used.

4.2 Experiments and results

4.2.1 Influence of data augmentation method

In our method, WGAN is employed to generate images with features similar to those in existing images and enlarge the size of the training dataset. To evaluate the performance of the data augmentation method, 300 raw images are selected from our dataset, and 900 fake images are generated based on the raw images with geometric methods such as image scaling (up and down), rotation and flipping, and the WGAN. In the training phase of the WGAN, the choice of iteration time can affect the quality of the generated fake images, and the number of iterations is 3000. Then, two YOLOv3 models are trained with the fake images generated by the geometric method and WGAN. The performance of the trained YOLOv3 models in classifying benign skin diseases reflects the quality of the generated images. One hundred images are chosen from our dataset as the test dataset, and the classification results are shown in Table 2.

The bottom row of the table shows the average accuracy of the classification, and it can be observed that the model trained with the images generated by the WGAN classifies the test dataset with an accuracy of 90.8%, while the other model only achieves an accuracy of 67.3%; this result indicates that the images generated by the WGAN contain more of the features of the original images than do those generated by the geometric method. In addition, the results also suggest that the YOLOv3 model can classify nail matrix nevi and pigment nevi with an accuracy of greater than 90%, and the accuracy of classifying malignant melanoma is much lower than that for benign nevi.

4.2.2 Influence of the quantity of experimental data

In this section, the impact of the size of the image dataset on the YOLOv3 model is analyzed. 10, 20, 60, 100, 200, 300, 400, 500, and 600 images are randomly selected from each of the three skin diseases to form training sets of 30, 60, 180, 300, 600, 900, 1200, 1500, and 1800 images. The precision rate for the models corresponding to training sets of different sizes is shown in Fig. 6.

From these experiments, one can conclude that the performance of the YOLOv3 model improves as the size of the training set increases. When the number of images exceeds 900, the size of the training set does not have a further significant influence on the performance of the model.

Table 2 Comparison of the classification accuracy of different data augmentation methods

Method	Geometric method			Our method		
Lesions	Classification accuracy	Accuracy of scores over 0.5(%)	Average confidence	Classification accuracy	Accuracy of scores below 0.5 (%)	Average confidence
Pigmented nevi	90	77.5	0.72	100	97.5	0.79
Nail matrix nevi	67.5	50	0.567	90	87.5	0.86
Malignant melanoma	45	32.5	0.64	82.5	80	0.82
Average	67.5	53.3	0.64	90.8	88.3	0.82

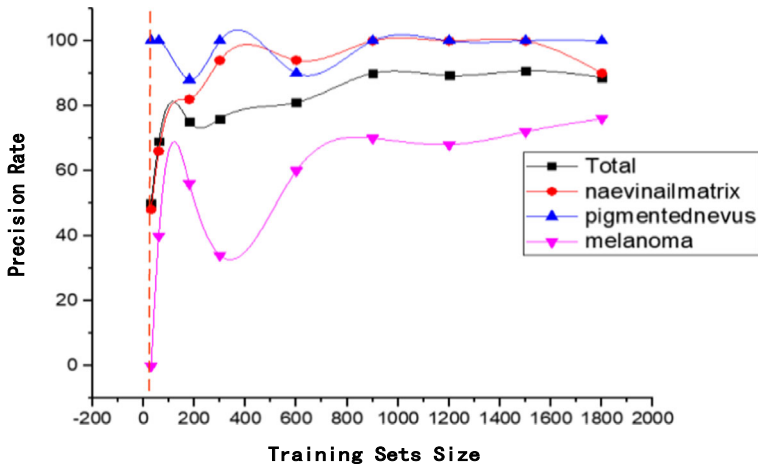**Fig. 6** Classification accuracy in the data incremental process

4.2.3 Influence of the learning rate

When a neural network is trained, the learning rate is usually the most important hyperparameter that needs to be adjusted. If the learning rate is too large, it will overshoot low-loss areas. If the learning rate is too small, the network may not be able to train. Therefore, in the training process of our model, a step-based decay learning rate scheme was used to find a more optimized landscape area loss. In the early stage, a reasonable set of weights should be found with a higher learning rate, these weights should be adjusted in the subsequent process, and a better weight should be found with a smaller learning rate. When Factor = 0.5, 0.33, and 0.25, the decay process of the learning rate was compared. As shown in Fig. 7.

As shown in Fig. 7, the step-based decay learning rate scheme is a piecewise function, and the smaller the factor is, the faster the learning rate will decay. When factor = 0.5, there was still a significant decline in the learning rate after 2500 epochs; when factor = 0.25, the learning rate leveled off after 1500 epochs. To improve the accuracy of our model and reduce the loss, fcator = 0.33 ≈ 1/3 was selected to better train the model weights.

4.2.4 Performance of the first phase classification method

In the first phase, 300 skin lesion images are selected, including 100 pigmented nevus, 100 nail matrix nevus, and 100 malignant melanoma images. An extension theory-based method is employed to classify skin lesion images into two classes: benign skin lesion images and suspected malignant skin lesion images. The parameters are defined below:

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

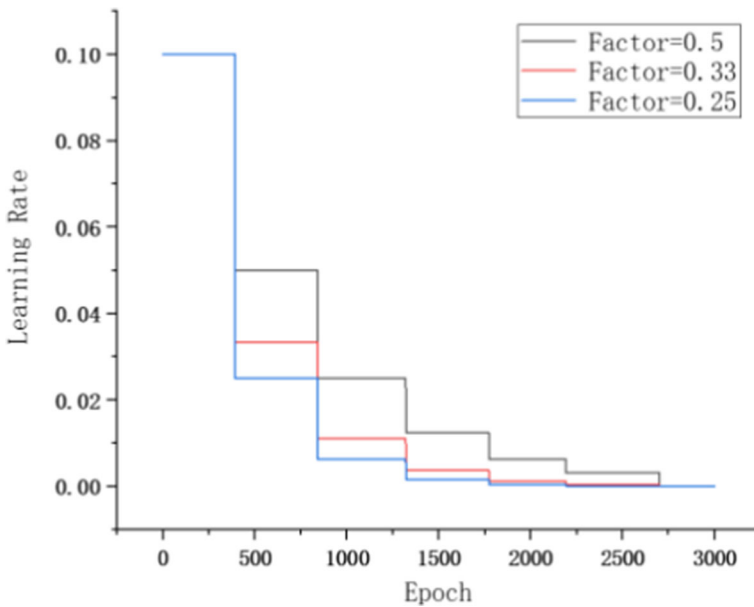$$Recall = \frac{TP}{TP + FN} \tag{12}$$

Fig. 7 Decay curve of learning rate

$$F_1 \boldsymbol{Store} = \frac{2*P*R}{P + R} \tag{13}$$

where TP indicates the number of suspected malignant skin lesion images successfully classified by the method. FP indicates the number of benign skin lesion images that are falsely classified as suspected malignant skin lesion images. FN indicates the number of suspected malignant skin lesion images that are falsely classified as benign skin lesion images. TN indicates the number of benign skin lesion images that are successfully classified. The classification results are shown in Table 3.

From the results of Table 3, P = 0.81 and R = 1 are calculated. And the result of the F1 score is 0.895, which indicates that the extension theory-based method performs well in classifying skin lesion images into benign and suspected malignant.

To verify the classification accuracy by using extension theory based on the BWV feature is much better, the performance of our method is compared with several typical malignant melanoma classification methods based on other features. Here, the methods will be evaluated based on the accuracy. The results are shown in Table 4.

The results showed that our method has a better effect. After the first phase classification, 300 skin lesion images are divided into 176 benign skin lesion images and 124 suspected malignant skin lesion images.

Table 3 The classification results of the first phase classification method

	Predicted class: suspected malignant	Predicted class: benign
Actual class: suspected malignant	100(TP)	0(FN)
Actual class: benign	24(FP)	176(TN)

Table 4 The comparison results of the methods based on different features

Methods	Accuracy
GVF-Snake (based on irregular contour)	0.782323
Watershed method (based on pigment reticular formation)	0.7471
SVM (based on irregular contour and pigment reticular formation)	0.8506
Our method (based on BWV)	0.92

4.2.5 Performance of the second phase classification method

In the second phase, YOLOv3 is used to classify benign skin lesion images, and the output is divided into pigmented nevi and nail matrix nevi. YoDyCK is used to classify suspected malignant skin lesion images, and the output is divided into pigmented nevi, nail matrix nevi, and malignant melanoma. The classification results are shown in Table 5.

The table shows that the benign skin lesion image classification accuracy by using YOLOv3 is 0.94. The suspected malignant skin lesion image classification accuracy by using YoDyCK is 0.92. The classification accuracy after the second phase is 0.93. The results show that our two-phase classification method can classify the three types of skin lesion images, and it has higher accuracy.

4.2.6 Comparison with other classification methods on our dataset and the ISBI 2016 dataset

To verify the effect of the two-phases classification method for skin lesion images proposed in this paper is better than other methods. Our method is compared with several state-of-the-art classification methods by using our dataset and the ISBI 2016 dataset. The methods will be evaluated based on three metrics: accuracy, sensitivity, and specificity. The results are shown in Table 6.

For skin lesion image classification, as shown in Table 6, it is clear that our method (two-phase classification) achieves the best results on all of the evaluation metrics, which

Table 5 Classification accuracy of the two-phases classification method

Items	Benign skin lesion images			Suspected malignant skin lesion images			Total
Diseases	Correct classification	Misclassification	Unable to classify	Correct classification	Misclassification	Unable to classify	
Pigmented nevi	90	0	1	8	0	1	100
Nail matrix nevi	76	4	5	14	1	0	100
Malignant melanomas	–	–	–	92	0	8	100
Total	166	4	6	114	1	9	300
	176			124			

Table 6 The comparison results of different classification methods on our dataset and ISBI 2016

Dataset	Methods	Accuracy	Sensitivity	Specificity
Our dataet	Two-level ensembles [57]	0.826	0.792	0.860
	Boosting ensemble [54]	0.773	0.743	0.803
	Faster-RCNN [45]	0.832	0.798	0.867
	Ours-YOLOv3-only	0.905	0.869	0.941
	Ours-YoDyCK-only	0.910	0.879	0.940
	Our method	0.930	0.919	0.943
ISBI 2016	Two-level ensembles [57]	0.829	0.783	0.875
	Boosting ensemble [54]	0.787	0.769	0.805
	Faster-RCNN [45]	0.887	0.835	0.939
	Ours-YOLOv3-only	0.925	0.883	0.967
	Ours-YoDyCK-only	0.937	0.895	0.979
	Our method	0.962	0.937	0.987

demonstrates the effectiveness of the proposed two-phase classification method. At the same time, in the methods of two-level ensembles, boosting ensemble, faster-RCNN, ours-YOLOv3-only, and ours-Yodyck-only, the training time of ours-Yodyck-only is the shortest. The training time of our method is 56.4 min faster than ours-Yodyck-only. Finally, the visualization of the classification results in this paper is shown in Fig. 8.
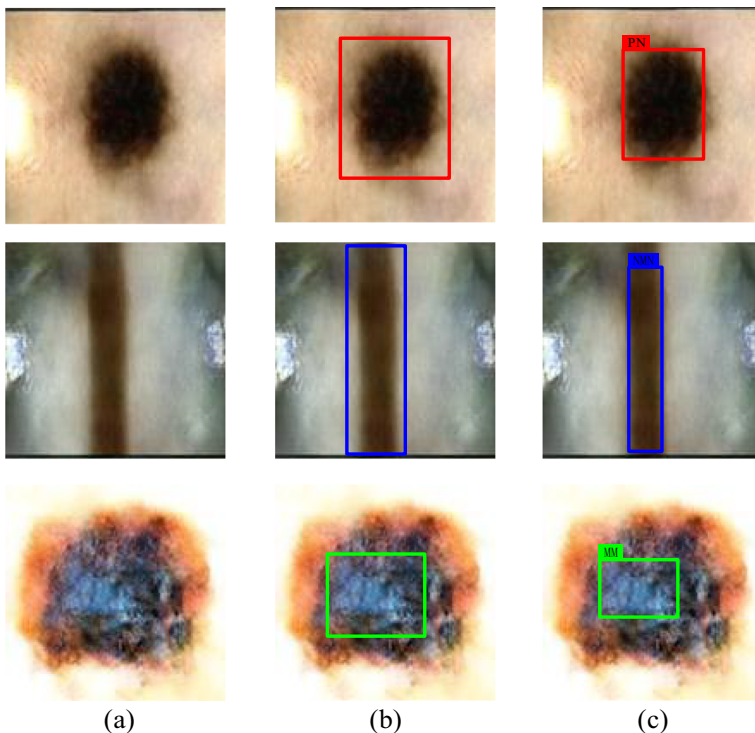

Fig. 8 Visualization of the classification results. **a** Original images of pigmented nevus (PN), nail matrix nevus (NMN), and malignant melanoma (MM). **b** Ground truth of pigmented nevus, nail matrix nevus, and malignant melanoma. **c** The classification results of our method

5 Conclusions

In this paper, automatically and accurately classifying skin lesion images is studied, with the goal of helping medical professionals make the diagnosis more accurately and efficiently. Our research solves three questions: (1) integrate medical domain features with deep learning features well within a unified framework; (2) detect the existence of BWV in a rapid and accurate way; and (3) prevent misdiagnosis of malignant skin diseases. In our method, a two-phase classification framework integrated with medical domain knowledge, deep learning, and a refined strategy is proposed to classify images into three classes: pigmented nevi, nail matrix nevi, and malignant melanoma. A BWV detection method based on extension theory is presented that can accurately explore visual information of malignant skin diseases. The YoDyCK method is proposed to extract the deep learning feature from skin lesion images and rapidly detect and locate the lesion area. On two skin lesion image datasets, the effectiveness of the proposed methods is demonstrated through studies. BWV is a critical medical domain feature for the diagnosis of malignant melanoma. Therefore, in our future studies, the multiarea and high-precision automatic segmentation of BWV and the edge of the lesion area will be further studied to better assist doctors in locating the resection area.

Appendix

Abbreviated form	Full form
YoDyCK	YOLOv3 optimized by Dynamic Convolution Kernel
BWV	Blue White Veil (a very critical feature which summarized based on expert experience for the diagnosis of malignant melanoma)
YOLOv3	the third version of You Only Look Once

Authors' contributions XB, HP, KZ, PL, LJ and CC conceived of the study. XB, HP and PL performed the collection and label dermoscopy images. XB and LJ carried out the experiment. XB, HP, KZ and CC analyzed the results. All authors read and approved the final manuscript.

Declarations

Competing interests The authors declare that they have no competing interests.

References

1. Alfed N, Khelifi F (2017) Bagged textural and color features for melanoma skin cancer detection in dermoscopic and standard images. Expert Syst Appl 90:101–110
2. Argenziano G, Fabbrocini G, Carli P, De Giorgi V, Sammarco E, Delfino M (1998) Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions: comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis. Arch Dermatol 134(12):1563–1570
3. Arjovsky M, Bottou L (2017) Towards principled methods for training generative adversarial networks. Stat 1050
4. Arroyo JLG, Zapirain BG, Zorrilla AM (2011) Blue-white veil and dark-red patch of pigment pattern recognition in dermoscopic images using machine-learning techniques. In: 2011 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), (12):196–201
5. Aswin RB, Jaleel JA, Salim S (2014) Hybrid genetic algorithm—Artificial neural network classifier for skin cancer detection. In: 2014 International conference on control, instrumentation, Communication and Computational Technologies (ICCICCT), (7):1304–1309.
6. Celebi ME, Iyatomi H, Stoecker WV, Moss RH, Rabinovitz HS, Argenziano G, Soyer HP (2008) Automatic detection of blue-white veil and related structures in dermoscopy images. Comput Med Imaging Graph 32(8):670–677
7. Chandy DA, Johnson JS, Selvan SE (2014) Texture feature extraction using gray level statistical matrix for content-based mammogram retrieval. Multimed Tools Appl 72:2011–2024
8. Chen H, Maharatna K (2020) An automatic R and T peak detection method based on the combination of hierarchical clustering and discrete wavelet transform. IEEE J Biomed Health Inform 24(10):2825–2832
9. Chen C-C, DaPonte JS, Fox MD (June 1989) Fractal feature analysis and classification in medical imaging. IEEE Trans Med Imaging 8(2):133–142. https://doi.org/10.1109/42.24861
10. Chen X, Bian X et al (2016) Construction method of uncertain type elementary dependent function in two nested regions. J Inner Mongolia Univ Nationalities 31(3):185–188
11. Chen X, Bian X et al (2018) Construction method of uncertain type elementary correlation function under three nested regions. *J Heilongjiang Univ Sci Technol* 28(1):124–128
12. Chen B et al (2020) Label co-occurrence learning with graph convolutional networks for multi-label chest X-ray image classification. IEEE J Biomed Health Inform 24(8):2292–2302
13. Choi YH, Tak YS, Rho S, Hwang E (2013) Skin feature extraction and processing model for statistical skin age estimation. Multimed Tools Appl 64:227–247
14. Codella NC, Gutman D, Celebi ME, Helba B, Marchetti MA, Dusza SW, ... Halpern A (2018) Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), (4):168–172
15. Di Leo G, Fabbrocini G, Paolillo A, Rescigno O, Sommella P (2009) Towards an automatic diagnosis system for skin lesions: estimation of blue-whitish veil and regression structures. In: 2009 6th international multi-conference on systems, Signals and Devices, (3):1–6
16. Di Leo G, Paolillo A, Sommella P, Fabbrocini G (2010) Automatic diagnosis of melanoma: a software system based on the 7-point check-list. In: 2010 43rd Hawaii international conference on system sciences, (11):1–10
17. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S (2017) Corrigendum: "Dermatologist-level classification of skin cancer with deep neural networks". *Nature* (546):686
18. Feng-Xu G, Wang K-J (2006) Study on extension control strategy of pendulum system. J Harbin Inst Technol 38(7):1146–1149
19. Florentin S, Victor V (2012) Applications of Extenics to 2D-Space and 3D Space," *The 6th Conference on Software, Knowledge, Information Management and Applications, Chengdu, China*, (12):9–11
20. Frid-Adar M, Diamant I, Klang E, Amitai M, Goldberger J, Greenspan H (2018) GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. Neurocomputing 321:321–331
21. Ganster H, Pinz P, Rohrer R, Wildling E, Binder M, Kittler H (2001) Automated melanoma recognition. IEEE Trans Med Imaging 20(3):233–239
22. Gao L, Pan H, Han Q et al (2015) Finding frequent approximate subgraphs in medical image database. IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 1004–1007.
23. Greenspan H, Van Ginneken B, Summers RM (2016) Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique. IEEE Trans Med Imaging 35(5):1153–1159
24. Haralick RM, Shanmugam K, Dinstein I (Nov. 1973) Textural features for image classification. IEEE Trans Syst Man Cybern SMC-3(6):610–621. https://doi.org/10.1109/TSMC.1973.4309314
25. He B, Zhu X (2005) Hybrid extension and adaptive control. Control Theory Appl 22(2):165–170

26. Isola P, Zhu JY, Zhou T, Efros AA (2017) Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 1125–1134
27. Kittler H, Pehamberger H, Wolff K, Binder MJTIO (2002) Diagnostic accuracy of dermoscopy. The Lancet Oncology 3(3):159–165
28. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems (25):1097–1105
29. Melbin K, Jacob Vetha Raj Y (2021) Integration of modified ABCD features and support vector machine for skin lesion types classification. Multimed Tools Appl 80(6):8909–8929
30. Mhaske HR, Phalke DA (2013) Melanoma skin cancer detection and classification based on supervised and unsupervised learning. In: 2013 International conference on circuits, Controls and Communications (CCUBE), (12):1–5
31. Nver HM, Ayan E (2019) Skin Lesion Segmentation in Dermoscopic Images with Combination of YOLO and GrabCut Algorithm. *Diagnostics* 9(3):72
32. Pan H, Li P, Li Q et al (2013) Brain CT image similarity retrieval method based on uncertain location graph. IEEE J Biomed Health Inform 18(2):574–584
33. Pang S et al (2019) A novel YOLOv3-arch model for identifying cholelithiasis and classifying gallstones on CT images. PLoS ONE 14(6):e0217647
34. Redmon J, Farhadi A (2017) YOLO9000: better, faster, stronger. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 7263–7271
35. Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 779–788
36. Ren S, He K, Girshick R, Sun J (2015) Faster r-cnn: towards real-time object detection with region proposal networks. In: Advances in neural information processing systems (28):91–99
37. Röhrich E, Thali M, Schweitzer W (2012) Skin injury model classification based on shape vector analysis. *BMC Med Imaging* 12:32. https://doi.org/10.1186/1471-2342-12-32
38. Roslin SE (2020) Classification of melanoma from Dermoscopic data using machine learning techniques. Multimed Tools Appl 79(5):3713–3728
39. Roth HR, Lu L, Liu J, Yao J, Seff A, Cherry K, … Summers RM (2015) Improving computer-aided detection using convolutional neural networks and random view aggregation. IEEE Trans Med Imaging 35(5):1170–1181
40. Seidenari S, Ferrari C, Borsari S, Benati E, Ponti G, Bassoli S, … Pellacani G (2010) Reticular grey-blue areas of regression as a dermoscopic marker of melanoma in situ. Br J Dermatol 163(2):302–309
41. Setiawan AW, Faisal A (2020) A study on JPEG compression in color retinal image using BT.601 and BT.709 standards: image quality assessment vs. file size. *2020 International Seminar on Application for Technology of Information and Communication (iSemantic)*, 436–441
42. Setiawan AW, Faisal A, Resfita N (2020) Effect of image downsizing and color reduction on skin cancer pre-screening. *2020 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, 148–151
43. Stoecker WV et al (2011) Detection of granularity in dermoscopy images of malignant melanoma using color and texture features. *Comput Med Imaging Graph* 35(2):144–147
44. Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, Liang J (2016) Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Trans Med Imaging 35(5): 1299–1312
45. Ulhaq A, Khan A, Robinson R (2020) Evaluating faster-RCNN and YOLOv3 for target detection in multi-sensor data. In: Statistics for Data Science and Policy Analysis, 185-193
46. Vitoria P, Sintes J, Ballester C (2019) Semantic image inpainting through improved Wasserstein generative adversarial networks. *14th International Conference on Computer Vision Theory and Applications*
47. Warsi F, Khanam R, Kamya S, Suárez-Araujo CP (2019) An efficient 3D color-texture feature and neural network technique for melanoma detection. Inform Med Unlocked 17:100176
48. Wen C (1983) Extension set and non-compatible problems. J Sci Explor 1:83–97
49. Wen C, Yong S (2006) Extenics, its significance in science and prospects in application. J Harbin Inst Technol 38(7):1079–1086
50. Yang C(2005) "The Methodology of Extenics", "Extenics: Its Significance in Science and Prospects in Application," *The 271th Symposium's Proceedings of Xiangshan Science Conference*, 12:35–38
51. Yang C, Wen C (2007) Extension engineering. Science Press, Beijing
52. Yang C, Weihua L, Xiaomei L (2011) Recent research Progress in theories and methods for the intelligent disposal of contradictory problems. J Guangdong Univ Technol 28:86–93
53. YOLOv3 Structure (n.d.), available online on: https://blog.csdn.net/qq_30815237/article/details/91949543. Accessed on 7-10-2020

54. Yun Y, Gu I (2013) Image Classification by Multi-Class Boosting of Visual and Infrared Fusion with Applications to Object Pose Recognition. *Swedish Symposium on Image Analysis (SSBA)*, (3):14–15
55. Zarit B, Super B, Quek F (n.d.) Comparison of five color models in skin pixel classification," Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, 58–63. https://doi.org/10.1109/RATFG.1999.799224
56. Zhang X, Zhu X (2019) Vehicle Detection in the Aerial Infrared Images via an Improved Yolov3 Network. 2019 *IEEE 4th International Conference on Signal and Image Processing (ICSIP). IEEE*
57. Zhang-China B, Pham-Australia TD (2010) Multiple Features Based Two-stage Hybrid Classifier Ensembles for Subcellular Phenotype Images Classification. Int J Biom Bioinforma 8:554–562
58. Zhao Yanwei S (2010) Extension Design. Science Press, Beijing