Melanoma Classification on Dermoscopy Images Using a Neural Network Ensemble Model

Fengying Xie,* Haidi Fan, Yang Li, Zhiguo Jiang, Rusong Meng, and Alan Bovik, *Fellow, IEEE*

Abstract—We develop a novel method for classifying melanocytic tumors as benign or malignant by the analysis of digital dermoscopy images. The algorithm follows three steps: first, lesions are extracted using a self-generating neural network (SGNN); second, features descriptive of tumor color, texture and border are extracted; and third, lesion objects are classified using a classifier based on a neural network ensemble model. In clinical situations, lesions occur that are too large to be entirely contained within the dermoscopy image. To deal with this difficult presentation, new border features are proposed, which are able to effectively characterize border irregularities on both complete lesions and incomplete lesions. In our model, a network ensemble classifier is designed that combines back propagation (BP) neural networks with fuzzy neural networks to achieve improved performance. Experiments are carried out on two diverse dermoscopy databases that include images of both the xanthous and caucasian races. The results show that classification accuracy is greatly enhanced by the use of the new border features and the proposed classifier model.

Index Terms—Dermoscopy image, neural network ensemble, feature extraction, lesion classification.

I. INTRODUCTION

MALIGNANT melanoma (MM) is the third most frequent type of skin cancer and one of the most malignant cancers [1], [2]. Invasive melanoma alone has an estimated incidence of 73,870 and an estimated total of 9,940 melanoma deaths occurred in the United States in 2015 [3]. As compared with the U.S., Europe and Australia, the incidence of skin cancer in China is lower, but it has still been increasing 3%-8% annually and has doubled over the past decade [4].

Dermoscopy is a non-invasive skin imaging technique which allows a magnified visualization of the skin surface and subdermal structures [5]. Dermoscopy images have played a significant role in increasing the survival rate of patients by assisting the early diagnosis of MM. However, diagnoses

that are made by human experts, while still the clear gold standards, are nevertheless subjective. Indeed, the accuracy and reproducibility of diagnosis is highly dependent on the expertise of the physician. It has been reported that dermoscopy may actually lower diagnostic accuracy when conducted by inexperienced dermatologists [6]. Computer-aided diagnosis (CAD) systems for skin cancer, which are not subjective, can assist the physician when making decisions via lesion border detection, quantifying diagnostic features, classifying lesions by type, etc [2]. According to Schindewolf *et al.* [7], the accuracy of diagnosis of MM can be increased from 75% up to as high as 92% when skilled dermatologists use CAD systems to evaluate skin cancer.

Usually, a computerized dermoscopy image analysis model includes four aspects: preprocessing, segmentation, feature extraction and classification. Studies related to classification of lesion objects can be found in the literature as early as 1987 [8]. Pigmented skin lesions are typically evaluated by dermatologists using the "ABCD" rule [9], [10], which analyzes the Asymmetry, Border irregularity, Color variation and Different structures of a lesion. Based on the "ABCD" rule, many classification methods have been developed on diverse datasets of dermoscopy images. In [11], Kusumoputro *et al.* extracted 18 shape and color features from dermoscopy images and used to train an artificial neural network to separate MM from benign lesions. In [12], Ganster *et al.* extracted a set of features from lesion images, including shape, border gradient and color colorbluedescriptors. The most discriminative features among these were then selected. A K-Nearest Neighbor (KNN) classifier was finally trained on these features to deliver a sensitivity of 87% with a specificity of 92%. In [13], 437 features descriptive of color, texture and shape were extracted, from which 18 optimal features were selected. A support vector machine (SVM) was then trained to classify the lesions into benign nevi or melanoma. Capdehourat *et al.* [14] characterized each candidate lesion by a set of features containing shape, color and texture information, which were used to train an AdaBoost classifier with C4.5 decision trees. This model delivered a specificity of 77% with sensitivity of 90% on an automatically segmented database, and a specificity of 85% with sensitivity of 90% on a manually segmented database.

Most of the existing literature regarding computer-assisted lesion classification has focused on feature extraction and classifier design on images that are either explicitly, or implicitly, assumed to contain a complete lesion object. However,

Fig. 1. Dermoscopy images containing incomplete lesion objects.

images may not always capture entire lesions, as shown in Fig. 1. Local features can be used to deal with the complex situations. Situ *et al.* [15] extracted local features from 16×16 image patches using wavelets and Gabor-like filters, then analyzed the responses using a bag-of-features (BoF) model to recognize MM. In [16], Barata *et al.* extracted texture and color features, then used a BoF model to code these features to classify lesions. They then improved classification performance by imposing a color constancy constraint [17]. The features in [16] and [17] were extracted from image patches, in order to better cope with incomplete lesions. However, border features were ignored, which are important for lesion diagnosis. Here, we proposed a dermoscopy tumor classification model that aims to handle incomplete tumor presentations. The model utilizes a set of tumor border features along with other tumor-descriptive features, which are fed to a neural network meta-ensemble model that is trained to differentiate malignant lesions from benign lesions.

The remainder of the paper is organized as follows. Section II describes preprocessing and segmentation. In Section III, feature extraction is presented. Section IV introduces the neural network ensemble model. In Section V and Section VI, we present and discuss the experimental results. Finally, Section VII gives concluding remarks.

II. PREPROCESSING AND SEGMENTATION

Hair removal is the purpose of the preprocessing stage. Dermoscopy image analysis can be greatly complicated by the presence of hair. For these images, hairs are detected and removed using the partial differential equation (PDE)-based image repair method presented in [18].

The accuracy of the segmentation process greatly affects subsequent feature extraction and classification. Although a number of automatic segmentation methods for dermoscopy images have been proposed, segmentation is usually achieved via a semi-automatic method involving manual interaction, or by completely manual segmentation [13], [17]. In our earlier work [19], an automatic method based on a self-generating neural network (SGNN) model was developed to segment dermoscopy images, and was able to obtain more accurate results under complex conditions than Otsu's threshold [20], k-means, fuzzy c-means and statistical region merging (SRM) [21]. Here, we use a SGNN model to segment the images as a proxy for either fully automatic or semi-automatic segmentation. Figure 2 shows two segmentation results obtained using the SGNN model of [19]. Since some of the experimental images were not accurately segmented by the SGNN method (for example Fig. 2(b)), we manually segmented these to obtain more accurate borders.
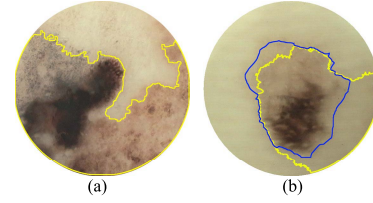

Fig. 2. Segmentation instances on two dermoscopy images (yellow line: SGNN, blue line: manual).
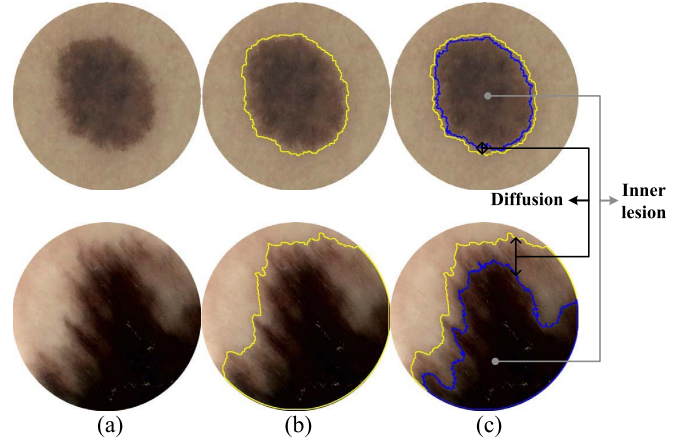

Fig. 3. Generation of two lesion regions. (a) Original lesion image. (b) Segmentation result on (a) Using SGNN. (c) Using SGNN followed by Otsu's threshold.

III. FEATURE EXTRACTION

The most common features mentioned in the literature on dermoscopic lesion classification involve color, texture, and shape [13], [22], [23]. Border features are less well-described [12], [24]. Since we are interested in handling images containing incomplete lesion objects in our dataset, shape features are abandoned in our model. In the following, we describe a set of widely used color and texture features as well as a set of lesion border features that are effective on incomplete lesions. These features are used in our classification method.

A. Region Division on Dermoscopy Images

In [13], Celebi *et al.* computed features over three regions: lesion, inner periphery, and outer periphery. However, on images containing incomplete lesions, some color and structural asymmetry features cannot be calculated correctly using the division proposed in [13]. Therefore, we instead divide the lesion object into two regions: a diffusion region and an inner lesion region. First, each lesion object is separated from the background skin using a combination of the SGNN method and manual interaction, as described in Section II. Then, Otsu's threshold [20] is used to automatically segment the lesion into these two region types. Figure 3 is an example showing generation of these two regions. In our method, features are extracted on both individual regions, as well as over the entire lesion region.

B. Description of Color and Texture Features

The color and texture features used in our model are the same as those described by Celebi *et al.* in [13]. We briefly summarize these features.

1) *Color features*:

(a) **RGB features:** A total of 24 color features are extracted from each dermoscopy image represented in RGB color space. Of these, 18 are calculated as follows: for each of the 3 RGB channels, 2 statistics (the channel mean and standard deviation) are computed for 3 segmented regions (diffusion, inner lesion, entire lesion object). The remaining 6 features are the ratios of corresponding statistics from inner and diffusion regions. Denoting μ_{region} and σ_{region} as the mean and standard deviation of the channel value within a region, 3 of these features are $\frac{\mu_{inner}}{\mu_{transition}}$ on 3 channels, with similar features defined using σ_{region}.

(b) **LUV histogram distances:** The LUV color space is coarsely quantized into $4 \times 8 \times 8$ bins. The color similarity between the inner region and the transition region is expressed as the L1-norm and L2-norm of the difference histograms, yielding 2 additional color features.

(c) **Color diversity:** RGB color space is divided into $16 \times 16 \times 16$ bins. Then the number of the bins, into which there are pixels to be divided, is calculated. The larger the number of bins, the more that likely the lesion is malignant. One feature is found here.

(d) **Centroidal distances:** Given a segmented lesion object, the distance is calculated between the geometric centroid of the object and the brightness centroid of the same. If the pigmentation of the object is homogeneous, the brightness and geometric centroids should be similar. This is true even for incomplete objects. Three features are obtained by computing this distance from each RGB channel.

2) *Texture Features:*
The spatial variations, randomness and correlations of local regional image brightness are important descriptions of possible malignancy. The gray level co-occurrence matrix (GLCM)-based texture descriptor is one of the most well-known and widely used methods in the literature [25]. In this paper, 5 statistical texture descriptors are calculated from the GLCM: the regional *energy, entropy, contrast, inverse difference moment* and *correlation*. These 5 statistics are also calculated over each of the 3 segmented regions (diffusion, inner lesion and the entire lesion object). Again, the ratios of the 5 statistics on the diffusion and inner lesion regions are also calculated. The total number of texture features extracted from each candidate lesion image is 20.

C. Description of Border Features

Border features [12], [24] that are used to characterize the degree of border irregularity of lesions are generally based on measurements of color, texture or brightness gradient. Ganster *et al.* [12] calculated the gradient and normalized brightness values of border regions. Various sample statistics, such as the minimum, maximum, average, and variance, were used as border features. Similar to the method in [12], the approach in [24] used a variety of gradient features to characterize border regions, including the mean and variance of the gradient magnitude along the candidate lesion border, as well as over 8 symmetric regions around the candidate
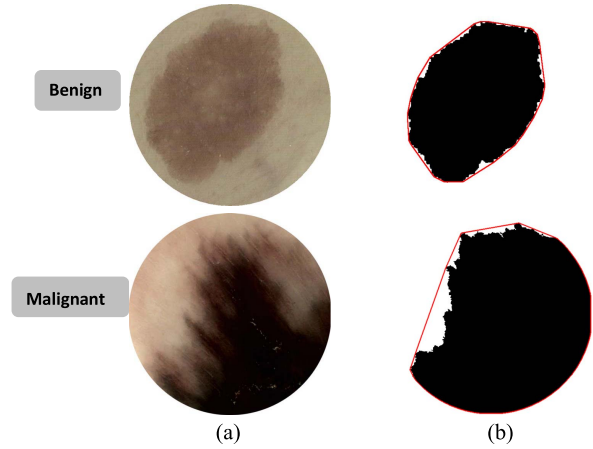

Fig. 4. Lesion regions and their convex hulls. (a) Dermoscopy images of benign (top) and malignant (bottom) lesions. (b) Segmented lesion regions, where red lines are superimposed convex hulls.
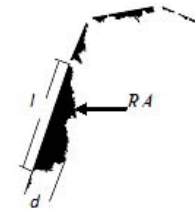

Fig. 5. Extracted concavities on the malignant lesion in Fig. 4.

border in three separate channels: color, texture and local skin darkness. These features [12], [24] were only extracted on complete lesion objects. In order to extract border features from incomplete objects, new border features are needed that can operate on incomplete lesions, thus weakening the dependence of lesion classification on perfect acquisition of dermoscopy images.

1) *Lesion Concavity Features:*
Benign lesions tend to have roughly elliptical shapes with roundish convex borders that exhibit few concavities. Malignant lesions commonly exhibit highly irregular borders that can be characterized as having more frequent concavities. As such, the convex hull is a useful tool for identifying border concavities. Figure 4 shows the convex hull of both malignant and benign segmented lesions. We use the lesion convex hull to define 6 feature parameters that are descriptive of the degree of concavities occurring on the lesion border.

Figure 5 shows concavities that were extracted from the malignant lesion image in Fig. 4. Figure 5 depicts ways of quantifying border concavities. Let N_r denotes the number of distinct concave segments along the border of the lesion object, and let l_i, d_i and RA_i be the span, depth and area of the ith concavity. Then 3 pairs of sample statistics are defined on the lesion object as follows:

(a) **Mean and standard deviation of span:**

$$\mu_s = \frac{1}{N_r} \sum_{i=1}^{N_r} l_i \tag{1}$$

$$\delta_s = \sqrt{\frac{1}{N_r} \sum_{i=1}^{N_r} (l_i - \mu_s)^2} \tag{2}$$

(b) **Mean and standard deviation of depth:**

$$\mu_d = \frac{1}{N_r} \sum_{i=1}^{N_r} d_i \tag{3}$$

$$\delta_d = \sqrt{\frac{1}{N_r} \sum_{i=1}^{N_r} (d_i - \mu_d)^2} \tag{4}$$

(c) **Mean and standard deviation of the average thickness:**

$$\mu_t = \frac{1}{N_r} \sum_{i=1}^{N_r} \frac{RA_i}{l_i} \tag{5}$$

$$\delta_t = \sqrt{\frac{1}{N_r} \sum_{i=1}^{N_r} \left(\frac{RA_i}{l_i} - \mu_t \right)^2}. \tag{6}$$

On benign lesions, concavities tend to be smaller, shallower, and more uniform in both span and depth. On malignant lesions, highly variable spans, depths and thicknesses tend to occur. Therefore, the values of these 6 feature parameters computed on malignant lesions are larger than on the benign lesion. The same observations are true on images of incomplete objects, keeping in mind that artificial boundaries at the frame of the image are not included in the computation.

2) Separation Between Inner and Outer Borders: As described in Section III-A, detected lesions are segmented and inner and diffusion lesion regions are identified as exemplified in Fig. 3. Benign lesions tend to have well-defined outer lesion borders (yellow contour in Fig. 3) and inner lesion borders (blue contour). The diffusion region between inner and outer borders are generally of uniform width, whereas malignant lesions exhibit diffusion regions that are less regular, with non-uniform diffusion widths. The separations between the inner and outer borders are highly variable on malignant lesions as compared to benign lesions.

Let Γ_{outer} and Γ_{inner} denote the outer and inner borders respectively, and $d(p_i, p_j)$ denote the Euclidean distance between any two pixels p_i and p_j. Then define the distance D_i from the ith pixel on the outer border to the inner border as:

$$D_i = min_j(d(p_i, p_j)), \quad p_i \in \Gamma_{outer}, \ p_j \in \Gamma_{inner} \tag{7}$$

The degree of variability of the separation between the inner and outer borders can then be represented as the variance of the distance from the outer border pixels to the inner border:

$$\delta_D = \sqrt{\frac{1}{N_p} \sum_{i=1}^{N_p} (D_i - \mu_D)^2} \tag{8}$$

where N_p is the length of the outer border (in pixels) and where

$$\mu_D = \frac{1}{N_p} \sum_{i=1}^{N_p} D_i. \tag{9}$$

As compared with the border characteristics of benign lesions, the inner and outer borders of malignant lesions tend to follow different paths causing highly variable separations between the inner and outer borders. Thus, δ_D tends to take much larger values on malignant lesions. Again, this measure applies equally well to incomplete lesion objects.

D. Feature Normalization

A total of 57 features (30 color, 20 texture and 7 border) have been described. While these features are highly descriptive of lesion type, the 57 features exhibit very different ranges of values. Thus we apply a process of feature normalization prior to classification by forming z-scores as follows:

$$z_{ij} = \frac{(x_{ij} - \mu_j)/(3\delta_j) + 1}{2} \tag{10}$$

where x_{ij} is the value of the jth feature of the ith sample, and μ_j and δ_j are the mean and standard deviation of the jth feature, respectively. Using (10), most of the z_{ij} values are forced into the range [0,1], while out-of-range values are clamped to either 0 or 1.

E. Feature Dimensionality Reduction

A careful process of feature reduction can eliminate redundant, irrelevant and noisy features while also improving classification performance. For example, in [13], Celebi *et al.* extracted 437 original features, then selected 18 optimal features from among them using the correlation based feature selection (CFS) method to increase lesion classification accuracy. However, this method does not completely remove redundancies. Principal components analysis (PCA) is a popular technique for dimensionality reduction. Given a set of data on n dimensions, PCA aims to find a linear subspace of dimension d lower than n such that the data points lie mainly on this linear subspace [26]. Here, feature reduction is accomplished by using PCA to filter out redundant, irrelevant and noisy features. The details are described in Section V.

IV. CLASSIFIER DESIGN BASED ON NEURAL NETWORK ENSEMBLE

Artificial nueral networks [11], SVMs [13], Adaboost [14] and KNN [12] have been widely used for lesion classification. Many factors influence classification performance, such as parameter settings, extracted features or feature combinations, and the quality of the experimental samples. On highly representative benign or malignant lesions, correct classification is easily achieved, yet on non-representative lesions, incorrect classifications often happen. Therefore, the difficult task is to achieve correct classification of non-representative lesions. An artificial neural network ensemble [27] is a composite model of multiple neural networks, with better generalization ability and stability than a single network. Bukhtoyarov *et al.* [31] improved classification performance by using a network ensemble model to provide "highly confident" classification results when predictions from a single network classifier were "not confident". Generally, network ensemble models achieve more reliable classification on non-representative objects. Here, we describe a neural network ensemble classifier that deploys a variety of individual networks to achieve more accurate lesion classification results.
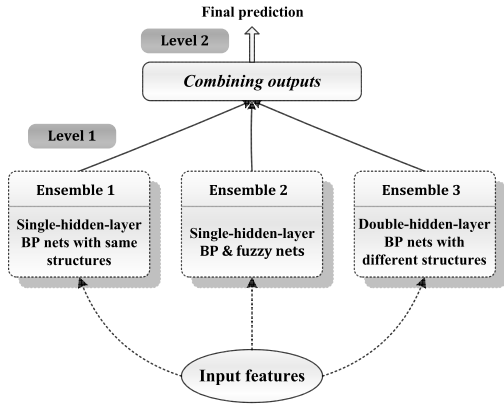
Fig. 6. Meta-ensemble model of multiple neural network ensembles.

A. Neural Network Ensemble

In a neural network ensemble, each network individual is initially trained using the same training sample. Then, the outputs of the individual networks are weighted to form the overall output of the ensemble:

$$y(\mathbf{x}, \mathbf{a}) = \sum_{i=1}^{p} a_i y_i(\mathbf{x}) \tag{11}$$

where $\mathbf{x}$ is the input feature vector and a_i is the weight applied to the output of the ith individual network.

Neural network ensembles are effective only when the individual nets make independent errors [27], [29]. A variety of algorithms have been developed to train network ensembles to achieve better generalization capability, which can be broadly classified according to the type of initial conditions, training data, network architecture and individual network types. In an ensemble, the network architecture and network type are completed in the model design stage, while the initial conditions and training data are developed in the generation stage of the individual nets.

B. Proposed Neural Network Meta-Ensemble Model

In the ensemble model design, varying the network structure is a very common approach [30], [31]. The individual nets in an ensemble are usually feed forward networks. Although using multiple, diverse types of networks is generally regarded as important to achieve the generalization ability of an ensemble, the individual nets are usually of the same network type.

A back propagation (BP) neural network is a kind of multilayer forward network, while a fuzzy neural network (FNN) is an inference system that combines fuzzy logic with an artificial neural network. Both can effectively handle uncertain, nonlinear, and other ill-posed problems. Here, we combine BP networks with FNNs to create the ensemble model, and input different features into the BP networks and FNNs. In addition, in order to further improve the performance of the ensemble, we design a meta-ensemble model, which is composed of three ensembles having different network structures/types, as shown in Fig. 6.

1) Ensemble 1:
A lesion is classified as either benign or malignant, which is a two-class problem. Ensemble 1 is designed to be a combination of single-hidden-layer BP networks, where all the BP individuals have the identical structure N-N-1 (N is the dimension of the input feature vector).

The individual nets need to be generated, and a standard boosting method [32], [33] is used here. Boosting generates a series of component networks whose training data sets are determined by the performance of the former networks. Training instances that are wrongly predicted by the former networks will play more important roles in the training of the later networks. In order to obtain diverse individuals, the initial parameters are randomly generated.

2) Ensemble 2:
Although using diverse types of networks is regarded as important to achieve individual diversity in an ensemble, the individual nets are usually of the same network type. In our model, ensemble 2 is designed as a combination of BP networks and FNNs. As with ensemble 1, the BPs in ensemble 2 have the N-N-1 structure, and each individual is generated using the boosting method. A standard FNN consists of 5 layers. By using the reduced feature set with four principle components described in section III-E, the FNN obtains a classification accuracy close to that obtained using all 57 original features. Therefore, to increase the difference between BP networks and FNNs, only the first four principle components are chosen as input to the FNN individuals in the ensemble, and their structures are 4-12-81-81-1. Similarly, the fuzzy net individuals are obtained by the boosting method. It is obvious that the designed ensemble 2 improves the individual diversity by using different network types.

3) Ensemble 3:
Using a different network structure is a common way to increase the generalization ability of an ensemble. In ensemble 3, double-hidden-layer BP networks with the N-H_1-H_2-1 structure are employed, where H_1 and H_2 are the number of units in the two hidden layers, respectively. For each BP individual, H_1 and H_2 are determined as follows:
- Given Q samples, randomly select k samples as the training set;
- While varying H_1 from $N/3 + 1$ to N and H_2 from 0 to $N/3$, train the network on the k samples and calculate the prediction accuracy using the remaining $Q - k$ samples. The values of H_1 and H_2 yielding the highest accuracy are taken as the optimal ones.

Since H_2 ranges from 0 to $N/3$, the individuals with one hidden layer will exist in ensemble 3 when H_2 is equal to 0. In addition, because the individuals have been trained when determining parameters H_1 and H_2, the boosting method is not needed here.

There are two main design strategies for neural network ensemble models [34]: the direct strategy and the overproduce and choose strategy. In our model, ensembles 1 and 2 belong to the direct strategy, while ensemble 3 is better described as belonging to the overproduce and choose strategy. Therefore, our designed meta-ensemble model may be regarded as a combination of these two strategies. By integrating the three ensembles with different structures and network types (see Fig. 6), the generalization ability of the

designed meta-ensemble model can be improved in regards to four aspects: initial conditions, training data, network architecture and individual network types.

C. Ensemble Size and Comprehensive Prediction

Many authors [27], [35] have discussed the importance of ensemble size, and most of them determined the optimal size through experiments. In [36] and [37], ranges of 5 to 11 individuals were suggested. In [16], Adaboost was used to classify melanomas, and the best result was achieved with combinations of only two to five weak classifiers. In this paper, the size of each ensemble in the proposed model is set to 5. For ensemble 2, because only 4 principle components are fed to the fuzzy networks, which is smaller than the feature set fed to the BP individuals, the number of BP individuals and FNN individuals are 3 and 2 respectively.

There are two levels of ensemble outputs in the proposed model shown in Fig. 6. The first level is the output of single ensembles, and the second level is the integrated output of all three ensembles. Two commonly used methods to pool or combine the multiple outputs are *voting* and *averaging*. Some studies [33], [38], [39] suggest that *voting* delivers better classification performance, while *averaging* is more suitable for regression. Here, *voting* is employed within ensembles 1 and 3 because their individual nets are of the same type, while *weighted averaging* is used for ensemble 2 since its individual nets are of two different types. Finally, *voting* is used to form the combined output of the meta-ensemble model.

V. EXPERIMENTAL RESULTS AND ANALYSIS

We deploy two dermoscopy datasets in our experiments: a xanthous race dataset and a caucasian race dataset. The xanthous race dataset includes 240 dermoscopy images (80 malignant and 160 benign), of which 90 are of incomplete lesion objects (49 are originally incomplete, while 41 others were rendered incomplete by random cuts). All of these images were supplied by the General Hospital of the Air Force of the Chinese Peoples Liberation Army and each image was classified by an experienced dermatologist (Rusong Meng). Of the 240 images in the xanthous race dataset, 174 images were completely segmented via SGNN, and 66 images were segmented by SGNN combined with manual interaction. The caucasian race dataset includes 360 dermoscopy images (120 malignant and 240 benign), in which there are 80 images of incomplete lesion objects (67 originally incomplete, 13 rendered incomplete by random cuts). These images were collected from the PH^2 [40] and EDRA datasets [5]. Of the 360 images, 298 were automatically segmented by SGNN, while 62 were segmented by SGNN combined with manual interaction.

In order to evaluate the effectiveness of the proposed method, the experiments were performed in regards to assessing the performance dependence on four aspects of the model: novel border features, feature reduction, the generalization ability of the designed meta-ensemble model, and the performance of the proposed classification framework. Three metrics including sensitivity, specificity and accuracy are used.

TABLE I
FOUR BORDER FEATURE SETS USED FOR COMPARISON

Featureset	Dimension	Reference	Detail
FeatureSet1	8	[12]	Minimum, maximum, average, and variance of the normalized intensity and the gradient values in the border band.
FeatureSet2	12	[24]	Mean and variance of the gradient magnitude along lesion border as well as over 8 symmetric regions around the border in three separate channels: color, texture and local skin darkness.
FeatureSet3	5	[13]	Convex-hull.
		[41]	Mean and deviation of the gradient magnitude for the entire border band.
		[42]	Fractal dimensions of the border line and the border band.
FeatureSet4	7		Proposed border features.

Sensitivity is the probability of correct detection of disease, while specificity is the probability that a benign lesion would not be diagnosed as malignant, and accuracy defines the ratio of the number of correctly classified samples to the total number of cases. The higher the values of the three metrics, the better the classification performance of the algorithm.

A. Experiment 1: Effectiveness of Border Features

In the proposed classification framework, the new border features are combined with 50 color and texture features to separate MM from benign lesions. In order to evaluate the proposed border features, we compared them with other border features extracted in [12], [13], [24], [41], and [42], as shown in Table I. In these other studies, shape/geometric features were extracted, which means that their border features were calculated on the complete lesions. In order to avoid unfair comparisons, invalid border areas are excluded when extracting border features from incomplete lesions using the compared methods.

We evaluated the classification efficiency of the different border features on the xanthous race dataset, of which 90 images have incomplete lesions and 150 have complete lesions. The classifier used was a single BP network with the N-N-1 structure, with expected error=0.03, learning rate=0.7, max epochs=1000, and where the activation function is a sigmoid function. To validate performance, 10 times 10-fold cross-validation was employed. Figure 7 shows the receiver operating characteristic (ROC) curves of the different border feature sets for all 240 images in the xanthous race dataset, where the left one only used border features and the right one used border features combined with 50 color and texture (denote by C&T) features. The closer to the top-left corner the ROC curve, the better the method. It can be seen that, when using only border features, our border feature vector greatly outperforms the compared border feature vectors. When combined with 50 C&T features, the advantage of our border feature vector is reduced. The AUC or area under the ROC curve, when high, indicates good algorithm performance. Table II shows the AUC values of different feature vectors.
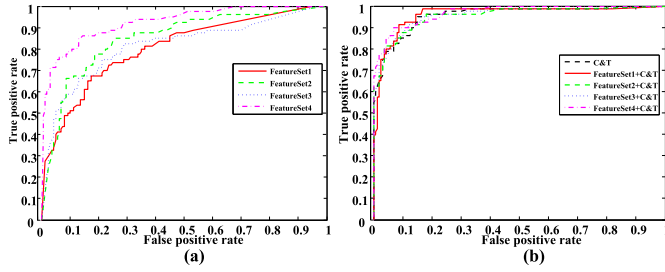
Fig. 7. ROC curves of different border features on xanthous race dataset. (a) using only border features; (b) border features combined with color&texture features.

TABLE II
AUC VALUES USING DIFFERENT BORDER FEATURES
ON XANTHOUS RACE DATASET

	C&T	Set1	Set2	Set3	Set4
Only border features	—	0.7484	0.8434	0.7771	0.9152
Combined with C&T	0.9559	0.9610	0.9548	0.9568	0.9698

Note: Set$n$ stands for FeatureSet$n$.

TABLE III
CLASSIFICATION ACCURACY USING DIFFERENT BORDER
FEATURES ON XANTHOUS RACE DATASET

Feature Set	Incomplete lesions(%)	Complete lesions(%)	All the lesions(%)
FeatureSet1	73.33	78.67	76.67
FeatureSet2	70.00	86.00	80.00
FeatureSet3	73.33	82.67	79.17
FeatureSet4	**84.44**	**88.00**	**86.67**
C&T	82.22	92.67	88.75
FeatureSet1+C&T	86.67	91.33	89.58
FeatureSet2+C&T	83.33	**94.67**	90.42
FeatureSet3+C&T	83.33	93.33	89.58
FeatureSet4+C&T	**87.78**	94.00	**91.67**

It can be seen that our border feature vector achieves the highest AUC values for only border features or for border features combined with 50 C&T features.

Table III gives the classification accuracy statistics for incomplete lesions, complete lesions and all the lesions (including incomplete lesions and complete lesions). It can be seen that incomplete lesions indeed decreased the classification performance. When classifying the lesions using only border features, our features achieved the best classification accuracy for both incomplete lesions and complete lesions. The classification results using only 50-D C&T features are also given in Table III. It can be seen that, when combining C&T features with these border features, the classification accuracy is improved, and the increase from using our border features is the most obvious. Therefore, our border features outperform the other border features.

Table IV shows the classification results using different feature sets on the caucasian race dataset. In the caucasian dataset, there are 80 images with incomplete lesions, and these generally have more severe border loss than the incomplete lesions in the xanthous dataset. It can be seen that FeatureSet 1 and FeatureSet 2, which are based on the normalized intensity or gradient values of the lesion border, were easily influenced by lesion border loss, and when lesion borders were severely

TABLE IV
CLASSIFICATION ACCURACY USING DIFFERENT BORDER
FEATURES ON CAUCASIAN RACE DATASET

Feature Set	Incomplete lesions(%)	Complete lesions(%)	All the lesions(%)
FeatureSet1	63.75	81.79	77.78
FeatureSet2	31.25	68.57	60.28
FeatureSet3	62.50	79.29	75.56
FeatureSet4	**73.75**	**85.71**	**83.06**
C&T	80.00	87.14	85.56
FeatureSet1+C&T	77.50	89.64	86.94
FeatureSet2+C&T	78.75	86.79	85.00
FeatureSet3+C&T	82.50	87.50	86.39
FeatureSet4+C&T	**86.25**	**90.00**	**89.17**

lost, combining them reduced the classification accuracy of the 50-D C&T features on the incomplete lesions. However, the features in FeatureSet 3 and FeatureSet 4 are based on fractal dimension, the convex hull or the difference between inner and outer borders, which are more robust. When combined with the 50-D C&T features, the classification accuracy was increased on both the incomplete lesions and the complete lesions. Among the four feature sets, our border feature set achieved the highest overall accuracy on the lesions in the caucasian dataset. Our border features efficiently described the lesion border irregularities.

B. Experiment 2: Effectiveness of Feature Dimensionality Reduction Using PCA

Our model deploys 57 features to differentiate malignant from benign skin lesions. Since not all features are equally valuable for the classification task, PCA was employed to remove feature redundancy and noise and to improve the classification accuracy. In this experiment, we studied the relative efficiency of using all 57 original features against using the reduced feature sets obtained via the PCA method, and selected the optimal reduced feature set for the subsequent classification. The classifier used was a single BP network, with the same parameters settings as in Section V-A. For each of the two datasets including the xanthous race and the caucasian race, 1/2 of the images were used as training samples to determine the optimal reduced feature set, and the remaining 1/2 of the images were used as test samples to verify the classifier performance in Sections V-C and V-D.

To obtain a reasonable way to reduce the dimensionality, a series of experiments was carried out by varying the number of principle components (NPCs) or the accumulated variance contribution rate (AVCR). Again, 10 times 10-fold cross-validation was used. Table V presents the classification results on the xanthous race dataset.

It can be seen that the two feature sets with 12 principle components (AVCR of 96.07%) and 15 principle components (AVCR of 97.46%) obtained the best accuracy among the reduced feature sets and the original 57-D feature vector. The reduced feature set with 15 principle components achieved a balance between sensitivity and specificity, with better performance than with 12 principle components. Therefore, the reduced feature set with 15 principle components was

TABLE V
CLASSIFICATION RESULTS OF REDUCED FEATURE SETS WITH
DIFFERENT DIMENSIONS AND 57-D FEATURE VECTOR
ON XANTHOUS RACE DATASET

NPCs(AVCR%)	$sen(\%)$	$spe(\%)$	$acc(\%)$
4 (80.01)	77.50	92.50	87.50
5 (84.58)	85.00	92.50	90.00
6 (88.07)	77.50	96.25	90.00
7 (90.06)	82.50	95.00	90.83
8 (91.61)	80.00	93.75	89.17
9 (93.12)	82.50	92.50	89.17
10 (94.47)	82.50	91.25	88.33
11 (95.38)	85.00	92.50	90.00
12 (96.07)	85.00	96.25	**92.50**
13 (96.60)	87.50	91.25	90.00
15 (97.46)	90.00	93.75	**92.50**
18 (98.36)	80.00	91.25	87.50
22 (99.09)	87.50	93.75	91.67
50 (99.99)	87.50	92.50	90.83
57-D feature	80.00	93.75	89.17

TABLE VI
CLASSIFICATION RESULTS OF DIFFERENT ENSEMBLES
ON XANTHOUS RACE DATASET

Ensemble	$sen(\%)$	$spe(\%)$	$acc(\%)$
Single BP	90.00	90.00	90.00
Ensemble1	90.00	93.75	92.50
Ensemble2	92.50	91.25	91.67
Ensemble3	92.50	92.50	92.50
Proposed	95.00	93.75	94.17

TABLE VII
CLASSIFICATION RESULTS OF DIFFERENT ENSEMBLES
ON CAUCASIAN RACE DATASET

Ensemble	$sen(\%)$	$spe(\%)$	$acc(\%)$
Single BP	75.00	93.33	87.22
Ensemble1	80.00	92.50	88.33
Ensemble2	80.00	94.17	89.45
Ensemble3	76.67	95.00	88.89
Proposed	83.33	95.00	91.11

adopted as the final input of the BP individuals in the designed classifier for the xanthous race dataset.

The accuracy value in the last line of Table V is different from that in the last line of Table III (both used the single BP classifier with the N-N-1 structure) because their experimental images are different.

As for the caucasian race dataset, the same experimental protocol as on the xanthous race dataset was adopted, and the reduced feature set with 22 principle components obtained the best classification result. The caucasian race dataset has different dimensions of the optimal reduced feature set than the xanthous race dataset, perhaps caused by different skin colors, different lesion characteristics, and even different image datasets.

C. Experiment 3: Generalization Ability of Different Ensembles

An artificial neural network ensemble has better generalization ability and stability than a single network. Here, we combined BP networks with FNNs, yielding a BP+fuzzy network meta-ensemble model. In a neural network ensemble, weak classifier individuals are permitted to increase individual diversity, therefore for BP individuals in the proposed model, the expected error was set to 0.05, which is looser than on the single BP classifier used in Sections V-A and V-B. In ensemble 3 of our model, BP individuals have two hidden layers and the numbers of the two hidden layers must be determined. Following the method described in Section IV-B, for each of the BP individuals, 70 percent of the training samples were randomly selected to train it while the remaining 30 percent were used to determine the optimal parameter H_1 and H_2.

Unlike traditional neural network ensemble models which have only one level of prediction, our designed meta-ensemble model, with three ensembles of different network structures/types, has two level of prediction. In order to verify the generalization ability of the proposed meta-ensemble model, the classification performances of the single BP, the three ensembles in our model, and the final meta-ensemble model, were compared on the test images. Table VI and Table VII

give the classification results using the optimal reduced feature sets for the xanthous race dataset and the caucasian race dataset, respectively. In Table VI, the classification result for the single BP with accuracy of 90% is different from the result obtained with the optimal reduced feature set with 15 principle components shown in Table V (which also used a single BP), because of the different experimental images used. From Tables VI and VII, it can be seen that all three ensembles improved the classification accuracy relative to single BP to some degree, although one of the specificity metrics decreased. When integrating the three ensembles together (the proposed classifier model), the classification performance was improved significantly relative to the single BP network. Therefore, the designed meta-ensemble model achieved better generalization ability than the three single ensembles.

D. Experiment 4: Performance Analysis of Lesion Classification

In this experiment, our meta-ensemble classifier model was compared with six common classifier models, including Random forests, FNN, KNN, structural SVM [43], SVM with the radial basis kernel function (RBF), and Gentle Adaboost [44], to verify the performance of the proposed classifier model. At the same time, two lesion classification systems using the BoF model [16], [17] were also tested to verify the performance of our lesion classification framework. All of these classifier models, including our model and the compared six common models were carried out using the optimal reduced feature sets on the two datasets. For the KNN classifier, we used the K = 3 nearest neighbors using the Euclidean distance. For the SVM classifier, the LIBSVM software package [45] was used. With the training images, the optimal parameters C and r obtained for the SVM with RBF were 2048 and 0.0313 for the xanthous race dataset, and 64 and 0.25 for the caucasian race dataset, respectively, using the grid search method.

Table VIII gives the classification results of the nine methods (our method, six common classifier methods, and two reference methods [16], [17] on the xanthous race dataset.

TABLE VIII
CLASSIFICATION RESULTS USING DIFFERENT CLASSIFIER
MODELS ON XANTHOUS RACE DATASET

Classifiers	$sen(\%)$	$spe(\%)$	$acc(\%)$
FNN	82.50	95.00	90.83
Random forests	80.00	**96.25**	90.83
KNN	77.50	**96.25**	90.00
Adaboost	80.00	92.50	88.33
RBF SVM	87.50	93.75	91.67
Structural SVM[43]	85.00	91.25	89.17
[16]	65.00	95.00	85.00
[17]	70.00	**96.25**	87.50
Proposed classifier	**95.00**	93.75	**94.17**

TABLE IX
CLASSIFICATION RESULTS USING DIFFERENT CLASSIFIER
MODELS ON CAUCASIANS RACE DATASET

Classifiers	$sen(\%)$	$spe(\%)$	$acc(\%)$
FNN	70.00	95.83	87.22
Random forests	73.33	95.83	88.33
KNN	51.67	**97.50**	82.22
Adaboost	71.67	95.83	87.78
RBF SVM	75.00	94.17	87.78
Structural SVM[43]	75.00	93.33	87.22
[16]	63.33	84.17	77.22
[17]	73.33	89.17	83.89
Proposed classifier	**83.33**	95.00	**91.11**

It can be seen that Random forests, KNN and the method in [17] delivered the highest specificity but their sensitivity was low, while our method obtained the best sensitivity of 95.00%, with an accuracy of 94.17%. The incorrect classification of a melanoma is regarded to be the gravest error, and in [16], the cost of an incorrectly classified melanoma was set to 1.5 times that of an incorrectly classified non-melanoma. With regards to sensitivity, the performance of our model was at least 7.5% higher than that of the compared classifier models. Therefore, with the highest accuracy and the best balance between sensitivity and specificity, our meta-ensemble model greatly outperforms the compared classifier models. In [16] and [17] where features are extracted on patch regions, the models are able to cope with incomplete lesion objects. In our method, new border features are proposed on which an meta-ensemble classifier model is designed. As may be seen in Table VIII, when compared with the systems in [16] and [17], by using our method the sensitivity and accuracy were greatly improved, which is a very positive outcome.

Table IX shows the classification results of the nine methods on the caucasian race dataset. It can be seen that, although the specificity of our method is moderate among these methods, it is still a high value, in addition, the sensitivity increases at least 8.33% more than other methods. With the best accuracy, our method is superior to the compared methods. Therefore, our proposed border features and designed classifier are also highly effective on the caucasian race dataset.

VI. DISCUSSION

In this paper, a framework was proposed for lesion classification, where new border features were proposed and a meta-ensemble model was designed. Regarding the proposed method, we have the following observations:

1) Although the proposed border features are able to describe border irregularities more efficiently than the compared methods, incomplete lesions still present a greater risk of incorrect classification than complete lesions. When a border irregularity is well preserved, this risk is lowered. Correct lesion classification depends on accurately obtaining a combination of features descriptive of color variations, texture patterns, shape asymmetries and border irregularities, following the ABCD rule. When the borders of a lesion are severely lost, local features are a good choice, as for example are used in the BoF model. BoF did not yield good performance in our experiment, likely because the basic features used in the BoF model are all low-level. Deep learning methods could be used to mine high-level features and could be expected to improve lesion classification performance dramatically, given a sufficiently large amount of data.

2) There are two main design strategies for neural network ensemble models: the direct strategy and the overproduce and choose strategy. Our designed meta-ensemble model is a combination of these two. Our main goal was to find a good border irregularity description method and to design a suitable classifier model for lesion classification. Therefore, in the designed meta-ensemble model, some parameters were set based on the existing literature. Our designed model is complex relative to some other common classifier models, such as standard BP, Adaboost, KNN and SVM, and requires more compute time and storage space. For example, the complexity of our proposed model is about 15 times of that using a single BP network (there are 15 net individuals in the proposed model).

VII. CONCLUSION

We have described a novel method for classifying skin lesion objects as malignant or benign. We began by identifying 57 descriptive features, including 50 color and texture features and 7 novel lesion border features. In clinical practice, images containing large and incomplete lesion objects are often obtained by dermoscopy, resulting in the failure of systems that rely on common shape features. The proposed border features described here were designed to be insensitive to the incompleteness of lesion objects. Feature dimensionality reduction was used to eliminate less relevant or noisy features, thereby improving classification performance. PCA was used to reduce the feature dimensions and to select the optimal feature set. An artificial neural network ensemble can be used to cope with this difficult classification problem in a robust and efficient way. A neural network meta-ensemble model was designed, by combining BP neural networks with fuzzy neural networks to increase individual net diversity. In the experiments, feature extraction and reduction were verified and classification performance was tested using FNN, Random forests, KNN, Gentle Adaboost, two SVM methods, and two systems using the BoF model, and the proposed meta-ensemble model on two datasets that respectively include xanthous race data and caucasian race data. The experimental results strongly suggest that the proposed lesion border features

are particularly beneficial for differentiating malignant from benign skin lesions. The classification results delivered by the designed model were shown to be more accurate than those by the compared methods.

REFERENCES

[1] G. Burg, *Das Melanom* (Serie Gesundheit). Piper/VCH, 1993.

[2] P. Schmid-Saugeon, J. Guillod, and J.-P. Thiran, "Towards a computer-aided diagnosis system for pigmented skin lesions," *Comput. Med. Imag. Graph.*, vol. 27, no. 1, pp. 65–78, 2003.

[3] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2015," *CA, Cancer J. Clin.*, vol. 65, no. 1, pp. 5–26, 2015.

[4] F. R. Liu, *Practice of Dermatology*, 3rd ed. Beijing, China: Medical Pub. House, 2005.

[5] G. Argenziano *et al.*, "Interactive atlas of dermoscopy," in *Proc. EDRA Med. Pub.*, 2004. [Online]. Available: http://www.dermoscopy.org

[6] M. Binder *et al.*, "A useful tool for the diagnosis of pigmented skin lesions for formally trained dermatologists," *Arch. Dermatol.*, vol. 131, no. 3, pp. 286–291, 1995.

[7] T. Schindewolf *et al.*, "Classification of melanocytic lesions with color and texture analysis using digital image processing," *Am. Soc. Cytol.*, vol. 15, no. 1, pp. 1–11, 1993.

[8] N. Cascinelli *et al.*, "A possible new tool for clinical diagnosis of melanoma: The computer," *J. Am. Acad. Dermatol.*, vol. 16, no. 2, pp. 361–367, 1987.

[9] W. Stolz *et al.*, "ABCD rule of dermatoscopy—A new practical method for early recognition of malignant-melanoma," *Eur. J. Dermatol.*, vol. 4, no. 7, pp. 521–527, 1994.

[10] X. Yuan, Z. Yang, G. Zouridakis, and N. Mullani, "SVM-based texture classification and application to early melanoma detection," in *Proc. IEEE 28th Annu. Int. Conf. Eng. Med. Biol. Soc. (EMBS)*, Aug. 2006, pp. 4775–4778.

[11] B. Kusumoputro and A. Ariyanto, "Neural network diagnosis of malignant skin cancers using principal component analysis as a preprocessor," in *Proc. IEEE World Congr. Comput. Intell. Neural Netw.*, May 1998, pp. 310–315.

[12] H. Ganster, P. Pinz, R. Rohrer, E. Wildling, M. Binder, and H. Kittler, "Automated melanoma recognition," *IEEE Trans. Med. Imag.*, vol. 20, no. 3, pp. 233–239, Mar. 2001.

[13] M. E. Celebi *et al.*, "A methodological approach to the classification of dermoscopy images," *Comput. Med. Imag. Graph.*, vol. 31, no. 6, pp. 362–373, 2007.

[14] G. Capdehourat *et al.*, "Toward a combined tool to assist dermatologists in melanoma detection from dermoscopic images of pigmented skin lesions," *Pattern Recognit. Lett.*, vol. 32, no. 16, pp. 2187–2196, Dec. 2011.

[15] N. Situ, X. Yuan, G. Chen, and J. Zouridakis, "Malignant melanoma detection by bag-of-features classification," in *Proc. 30th IEEE EMBS Annu. Int. Conf.*, Aug. 2008, pp. 3110–3113.

[16] C. Barata, M. Ruela, M. Francisco, T. Mendonça, and J. S. Marques, "Two systems for the detection of melanomas in dermoscopy images using texture and color features," *IEEE Syst. J.*, vol. 8, no. 3, pp. 965–979, Sep. 2014.

[17] C. Barata, M. E. Celebi, and J. S. Marques, "Improving dermoscopy image classification using color constancy," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 3, pp. 1146–1152, May 2015.

[18] F.-Y. Xie, S.-Y. Qin, Z.-G. Jiang, and R.-S. Meng, "PDE-based unsupervised repair of hair-occluded information in dermoscopy images of melanoma," *Comput. Med. Imag. Graph.*, vol. 33, no. 4, pp. 275–282, 2009.

[19] F. Y. Xie and A. C. Bovik, "Automatic segmentation of dermoscopy images using self-generating neural networks seeded by genetic algorithm," *Pattern Recognit.*, vol. 46, no. 3, pp. 1012–1019, Mar. 2013.

[20] N. Otsu, "An automatic threshold selection method based on discriminate and least squares criteria," *Automatica*, vol. 63, pp. 349–356, Jun. 1979.

[21] M. E. Celebi *et al.*, "Border detection in dermoscopy images using statistical region merging," *Skin Res. Technol.*, vol. 14, no. 3, pp. 347–353, Aug. 2008.

[22] K. Korotkov and R. Garcia, "Computerized analysis of pigmented skin lesions: A review," *Artif. Intell. Med.*, vol. 56, no. 2, pp. 69–90, 2012.

[23] M. Hintz-Madsen *et al.*, "A probabilistic neural network framework for detection of malignant melanoma," in *Artificial Neural Networks in Cancer Diagnosis, Prognosis, and Patient Management*, vol. 5. 2001, pp. 3262–3266.

[24] P. G. Cavalcanti and J. Scharcanski, "Automated prescreening of pigmented skin lesions using standard cameras," *Comput. Med. Imag. Graph.*, vol. 35, no. 6, pp. 481–491, 2011.

[25] D. A. Clausi, "An analysis of co-occurrence texture statistics as a function of grey level quantization," *Can. J. Remote Sens.*, vol. 28, no. 1, pp. 45–62, 2002.

[26] A. Ghodsi, "Dimensionality reduction a short tutorial," Dept. Statist. Actuarial Sci., Univ. Waterloo, Waterloo, ON, Canada, Tech. Rep., 2006.

[27] L. K. Hansen and P. Salamon, "Neural network ensembles," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 10, pp. 993–1001, Oct. 1990.

[28] V. Bukhtoyarov and E. Semenkin, "Neural networks ensemble approach for detecting attacks in computer networks," in *Proc. IEEE World Congr. Comput. Intell.*, Jun. 2012, pp. 10–15.

[29] D. Partridge, "Network generalization differences quantified," *Neural Netw.*, vol. 9, no. 2, pp. 263–271, 1996.

[30] V. V. Bukhtoyarov and O. E. Semenkina, "Comprehensive evolutionary approach for neural network ensemble automatic design," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Jul. 2010, pp. 1–6.

[31] V. Bukhtoyarov and E. Semenkin, "Neural networks ensemble approach for detecting attacks in computer networks," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Jun. 2012, pp. 1–6.

[32] H. Drucker *et al.*, "Boosting and other ensemble methods," *Neural Comput.*, vol. 6, no. 6, pp. 1289–1301, 1994.

[33] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, Aug. 1997.

[34] G. Giacinto, F. Roli, and G. Fumerga, "Unsupervised learning of neural network ensembles for image classification," in *Proc. Neural Netw.*, vol. 3. Jul. 2000, pp. 155–159.

[35] H. L. Yu *et al.*, "A novel neural network ensemble method based on affinity propagation clustering and Lagrange multiplier," in *Proc. Comput. Intell. Softw. Eng.*, 2009, pp. 1–5.

[36] G. P. Zhang, "Neural network ensemble method with jittered training data for time series forecasting," *Inf. Sci.*, vol. 177, no. 23, pp. 5329–5346, 2007.

[37] J. Yang, X. Zeng, S. Zhong, and S. Wu, "Effective neural network ensemble approach for improving generalization performance," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 6, pp. 878–887, Jun. 2013.

[38] D. W. Opitz and J. W. Shavlik, "Actively searching for an effective neural network ensemble," *Connection Sci.*, vol. 8, nos. 3–4, pp. 337–354, 1996.

[39] M. P. Perrone and L. N. Cooper, "When networks disagree: Ensemble methods for hybrid neural networks," in *Proc. Neural Netw. Speech Image Process.*, 1993, pp. 126–142.

[40] T. Mendonça, P. M. Ferreira, J. S. Marques, A. R. S. Marcal, and J. Rozeira, "PH2-A dermoscopic image database for research and benchmarking," in *Proc. 35th Annu. Int. Conf. Eng. Med. Biol. Soc. (EMBC)*, Jul. 2013, pp. 5437–5440.

[41] J. F. Alcon *et al.*, "Automatic imaging system with decision support for inspection of pigmented skin lesions and melanoma diagnosis," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 1, pp. 14–25, Feb. 2009.

[42] J. Blackledge and D. Dubovitskiy, "Object detection and classification with applications to skin cancer screening," *ISAST Trans. Intell. Syst.*, vol. 1, no. 2, pp. 1797–1802, 2008.

[43] I. Tsochantaridis, T. Joachims, T. Hofmann, and Y. Altun, "Large margin methods for structured and interdependent output variables," *J. Mach. Learn. Res.*, vol. 6, pp. 1453–1484, Sep. 2005.

[44] *The Gental Adaboost Package.* [Online]. Available: http://graphics.cs.msu.ru/en/science/research/machinelearning/adaboosttoolbox

[45] *The LIBSVM Package.* [Online]. Available: http://www.csie.ntu.edu.tw/cjlin/libsvm/