Received November 19, 2019, accepted December 20, 2019, date of publication January 6, 2020, date of current version February 14, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2964276

Breast Cancer Image Classification via Multi-Network Features and Dual-Network Orthogonal Low-Rank Learning

YONGJUN WANG[1,2], BAIYING LEI[1,2], (Senior Member, IEEE), AHMED ELAZAB[1,2,3], EE-LENG TAN[4], WEI WANG[1,2], FANGLIN HUANG[1,2], XUEHAO GONG[5], AND TIANFU WANG[1,2]

[1]National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Shenzhen University, Shenzhen 518060, China
[2]Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, Health Science Center, School of Biomedical Engineering, Shenzhen University, Shenzhen 518060, China
[3]Computer Science Department, Misr Higher Institute for Commerce and Computers, Mansoura 35516, Egypt
[4]School of Electric and Electronical Engineering, Nanyang Technological University, Singapore 639798
[5]First Affiliated Hospital of Shenzhen University, Shenzhen University, Second People's Hospital of Shenzhen, Shenzhen 518035, China

Corresponding authors: Xuehao Gong (fox_gxh@sina.com) and Tianfu Wang (tfwang@szu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61871274, Grant 61801305, and Grant 81571758, in part by the Guangdong Pearl River Talents Plan under Grant 2016ZT06S220, in part by the Shenzhen Peacock Plan under Grant KQTD2016053112051497 and Grant KQTD2015033016104926, in part by the Shenzhen Key Basic Research Project under Grant JCYJ20170413161913429, Grant JCYJ20180507184647636, Grant JCYJ2017 0818142347251, Grant JCYJ20170302153337765, Grant CYJ 20170302150411789, Grant JCYJ20170302142515949, Grant GCZX2017040715180580, Grant GJHZ 20180418190529516, and Grant JSGG 20180507183215520, in part by the Sanming Project of Medicine in Shenzhen under Grant SZSM201612027, and in part by the SZU Medical Young Scientists Program under Grant 71201-000001.

ABSTRACT Histopathological image analysis is an important technique for early diagnosis and detection of breast cancer in clinical practice. However, it has limited efficiency and thus the detection of breast cancer is still an open issue in medical image analysis. To improve the early diagnostic accuracy of breast cancer and reduce the workload of doctors, we devise a classification framework based on histology images by combining deep learning with machine learning methodologies in this paper. Specifically, we devise a multi-network feature extraction model by using pre-trained deep convolution neural networks (DCNNs), develop an effective feature dimension reduction method and train an ensemble support vector machine (E-SVM). First, we preprocess the histological images via scale transformation and color enhancement methods. Second, the multi-network features are extracted by using four pre-trained DCNNs (e.g., DenseNet-121, ResNet-50, multi-level InceptionV3, and multi-level VGG-16). Third, a feature selection method via dual-network orthogonal low-rank learning (DOLL) is further developed for performance boosting and overfitting alleviation. Finally, an E-SVM is trained via fused features and voting strategy to perform the classification task, which classifies the images into four classes (i.e., benign, in situ carcinomas, invasive carcinomas, and normal). We evaluate the proposed method on the public ICIAR 2018 Challenge dataset of histology images of breast cancer and achieve a high classification accuracy of 97.70%. Experimental results show that our method can achieve quite promising performance and outperform state-of-the-art methods.

INDEX TERMS Breast cancer image classification, deep convolutional neural network, multi-network features, low-rank learning, ensemble support vector machine.

I. INTRODUCTION

Breast cancer is one of the most common types of cancer and the main leading cause of cancer death among women worldwide [1]. The cornerstone of breast cancer control is early diagnosis, which helps to increase the survival rate of breast cancer. Currently, the early diagnosis of breast cancer is usually performed by biopsy. In clinical practice, biopsy has three main steps. First, the biopsy materials of breast cancer are obtained by drill-biopsy. Second, histopathology images are stained by hematoxylin and eosin (H&E) staining. Third, pathologists perform early diagnosis of breast cancer

The associate editor coordinating the review of this manuscript and approving it for publication was Keivan Navaie.

by observing the histology images. However, the diagnostic performance relies on the doctors' professional skills and experience, which is typically subjective and maybe inconsistent across different pathologists. To reduce these adverse effects improve early diagnostic efficiency, and alleviate the workload burden, the computer-aided diagnosis (CAD) systems are developed [2]–[5] utilizing image analysis methods.

With the latest development of machine learning and deep learning techniques [6]–[8], the CAD systems can potentially offer more reliable classification methods for the histology images of breast cancer [9]–[15]. These methods are mainly learnt models to classify the histology images of breast cancer into two classes (e.g., carcinomas and non-carcinomas) or four classes (e.g., benign, in situ carcinomas, invasive carcinomas, and normal, Fig. 1). As seen from Fig. 1, these histology images have very large size, uneven H&E staining, and the great differences between pathological images from different patients, which have high the intra-class differences and low the inter-class differences [16]. Therefore, the image preprocessing [12] is desirable by utilizing scale transformation and color enhancement [17], where scale transformation (e.g., downscale and randomly cropped image) is used to solve the problem of very large image size and color enhancement is used to address the problem of uneven H&E staining. The image preprocessing helps to narrow the intra-class differences and increase the inter-class differences.

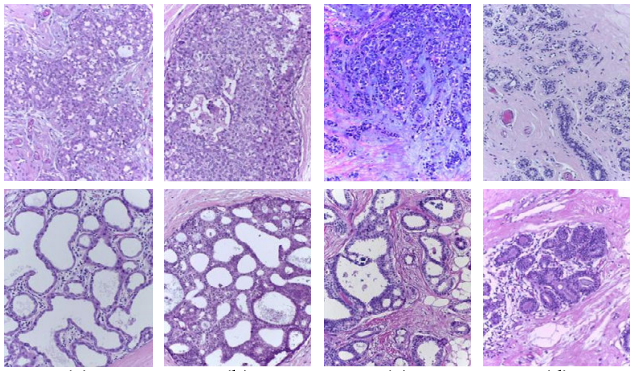


FIGURE 1. Examples of microscopic breast cancer histological images: (a) benign, (b) in situ carcinoma, (c) invasive carcinoma, (d) normal.

It is known that deep convolution neural networks (DCNNs) have been widely used in many image classification tasks with great successes [18]–[21]. The successes are mainly attributed to the powerful learning ability of DCNN, which can obtain more important feature information. However, they usually only employ single-network features, which can hardly get comprehensive image features and have poor generalization ability. Therefore, the multi-network features are obtained by using different DCNNs and different convolution layers, respectively, which can improve performance and enhance generalization ability. As a result, we design a multi-network feature model by using four classical DCNNs (e.g., VGG-16 [18], InceptionV3 [22], ResNet-50 [21], and DenseNet-121 [23]). Accordingly, different networks have unique advantages in their network structures, which allows effective capturing of complemented features.

The VGG-16 network uses a smaller convolution kernel and piecewise convolution to extract the detailed local feature information. In the InceptionV3 network, the mixed modules decompose two-dimensional convolution into two one-dimensional convolutions, which increases the nonlinearity and the width of the network to eliminate representation bottleneck. The ResNet-50 network has a deeper network structure and the residual module [21], which may solve the degradation problem in the optimization process and enhance learning ability. In the DenseNet-121 network, an innovative dense module [23] is proposed, which connects each layer to every other layer in a feed-forward fashion. It relieves the gradient vanishing problem, reinforces feature propagation, encourages feature reuse and substantially cuts down the number of parameters.

To further boost the classification performance, this paper adopts the multi-level InceptionV3 (ML-InceptionV3) network and the multi-level VGG-16 (ML-VGG-16) [12] network by extracting the feature maps of the intermediate layer of InceptionV3 and VGG-16 networks, which are combined with DenseNet-121 and ResNet-50 to obtain multi-network features. Afterward, the features from the preprocessing of scale transformation and color enhancement are encoded via 3-norm pooling [24] and average encoding methods.

For the medical image data with small samples, it is well-known that the DCNN models of end-to-end architectures are prone to overfitting. Additionally, the multi-network features have high feature dimensions, which leads to higher computational costs. To address these issues, feature selection methods are effective ways. However, the existing methods [25]–[27] have limitations since they mainly consider the relation between the features from the same source and the response variables, which fail to consider the relations among response variables and the complementary relations among the features of two different DCNNs (dual-network). Therefore, we propose a dimension reduction method called dual-network orthogonal low-rank learning (DOLL) inspired by the previous work [28]. Our proposed method utilizes joint low-rank learning and orthogonal rotation among dual-network features, which can consider three relations (e.g., the relation among the features and the response variables, the relation among response variables, and the complementary relation between dual-network features). It can effectively remove redundant features and select important feature information. Also, the softmax layer in DCNN is replaced by ensemble support vector machine (E-SVM) classifiers for classification performance boosting.

In summary, we propose a new classification framework to classify breast cancer histological images, which uses multi-network features, DOLL method and E-SVM classifier. The main contributions of this paper are as below.

1) We devise a multi-network feature extraction model to obtain more comprehensive feature representations of breast cancer histological images.

2) We develop a DOLL feature selection method, which considers three relationships to remove the redundant features and obtain complementary features.

3) We train an E-SVM classifier with fused features and voting strategy to improve the classification performance.

The rest of this paper is organized as follows. In Section II, we briefly review related work of the early diagnosis of breast cancer. In Section III, we describe our proposed method in details. The experiments and comparison results are given in Section IV followed by discussions in Section V. Finally, our conclusions are presented in Section VI.

II. RELATED WORK

In this section, we briefly review the development of the early diagnosis of breast cancer using CAD technologies. Currently, in breast cancer early screening, the CAD systems are based on breast mammography [2], [5]. In the clinical diagnosis of breast cancer, the puncture biopsy is still the mainstream method. However, the performance relies on the doctors' professional skills and experience, which is typically subjective. So the diagnostic accuracy is usually unsatisfactory due to many challenges [16] of the breast cancer histology image analysis. Therefore, researchers tried to propose reliable methods to address these drawbacks to improve the efficiency of early diagnosis of breast cancer.

At present, the traditional machine learning methods have been widely used in the image analysis tasks of breast cancer. For instance, Wang *et al.* [15] proposed a double-strategy splitting model with adaptive mathematical morphology and curvature scale space corner detection, which utilized shape and textural features to achieve the classification of cell nuclei in pathological images of breast cancer. This approach could only perform the classification of normal and malignant without further differentiation of specific types of breast cancer. He *et al*. [29] proposed a fusion strategy of heterogeneous features from stacked sparse auto-encoder to boost the accuracy of histopathological image analysis. Their method focused on the rank-level fusion of local and holistic features to assist image-guided diagnosis of breast cancer. Recently, Zheng *et al.* [30] proposed a classification framework of whole slide images of breast cancer. It shares the advantages of both histopathological image classification and content-based histopathological image retrieval. The framework could recognize the malignant regions using a probability map. Sudharshan *et al.* [31] proposed a weakly supervised learning framework of multiple instance learning, without the need to label all the instances. Although these traditional machine learning methods have achieved good results, the diagnosis performance is still unsatisfactory.

To further improve the diagnosis performance, deep learning [7], especially DCNNs, also has been widely used in many medical image analysis tasks [32]. Therefore, many researchers also have studied the classification of breast cancer using deep learning methods [33]–[38]. For example, Han *et al.* [14] proposed a structured deep learning model for automatic multi-classification of breast cancer from histopathological images, and achieved remarkable performance on a large-scale dataset. Similarly, Gandomkar *et al.* [35] carried out the multi-classification of breast cancer using deep residual learning. They implemented classification for benign or cancer, and then categorized cancer and benign cases into four different subtypes. However, these deep learning methods often suffer from overfitting due to the limited training samples.

Unfortunately, unlike the case of natural image classification tasks, there are less labeled images available for effective training of a deep learning network for early diagnose of breast cancer. Therefore, in recent studies, researchers have combined deep learning with traditional machine learning to exploit their individual merits. For instance, Araújo *et al.* [13] extracted the features by the DCNN and used these features to train an SVM classifier to categorize pathological images of breast cancer. Vo *et al.* [11] proposed an incremental boosting DCNN that was strengthened by a stepwise combination of weak and strong classifiers to extract more effective features than traditional machine learning approaches. They achieved superior results in the classification of breast cancer image than using traditional machine learning or deep learning methods alone. Nevertheless, they only used the image features of the single network instead of the multi-network features, which leads to the inability to obtain informative feature representation and reduces classification performance. Hence, we propose a multi-network feature model based on four different DCNNs to extract more comprehensive image features.

It is worth noting that the high-dimensional features will cause the classifier to over fit, high computational costs and affect classification accuracy. To solve these problems, many researchers focused on the research of efficient feature selection algorithm [25], [26], [28]. Commonly used classical feature selection methods include principal component analysis (PCA) [26] and locally linear embedding (LLE) [27]. In addition, Tomioka and Sugiyama [39] developed a dual-augmented Lagrangian method (DALM) to achieve feature selection. Zhu *et al.* [40] proposed a multi-relational regularization (MRR) feature dimensionality reduction method. The MRR method considers three relationships to select the most representative feature information. Nie *et al.* [25] proposed an unsupervised feature selection algorithm. This method performs both feature selection and local structure learning. Moreover, the similarity matrix is constrained to contain more accurate data structure information so that it can select valuable features. However, these methods focus on the features from the same source without considering the complementary relation among dual-network features. For this reason, we propose the DOLL method to reduce the feature dimension.

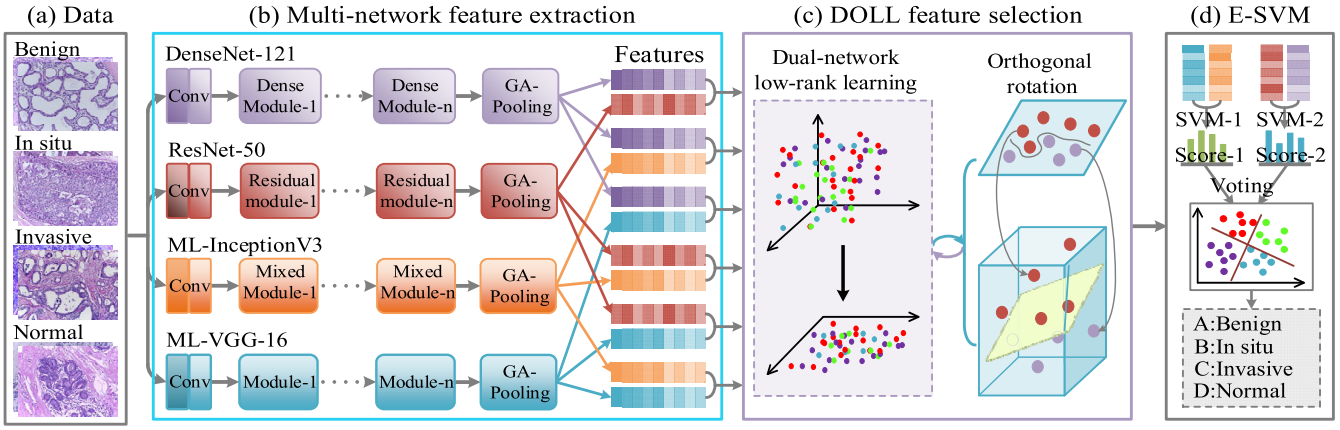


FIGURE 2. Breast cancer pathological image classification framework. (a) Input images; (b) Feature extraction using DenseNet-121, ResNet-50, ML-InceptionV3, and ML-VGG-16; (c) Feature selection via the DOLL method; (d) E-SVM classifier.

III. METHODOLOGY

In this paper, we perform the classification of the breast cancer histopathological images by using multi-network features, the DOLL feature selection method, and E-SVM classifier. Fig. 2 shows the architecture of our proposed classification model, which comprises three steps: feature extraction, feature selection, and classification. In the first step, we first perform image preprocessing by random cropping and color enhancement. Then, we use pre-trained ResNet-50, DenseNet-121, ML-InceptionV3, and ML-VGG16 as feature extractors to obtain multi-network features. These features from different image crops or different color-enhanced images are encoded via utilizing 3-norm pooling and average encoding, respectively. This helps to reduce the effect of very large image size and uneven H&E staining on classification performance. In the second step, we apply the DOLL method to perform feature reduction, which has fast training speed, low computational cost and alleviates classifier overfitting. In the last step, breast cancer pathological images are distinguished using E-SVM classifier. We use dual-network fused features and voting strategy between different dual-network to train E-SVM classifiers, which can effectively improve the classification performance.

A. NOTATIONS

In this paper, uppercase boldface letters represent matrices, lowercase boldface letters represent vectors, and ordinary italics represent scalars. For the matrix $\mathbf{X} = [x_{i,j}]$, its i-th row and j-th column are denoted as $\mathbf{x}^i$ and $\mathbf{x}_j$. The Frobenius norms ℓ_1, ℓ_2, $\ell_{2,1}$, and ℓ_F-norm of a matrix $\mathbf{X}$ are denoted as $\|\mathbf{X}\|_1 = \sum_i |\mathbf{x}^i|$, $\|\mathbf{x}^i\|_2 = \sqrt{\sum_j x_{i,j}^2}$, $\|\mathbf{X}\|_{2,1} = \sum_i \|\mathbf{x}^i\|_2 = \sum_i \sqrt{\sum_j x_{i,j}^2}$, and $\|\mathbf{X}\|_F = \sqrt{\sum_i \|\mathbf{x}^i\|_2^2} = \sqrt{\sum_j \|\mathbf{x}_j\|_2^2}$, respectively. Moreover, $\mathbf{X}^T$, Tr($\mathbf{X}$), rank($\mathbf{X}$), and $\mathbf{X}^{-1}$ denoted as the transpose, trace, rank, and inverse operators of $\mathbf{X}$, respectively.

B. FEATURE EXTRACTION MODEL

In this section, we briefly introduce the image preprocessing and describe how to extract the multi-network features using the four pre-trained DCNNs, and perform the feature encoding, as shown in Fig. 2 (b) and Fig. 3.

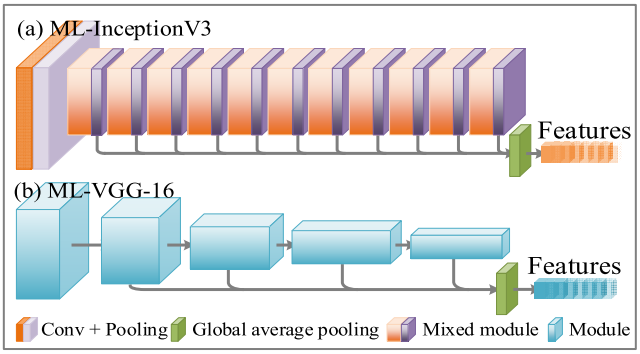


FIGURE 3. Feature extraction models of the ML-InceptionV3 and ML-VGG-16.

1) IMAGE PREPROCESSING

In this study, we first downscale the original images (2048×1536 pixels) by a factor of two, and perform color space conversion [17]. By changing the weight of the RGB color component of the image, many color enhanced images of each original input images are obtained. And encoding the features of color enhanced images can correct the uneven H&E staining, and narrow the intra-class differences. In the next step, the image crops of different sizes are randomly extracted, which can help to obtain local details and global feature information and reduce computation costs. In our experiments, the cropped image sizes are preset to be 400 × 400 pixels or 700 × 700 pixels, the number of images to be cropped is 20, and the number of color-enhanced images is 50.

2) MULTI-NETWORK FEATURE EXTRACTION

The accuracy and robustness of the deep learning methods rely on a large number of training samples. However, the availability of large collections of annotated medical images is rare and a small dataset is insufficient to train a deep learning model with high precision and strong robustness. The over-sized and limited training images may lead to over-fitting of the end-to-end DCNNs. To solve these problems, we use four DCNNs (e.g., DenseNet-121, ResNet-50, ML-InceptionV3, and ML-VGG-16) to extract multi-network features. In this paper, these DCNNs are initialized with ImageNet pre-trained weights [41] to fit our learning task. After that, we feed the cropped images into these pre-trained DCNNs while remove the fully connected layer and softmax layers and use global average pooling to encode the feature map of the channels. For DenseNet-121, we convert the last convolutional layer consisting of 1024 channels via global average pooling into a one-dimensional feature vector with a length of 1024. For ResNet-50, we convert the last convolutional layer consisting of 2048 channels via global average pooling into a one-dimensional feature vector with a length of 2048.

Based on InceptionV3 and VGG-16, we design ML-InceptionV3 network and ML-VGG-16 network to extract image features, and their networks are shown in Fig. 3. For ML-InceptionV3, we apply the global average pooling operation to feature maps of the connection layer in the internal mixed modules of InceptionV3, and then concatenate them into one vector, as shown in Fig. 3 (a). For ML-VGG-16, we apply the global average pooling operation to the four convolutional modules of VGG-16, and concatenate them into a single vector, as shown in Fig. 3 (b). Namely, the motivation of our design is that DenseNet-121 and ResNet-50 are endowed with a deep network structure that utilizes dense and residual modules to obtain the abstract semantic features. ML-InceptionV3 and ML-VGG-16 can extract more superficial local detailed feature information (e.g., shape, texture, color and others). The features of four DCNNs are combined in pairs to obtain the features of six dual-networks; DenseNet-121 and ResNet-50 (D-R), DenseNet-121 and ML-InceptionV3 (D-MI), DenseNet-121 and ML-VGG-16 (D-MV), ResNet-50 and ML-InceptionV3 (R-MI), ResNet-50 and ML-VGG-16 (R-MV), and ML-InceptionV3 and ML-VGG-16 (MI-MV). In the next step, the dimension of dual-network features is reduced by using the DOLL method.

3) FEATURE ENCODING

The features of color-enhanced images and cropped images obtained by each DCNN are encoded by performing 3-norm pooling [24] and average encoding, respectively. Let $\mathbf{E} = \left[\mathbf{e}^1; \mathbf{e}^2; \ldots\ldots; \mathbf{e}^s\right] \in \mathbb{R}^{s\times d}$ be the feature matrix of image crops, where s is the number of image crops, d is the feature dimension. Let $\mathbf{V} = \left[\mathbf{v}^1; \mathbf{v}^2; \ldots\ldots; \mathbf{v}^l\right] \in \mathbb{R}^{l\times d}$ be the feature matrix of color-enhanced images, where l is the number of color-enhanced images. We define 3-norm pooling method as

$$\mathbf{v}^j = \left(\frac{1}{s}\sum\nolimits_{i=1}^{s}\left(\mathbf{e}^i\right)^p\right)^{\frac{1}{p}}, \tag{1}$$

where $(\mathbf{e}^i)^p$ is defined as $(\mathrm{e}_{i,1}^p, \mathrm{e}_{i,2}^p, \ldots \mathrm{e}_{i,j}^p \ldots \mathrm{e}_{i,d}^p)$, p is the hyper-parameter, and it is set to 3 as suggested in [24].

Average encoding is performed by calculating the mean value of feature vectors of l color-enhanced images. Denote $\mathbf{X} = \left[\mathbf{x}^1; \mathbf{x}^2; \ldots\ldots; \mathbf{x}^n\right] \in \mathbb{R}^{n\times d}$ as a feature matrix of input image containing n samples. Average encoding is defined as

$$\mathbf{x}^k = \frac{1}{l}\sum\nolimits_{j}^{l}\mathbf{v}^j. \tag{2}$$

These two methods not only can effectively combine local texture, boundary, and color information of the original image feature information, but also dramatically alleviate the uneven staining issue of pathological images.

C. DOLL FEATURE SELECTION METHOD

Full features are extracted by utilizing different DCNNs and different convolutional layers in the same DCNN. However, these features can cause information redundancy and overfitting and reduce the classification accuracy of breast cancer histological images. Therefore, we propose a DOLL feature selection method.

1) LOW-RANK REGRESSION

Let $\mathbf{X} = \left[\mathbf{x}^1; \mathbf{x}^2; \ldots\ldots; \mathbf{x}^n\right] \in \mathbb{R}^{n\times d}$ be the feature matrix, and $\mathbf{Y} = [\mathbf{y}^1; \mathbf{y}^2; \cdots\cdots; \mathbf{y}^n] \in \mathbb{R}^{n\times c}$ be the label matrix, where n is the total number of samples, d is the feature dimension, and c is the number of breast cancer classes. We denote the vector $\mathbf{y}^i = \left\{\mathrm{y}_{i,1}, \mathrm{y}_{i,2}, \cdots\cdots, \mathrm{y}_{i,c}\right\} \in \{0, 1\}^c$ as a representation of the class label of the i-th sample $\mathbf{x}^i$. In this way, when $\mathbf{x}^i$ belongs to the j-th class, the j-th element in $\mathbf{y}^i$ is set to one (*e.g.*, $\mathrm{y}_{i,j} = 1$), and other elements are set to zero.

The low-rank regression [42], [43] of the measurements between the response variables and the feature variables formula is defined as

$$\mathbf{Y} = \mathbf{X}\mathbf{W} + \mathbf{a}\mathbf{b}, \tag{3}$$

where $\mathbf{W} \in \mathbb{R}^{d\times c}$ represents the coefficient matrix, $\mathbf{b} \in \mathbb{R}^{1\times c}$ represents the bias term, and $\mathbf{a} \in \mathbb{R}^{n\times 1}$ is the column vector of all elements set to ones. Then, we can obtain the solution of $\mathbf{W}$ via least squares method, which is given as

$$\mathbf{W} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\left(\mathbf{Y} - \mathbf{a}\mathbf{b}\right). \tag{4}$$

However, Eq. (4) only considers the relation between the feature variables and the response variables, while the relation among the response variables is ignored. To address this issue, we impose a representation on the rank of $\mathbf{W}$(*i.e.*, $\mathrm{rank}\,(\mathbf{W}) \leq \min(d, c)$, and suppose $\mathrm{rank}\,(\mathbf{W}) = r$). In general, the low-rank representation (LRR) [44] of the data matrix $\mathbf{D}$ can be represented as a linear combination of the

dictionary matrix $\mathbf{A}$ and the regression parameter matrix $\mathbf{Z}$. It aims at finding the lowest-rank representation between all the linear combinations, which is formulated as

$$\min_{\mathbf{Z}} \|\mathbf{Z}\|_*, \quad \text{s.t.} \mathbf{D} = \mathbf{AZ}, \tag{5}$$

where $\|\cdot\|_*$ is the nuclear norm of the matrix, which is the sum of the singular values of the matrix. Therefore, LRR on $\mathbf{W}$ is represented as

$$\mathbf{W} = \mathbf{AZ}, \tag{6}$$

where $\mathbf{A} \in \mathbb{R}^{d\times r}$, $\mathbf{Z} \in \mathbb{R}^{r\times c}$. When r is fixed, we replace $\mathbf{W}$ via $\mathbf{AZ}$ in Eq. (3). Thus, the LRR is redefined as

$$\min_{\mathbf{A},\mathbf{Z},\mathbf{b}} \|\mathbf{Y} - \mathbf{XAZ} - \mathbf{ab}\|_F^2. \tag{7}$$

Based on Eq. (7), we can project the feature matrix $\mathbf{X}$ into an r-dimensional space. In other words, the low-rank regression is viewed as the subspace learning on $\mathbf{X}$. Therefore, when $r \leq d$, LRR is an effective way to achieve dimension reduction. Ultimately, the low-rank regression not only considers the relationship between the feature variables and the response variables, but also the relationship among the response variables. However, the deficiency is that the complementarity relation of the features among the dual-network is not considered.

2) DOLL FEATURE SELECTION

To consider a complementary relationship, we develop a dual-network low-rank learning (DLL). In general, feature dimensions of different DCNNs are unequal. Therefore, the PCA method is first used to reduce feature dimension ($d = \min(d_1, d_2)$), where d_1 and d_2 represent feature dimensions of the dual-network. Let $\mathbf{X}_1 \in \mathbb{R}^{n\times d}$, $\mathbf{X}_2 \in \mathbb{R}^{n\times d}$ be two feature matrices of the dual-network. $\mathbf{W}_1 \in \mathbb{R}^{d\times c}$, $\mathbf{W}_2 \in \mathbb{R}^{d\times c}$ are two coefficient matrices, which are replaced by LRR in Eq. (6). Therefore, we define DLL as

$$\min_{\mathbf{A_1},\mathbf{Z_1},\mathbf{b_1},\mathbf{A_2},\mathbf{Z_2},\mathbf{b_2},} \|\mathbf{Y}-\mathbf{X}_1\mathbf{A}_1\mathbf{Z}_1-\mathbf{ab}_1\|_F^2+\|\mathbf{Y}-\mathbf{X}_2\mathbf{A}_2\mathbf{Z}_2-\mathbf{ab}_2\|_F^2, \tag{8}$$

where $\mathbf{A}_1 \in \mathbb{R}^{d\times r}$ and $\mathbf{A}_2 \in \mathbb{R}^{d\times r}$ represent two low-rank dictionary matrices, $\mathbf{Z}_1 \in \mathbb{R}^{r\times c}$ and $\mathbf{Z}_2 \in \mathbb{R}^{r\times c}$ represent two regression parameter matrices, $\mathbf{b}_2 \in \mathbb{R}^{1\times c}$ and $\mathbf{b}_1 \in \mathbb{R}^{1\times c}$ represent two bias terms, respectively. Hence, the DLL considers three relationships, simultaneously.

The DLL can linearly represent the relation among the response variables utilizing r latent factors obtained from d feature variables of $\mathbf{X}_1$ and $\mathbf{X}_2$. However, the multi-network features contain a large number of redundant features. These redundant features may not be useful in prediction and affect the calculation of r latent factors. In this way, when $\mathbf{X}_1$ and $\mathbf{X}_2$ are mapped into a low-dimensional space, we perform subspace learning and explain response variables for these features to achieve the feature dimension reduction. To this end, we utilize two $\ell_{2,1}$-norm items of the dual-network, respectively. Since the feature matrices mapped to low-dimension space can change the distribution of original feature in high-dimension space, we add the orthogonal constraints and Eq. (8) is reformulated as

$$\begin{aligned}\min_{\mathbf{A_1},\mathbf{Z_1},\mathbf{b_1},\mathbf{A_2},\mathbf{Z_2},\mathbf{b_2}} & \|\mathbf{Y}-\mathbf{X}_1\mathbf{A}_1\mathbf{Z}_1-\mathbf{ab}_1\|_F^2+\alpha_1\|\mathbf{A}_1\|_{2,1} \\ & +\|\mathbf{Y}-\mathbf{X}_2\mathbf{A}_2\mathbf{Z}_2-\mathbf{ab}_2\|_F^2+\alpha_2\|\mathbf{A}_2\|_{2,1}, \\ \text{s.t.,}\ & \mathbf{Z}_1\mathbf{Z}_1^T=\mathbf{I}_r, \mathbf{Z}_2\mathbf{Z}_2^T=\mathbf{I}_r,\end{aligned} \tag{9}$$

where $\mathbf{I}_r \in \mathbb{R}^{r\times r}$, α_1 and α_2 are two tuning parameters. The $\ell_{2,1}$-norm items on $\mathbf{A}_1$ and $\mathbf{A}_2$ penalize the coefficients of $\mathbf{A}_1$ and $\mathbf{A}_2$ in a row-wise manner for joint selection or unselection of the features in predicting the response variables. It is worth noting that, the column-wise low-rank representations and the row-wise $\ell_{2,1}$-norm on $\mathbf{A}_1$ and $\mathbf{A}_2$ have the effects of conducting subspace learning and feature selection on $\mathbf{X}_1$ and $\mathbf{X}_2$, respectively.

To exploit the advantage of the relation among the dual-network features, $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are replaced by a shared regression parameter matrix $\mathbf{Z} \in \mathbb{R}^{r\times c}$ of the dual-network. Thus, Eq. (9) is rewritten as

$$\begin{aligned}\min_{\mathbf{A_1},\mathbf{b_1},\mathbf{A_2},\mathbf{b_2},\mathbf{Z}} & \|\mathbf{Y}-\mathbf{X}_1\mathbf{A}_1\mathbf{Z}-\mathbf{ab}_1\|_F^2+\alpha_1\|\mathbf{A}_1\|_{2,1} \\ & +\|\mathbf{Y}-\mathbf{X}_2\mathbf{A}_2\mathbf{Z}-\mathbf{ab}_2\|_F^2+\alpha_2\|\mathbf{A}_2\|_{2,1}, \\ \text{s.t.,}\ & \mathbf{ZZ}^T=\mathbf{I}_r.\end{aligned} \tag{10}$$

Specifically, the complemented features of the dual-network have different feature representations of breast cancer pathological images. Meanwhile, adding orthogonal rotation $\mathbf{Z}$ can transfer low-dimension feature representations into the original label space spanned by the high-dimension response variable $\mathbf{Y}$. Such an orthogonal rotation step naturally takes advantage of DLL and explores the relation among the dual-network. Therefore, we use the orthogonal rotation to consider the relation of complementarity among dual-network in DLL method. At last, the DOLL is utilized for feature selection by considering three relations to efficiently remove redundant features and select many complementary features to improve the classification accuracy of breast cancer.

D. OPTIMIZATION

This Section describes the optimization process, which determines optimal parameters (e.g., $\mathbf{Z}$, $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{A}_1$, and $\mathbf{A}_2$). Specifically, we iteratively conduct the following two steps until satisfying predefined conditions. (1) Update $\mathbf{Z}$ with fixed $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{A}_1$ and $\mathbf{A}_2$. (2) Update $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{A}_1$ and $\mathbf{A}_2$ with fixed $\mathbf{Z}$.

1) UPDATE $\mathbf{Z}$ WITH FIXED $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{A}_1$ AND $\mathbf{A}_2$

When $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{A}_1$ and $\mathbf{A}_2$ are fixed to optimize $\mathbf{Z}$, Eq. (10) can be rewritten as

$$\begin{aligned}\min_{\mathbf{Z}} & \|\mathbf{Y}-\mathbf{X}_1\mathbf{A}_1\mathbf{Z}-\mathbf{ab}_1\|_F^2+\|\mathbf{Y}-\mathbf{X}_2\mathbf{A}_2\mathbf{Z}-\mathbf{ab}_2\|_F^2, \\ \text{s.t.,}\ & \mathbf{ZZ}^T=\mathbf{I}_r.\end{aligned} \tag{11}$$

After simple mathematical manipulation, Eq. (11) is equivalent to the following formula

$$\min_{\mathbf{Z}} \left\|\mathbf{Y}' - \mathbf{X}'\mathbf{Z}\right\|_F^2, \quad \text{s.t.,}\ \mathbf{Z}\mathbf{Z}^T = \mathbf{I}_r, \tag{12}$$

where $\mathbf{Y}' = \begin{bmatrix} \mathbf{Y}' - \mathbf{a}\mathbf{b}_1 \\ \mathbf{Y}' - \mathbf{a}\mathbf{b}_2 \end{bmatrix} \in \mathbb{R}^{2n\times c}$ and $\mathbf{X}' = \begin{bmatrix} \mathbf{X}_1\mathbf{A}_1 \\ \mathbf{X}_2\mathbf{A}_2 \end{bmatrix} \in \mathbb{R}^{2n\times r}$. The optimization problem of Eq. (12) is actually an orthogonal Procrustes problem [45]. The optimal solution of $\mathbf{Z}$ is $(\mathbf{U}\mathbf{V}^T)^T$, where $\mathbf{U} \in \mathbb{R}^{c\times r}$ and $\mathbf{V} \in \mathbb{R}^{r\times r}$ are obtained from the singular value decomposition of $\mathbf{Y}'^T\mathbf{X}' = \mathbf{U}\mathbf{G}\mathbf{V}^T$, and $\mathbf{G} \in \mathbb{R}^{r\times r}$ is a diagonal matrix.

2) UPDATE $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{A}_1$ AND $\mathbf{A}_2$ WITH FIXED $\mathbf{Z}$

When $\mathbf{Z}$ is fixed to optimize $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{A}_1$ and $\mathbf{A}_2$, Eq. (10) can be rewritten as

$$\begin{aligned}\min_{\mathbf{A}_1,\mathbf{b}_1,\mathbf{A}_2,\mathbf{b}_2} &\|\mathbf{Y}-\mathbf{X}_1\mathbf{A}_1\mathbf{Z} - \mathbf{a}\mathbf{b}_1\|_F^2 + \alpha_1 \|\mathbf{A}_1\|_{2,1} \\ &+ \|\mathbf{Y}-\mathbf{X}_2\mathbf{A}_2\mathbf{Z} - \mathbf{a}\mathbf{b}_2\|_F^2 + \alpha_2 \|\mathbf{A}_2\|_{2,1}.\end{aligned} \tag{13}$$

Taking the partial derivative of $\mathbf{b}_1$ and $\mathbf{b}_2$ in Eq. (10), and then setting the derivative to zero, we have

$$\mathbf{a}^T\mathbf{X}_1\mathbf{A}_1\mathbf{Z} + \mathbf{a}^T\mathbf{a}\mathbf{b}_1 - \mathbf{a}^T\mathbf{Y} = 0, \tag{14}$$

$$\mathbf{a}^T\mathbf{X}_2\mathbf{A}_2\mathbf{Z} + \mathbf{a}^T\mathbf{a}\mathbf{b}_2 - \mathbf{a}^T\mathbf{Y} = 0. \tag{15}$$

After simple mathematical manipulation, we can get optimal solutions for $\mathbf{b}_1$ and $\mathbf{b}_2$

$$\mathbf{b}_1 = \frac{1}{n}(\mathbf{a}^T\mathbf{Y} - \mathbf{a}^T\mathbf{X}_1\mathbf{A}_1\mathbf{Z}), \tag{16}$$

$$\mathbf{b}_2 = \frac{1}{n}\left(\mathbf{a}^T\mathbf{Y} - \mathbf{a}^T\mathbf{X}_2\mathbf{A}_2\mathbf{Z}\right). \tag{17}$$

By replacing $\mathbf{b}_1$ and $\mathbf{b}_2$ in Eq. (13) with Eq. (16) and Eq. (17), we compute the optimal solutions of $\mathbf{A}_1$ and $\mathbf{A}_2$, and define $\mathbf{H} = \mathbf{I}_n - \frac{1}{n}\mathbf{a}\mathbf{a}^T \in \mathbb{R}^{n\times n}$, where $\mathbf{I}_n \in \mathbb{R}^{n\times n}$ is an identity matrix. Eq. (13) can be rewritten as

$$\begin{aligned}\min_{\mathbf{A}_1,\mathbf{A}_2} &\|\mathbf{H}\mathbf{Y}-\mathbf{H}\mathbf{X}_1\mathbf{A}_1\mathbf{Z}\|_F^2 + \alpha_1 \|\mathbf{A}_1\|_{2,1} \\ &+ \|\mathbf{H}\mathbf{Y}-\mathbf{X}\mathbf{H}_2\mathbf{A}_2\mathbf{Z}\|_F^2 + \alpha_2 \|\mathbf{A}_2\|_{2,1}.\end{aligned} \tag{18}$$

The orthogonal transformation of $\mathbf{Z}$ is used to simplify Eq. (18), the objective function is

$$\begin{aligned}\min_{\boldsymbol{A}_1,\boldsymbol{A}_2} &\left\|\mathbf{H}\mathbf{Y}\mathbf{Z}^T - \mathbf{H}\mathbf{X}_1\mathbf{A}_1\right\|_F^2 + \alpha_1 \|\mathbf{A}_1\|_{2,1} \\ &+ \left\|\mathbf{H}\mathbf{Y}\mathbf{Z}^T - \mathbf{X}\mathbf{H}_2\mathbf{A}_2\right\|_F^2 + \alpha_2 \|\mathbf{A}_2\|_{2,1}.\end{aligned} \tag{19}$$

We use iteratively reweighted least square to optimize Eq. (19), which is expressed as

$$\begin{aligned}\min_{\mathbf{A}_1,\mathbf{A}_2} &\left\|\mathbf{H}\mathbf{Y}\mathbf{Z}^T - \mathbf{H}\mathbf{X}_1\mathbf{A}_1\right\|_F^2 + \alpha_1 \mathrm{tr}\left(\mathbf{A}_1^T\mathbf{P}\mathbf{A}_1\right) \\ &+ \left\|\mathbf{H}\mathbf{Y}\mathbf{Z}^T - \mathbf{X}\mathbf{H}_2\mathbf{A}_2\right\|_F^2 + \alpha_2 \mathrm{tr}\left(\mathbf{A}_2^T\mathbf{Q}\mathbf{A}_2\right),\end{aligned} \tag{20}$$

where $\mathbf{P} \in \mathbb{R}^{d\times d}$ and $\mathbf{Q} \in \mathbb{R}^{d\times d}$ are diagonal matrices with $p_{jj} = \frac{1}{2\left\|\mathbf{A}_{1_j}\right\|_2^2}$ and $q_{jj} = \frac{1}{2\left\|\mathbf{A}_{2_j}\right\|_2^2}$, $j = 1,\ldots,d$, respectively. By setting partial derivatives of $\mathbf{A}_1$ and $\mathbf{A}_2$ in Eq. (20) to be zero, $\mathbf{A}_1$ and $\mathbf{A}_2$ becomes

$$\mathbf{A}_1 = (\mathbf{X}_1^T\mathbf{H}\mathbf{X}_1 + \alpha_1\mathbf{P})^{-1}\mathbf{X}_1^T\mathbf{H}\mathbf{Y}\mathbf{Z}^T), \tag{21}$$

$$\mathbf{A}_2 = (\mathbf{X}_2^T\mathbf{H}\mathbf{X}_2 + \alpha_2\mathbf{Q})^{-1}\mathbf{X}_2^T\mathbf{H}\mathbf{Y}\mathbf{Z}^T). \tag{22}$$

Algorithm 1 Pseudo Code of Solving Eq. (10)

Input: $\mathbf{X}_1 \in \mathbb{R}^{n\times d}$, $\mathbf{X}_2 \in \mathbb{R}^{n\times d}$, $\mathbf{Y} \in \mathbb{R}^{n\times c}$, α_1, α_2;
Output: $\mathbf{W}_1$, $\mathbf{W}_2$;

```
Initialize t = 1, a as a column vector of all elements set to ones, and b_1, b_2 as two random vectors;
Initialize W_1, W_2, Z_1, Z_2, as four random matrices;
repeat
    Update Z (t+1) via Eq. (12);
    Update b_1 (t+1) via Eq. (16);
    Update b_2 (t+1) via Eq. (17);
    Update P (t+1) via p(t+1)_jj = 1 / (2 ||A(t+1)_1j||_2^2), j = 1, ..., d;
    Update Q (t+1) via q(t+1)_jj = 1 / (2 ||A(t+1)_2j||_2^2), j = 1, ..., d;
    Update A_1 (t+1) via Eq. (21);
    Update A_2 (t+1) via Eq. (22);
    t = t + 1;
Until The difference between the objective function values of Eq. (10) within two sequential iterations is less than 10^-8;
Update W_1 (t+1) and W_2 (t+1) via Eq. (6);
```

Finally, we can obtain optimal parameters $\boldsymbol{Z}$, $\boldsymbol{b_1}$, $\boldsymbol{b_2}$, $\boldsymbol{A_1}$ and $\boldsymbol{A_2}$, and then calculate the coefficient matrix $\boldsymbol{W_1}$ and $\boldsymbol{W_2}$ via Eq. (6) to obtain the complementary features from six dual-networks, respectively. Algorithm 1 summarizes the optimization processes of solving Eq. (10).

E. TRAIN E-SVM CLASSIFIERS

As shown in Fig. 4, we train the classifiers via fused features and voting strategy to conduct the breast cancer classification. Specifically, the dual-network features are first fused to train SVM classifiers, as shown in Fig. 4 (a). E-SVM classifier is trained by voting strategy as shown in Fig. 4 (b). For the features of each dual-network, we can train two SVMs and vote on their predicted score.

We train an E-SVM classifier with high precision and strong robustness by considering the fused features of dual-network and voting strategy for the predicted score of two dual-networks as depicted in Fig. 4 (c). In our experiments, three E-SVM classifiers are trained. The reason why we only train three E-SVM classifiers is that the features of training each E-SVM classifier must be derived from different DCNNs rather than sharing the features of the same DCNN (e.g., D-R and D-MV shares the features of DenseNet-121), which avoids introducing redundant features.

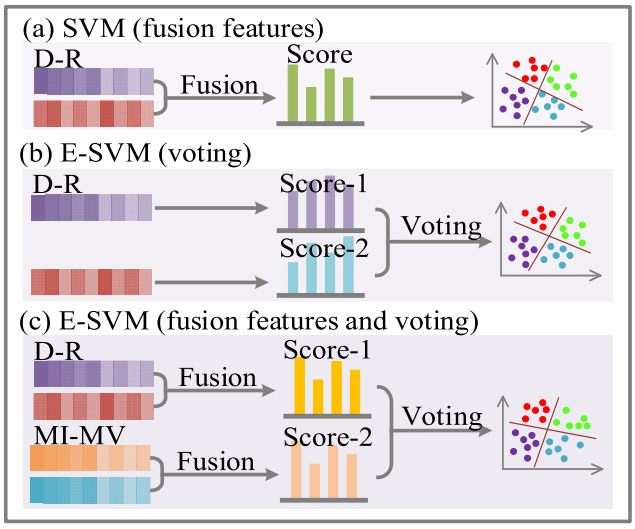


FIGURE 4. Training classifiers: (a) training SVM using fused features of the dual-network; (b) training E-SVM using voting strategy between the dual-network; (c) training E-SVM using fused features and voting strategy between different dual-networks.

IV. EXPERIMENTAL SETUP AND RESULTS

A. DATASETS

The public dataset of the ICIAR 2018 grand challenge [20] is used in our experiments. It includes 400 H&E stained images of breast histology microscopy (2048 × 1536 pixels). All the images are digitized with the same acquisition conditions, with a magnification of 200× and pixel size of 0.42μm × 0.42μm. Each image is labeled with one of the four balanced classes: benign, in situ carcinoma, invasive carcinoma, and normal. Each class has 100 images and a typical image of each predominant cancer type is shown in Fig. 1. Moreover, two expert pathologists perform the image-wise annotation of ICIAR 2018 dataset [20].

B. EVALUATION METRICS

In this paper, we adopt a nested 10-fold cross-validation method to complete all the experimental verification. We evaluate the performances of our proposed breast cancer histopathological image classification model in ICIAR 2018 challenge dataset [20] in terms of accuracy (ACC), area under receiver operating characteristic (ROC) curve (AUC), precision (Pre), Recall, and F1 score. The metrics for evaluating classification results are defined as follow

$$\mathrm{ACC} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{FP}+\mathrm{TN}+\mathrm{FN}}, \tag{23}$$

$$\mathrm{AUC} = \frac{\sum rank_i - \frac{M\times(M+1)}{2}}{M \times N}, \tag{24}$$

$$\mathrm{Pre} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}, \tag{25}$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}, \tag{26}$$

$$\mathrm{F1} = \frac{2 \times \mathrm{Pre} \times \mathrm{Recall}}{\mathrm{Pre}+\mathrm{Recall}}, \tag{27}$$

where TP, TN, FP, and FN are the number of true positive, true negative, false positive, and false negative, respectively. $\sum rank_i$ represents the sum of the serial number of positive samples. M and N represent the number of positive and negative samples, respectively.

C. EXPERIMENTAL RESULTS

1) EXPERIMENTS ON THE BASELINE DCNNS

We evaluate six DCNNs (e.g., DenseNet-121, ResNet-50, ML-InceptionV3, ML-VGG-16, InceptionV3, and VGG-16) using two different pre-set crop sizes (e.g., 400×400 pixels and 700 × 700 pixels) and fusing these features of two different image crop sizes. The SVM classifiers are trained to perform classification tasks, the results are shown in Fig. 5. It can be noticed that, by comparing the results of DenseNet-121 and ResNet-50 of deeper network structure, the classification accuracy can be significantly improved using fused features between different image crop sizes. The features of different image crop sizes are different. Hence, more comprehensive features are obtained by fused features.

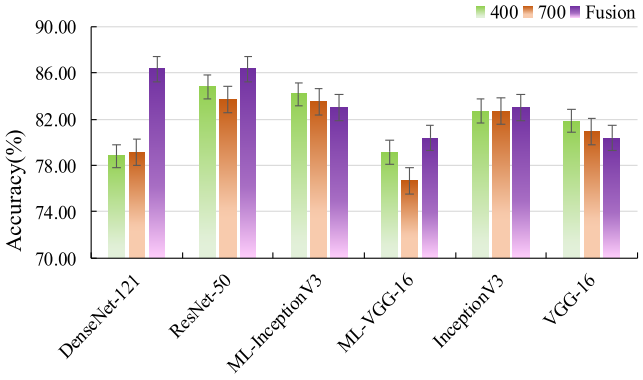


FIGURE 5. Experimental results of different scenarios: green represents the image crops with a size of 400 × 400 pixels, orange represents the image crops with a size of 700 × 700 pixel, and purple represents the features fusion between different crop sizes.

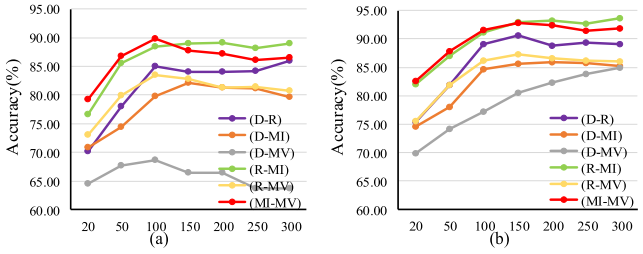


FIGURE 6. Experimental results of different feature dimensions in six dual-network: (a) Using the fused features strategy to train SVM, (b) Using the voting strategy to train E-SVM.

In addition, the classification accuracy of ML-InceptionV3 is better than InceptionV3, but ML-VGG-16 is worse than VGG-16. This is probably because ML-VGG-16 has many redundant and noisy features, which can

TABLE 1. Experimental results using different fused features strategies to train SVM classifier (± 0.50%).

Dual-net work	100					150					200				
	ACC	AUC	Pre	Recall	F1	ACC	AUC	Pre	Recall	F1	ACC	AUC	Pre	Recall	F1
D-R	84.95	89.91	85.96	85.15	85.55	84.10	89.38	85.55	84.21	84.87	84.00	89.33	85.44	84.05	84.74
D-MI	79.75	86.50	81.31	79.83	80.56	82.15	88.18	83.70	82.38	83.03	81.30	87.52	82.84	81.47	82.15
D-MV	68.70	76.07	81.31	79.83	80.56	66.40	77.72	83.70	82.38	83.03	66.45	77.51	82.84	81.47	82.15
R-MI	88.45	91.23	89.79	88.65	89.22	89.00	92.79	90.02	89.11	89.56	89.10	92.66	90.06	89.14	89.60
R-MV	83.45	88.95	84.63	83.54	84.08	82.80	88.53	84.09	82.88	83.48	81.25	87.56	80.98	81.30	81.14
MI-MV	89.75	93.13	90.83	89.85	90.34	87.80	91.86	89.32	88.17	88.74	87.15	91.53	88.06	87.25	87.65

TABLE 2. Experimental results using the voting strategy to train E-SVM classifiers (± 0.50%).

Dual-net work	100					150					200				
	ACC	AUC	Pre	Recall	F1	ACC	AUC	Pre	Recall	F1	ACC	AUC	Pre	Recall	F1
D-R	89.15	92.71	90.40	89.28	89.84	90.60	93.66	90.48	90.67	90.57	88.75	92.40	91.31	89.01	90.15
D-MI	84.65	89.79	86.27	84.88	85.57	85.65	90.43	86.83	85.80	86.31	85.85	90.62	87.12	85.88	86.50
D-MV	77.25	84.80	79.08	77.42	78.24	80.50	86.78	82.24	80.74	81.48	82.25	88.22	83.63	82.21	82.91
R-MI	91.20	94.16	92.10	91.45	91.77	92.90	95.36	93.41	92.91	93.16	93.25	95.45	93.93	93.33	93.63
R-MV	86.25	90.87	87.52	86.33	86.92	87.35	91.50	88.53	87.53	88.03	86.65	91.15	82.66	86.73	84.65
MI-MV	91.55	94.27	92.50	91.67	92.08	92.80	95.16	93.62	92.84	93.23	92.35	94.89	93.17	92.46	92.81

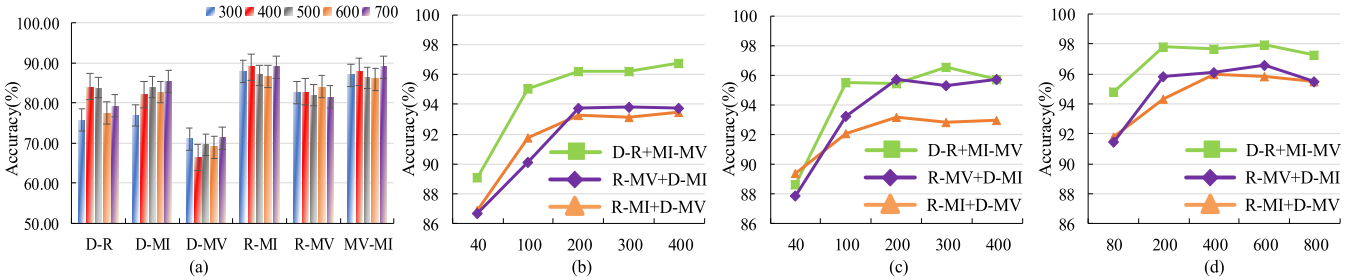


FIGURE 7. Experimental results: (a) Different image crop sizes (e.g., 20 image crops, 50 color-enhanced images, the feature dimension is 300; (b) E-SVM classifier trained using image crop size of 400 × 400 pixels; (c) E-SVM classifier trained using image crop size of 650 × 650 pixels; (d) E-SVM classifier trained using features fusion between two image crop sizes.

affect the performance of the classifier. However, this is avoided in our proposed method as we develop a feature selection algorithm to reduce redundant and noisy features.

2) EXPERIMENTS ON THE DOLL FEATURE SELECTION METHOD

We evaluate the effectiveness of the DOLL feature selection method using fused features of the dual-network and voting strategy between two dual-networks. We set the image preprocessing parameters as follows: 50 color-enhanced images, cropped image size of 400 × 400 pixels, and 20 cropped images.

The classification accuracies of different feature dimensions are shown in Fig. 6. When feature dimensions are between 100 and 150, the best accuracy is obtained, and the accuracy hardly improves with the increase of feature dimensions. Meanwhile, we note that three dual-networks of MI-MV, R-MI, and D-R obtain better accuracy using either SVM or E-SVM classifiers. Also, the performance of three different dimensional features (e.g. 100, 150, and 200) is verified in detail. Fused features of each dual-network are then used to train SVM classifiers (e.g., Fig. 4 (a)), the results are shown in Table 1. Compared with the experimental results, we find that the classification accuracy of the four dual-networks of D-MI, R-MI, R-MV, and MI-MV has no significant improvement, and the results of the D-R and D-MV are even worse. The reason is that the poor correlation between deeper and shallower networks, which limits the ability of the proposed feature selection algorithm. Last, E-SVM classifiers (e.g., Fig. 4 (b)) are trained by the voting strategy, the results are shown in Table 2. The performance is worse when the feature dimension is reduced. Meanwhile, by comparing Table 1 with Table 2, we observe that the performance of the E-SVM classifiers is better. It is also shown that the voting strategy can improve the accuracy of classification.

In conclusion, at the expense of a small amount of precision, the feature dimensions can be greatly reduced, a significant amount of noise and redundant features are removed, and overfitting is avoided, which shows that the proposed feature selection method is effective. We also prove that fused features and voting strategies are effective in improving classification performance.

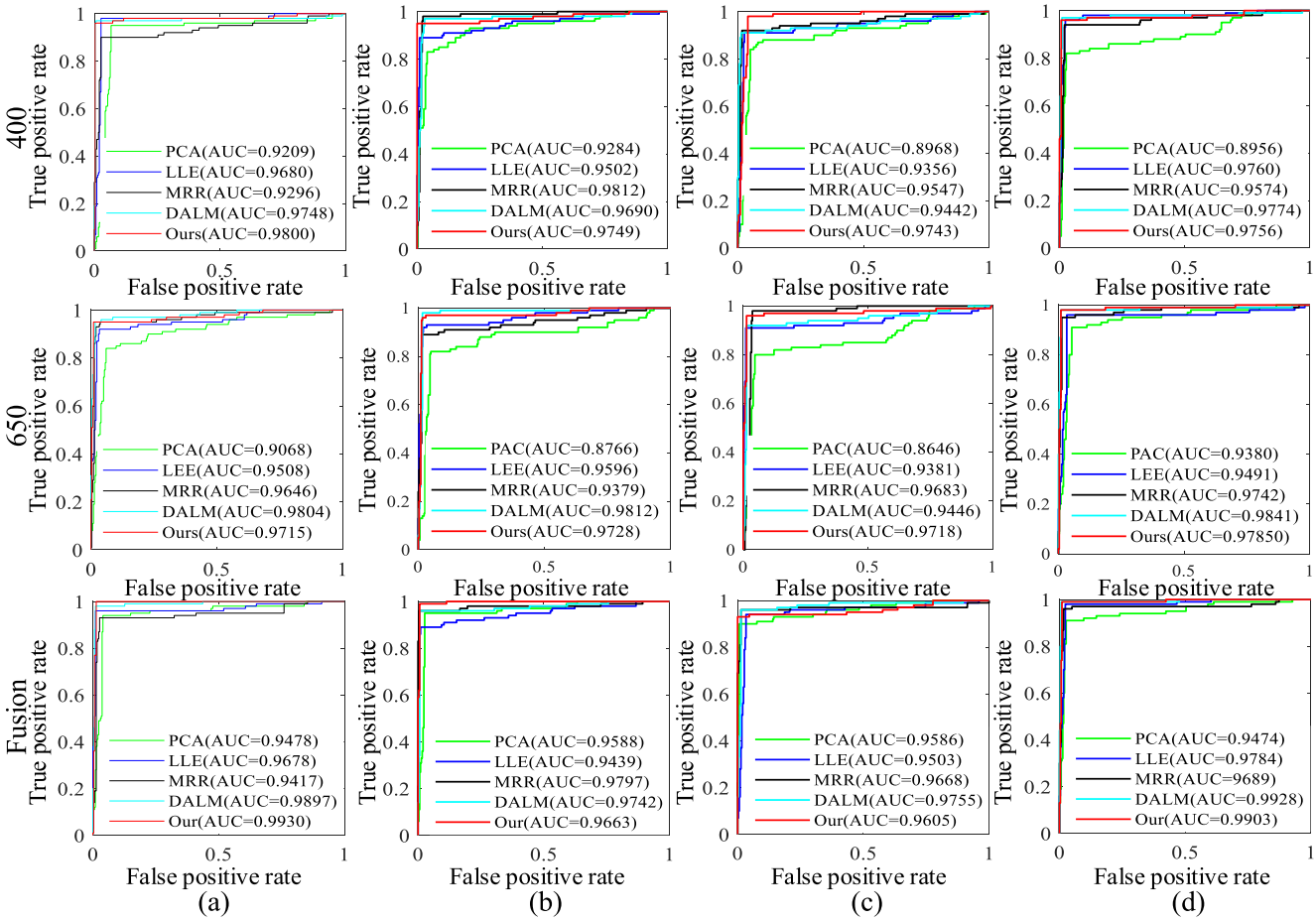


FIGURE 8. The ROC curves of the experimental results of the E-SVM classifiers with D-R and MI-MV. From top to bottom: the image crop size is 400 × 400 pixels and the feature dimensions is 200; the image crop size is 650 × 650 pixels and the feature dimensions is 200; fused features between 400 × 400 pixels and 650 × 650 pixels, and the feature dimensions is 200. From left to right, (a) is benign, (b) is in situ carcinoma, (c) is invasive carcinoma, and (d) is normal.

3) EXPERIMENTS OF FUSED FEATURES AND E-SVM CLASSIFIER

In this sub-section, we first explore the impact of different image crop sizes. The fused features of each dual-network are used to train SVM classifiers (e.g., Fig. 4 (a)), the results are shown in Fig. 7 (a). We have the following observations. On the one hand, when the cropped image size was 400 × 400 pixels, we get the best classification accuracy and while cropped size between 600 and 700 pixels, the accuracy becomes unstable. On the other hand, we can find that D-MV produces low accuracy. The main reasons are as follows. When smaller size image crops are resized 224 × 224 pixels, we retain more local detailed feature information. When larger size image crops are resized 224 × 224 pixels, more global feature information can be retained. Therefore, in the following experiments, we fuse the features between 400 × 400 pixels and 650 × 650 pixels to improve the classification accuracy.

Next, we combine the image preprocessing (e.g., image crop sizes of 400 × 400 or 650 × 650 pixels, 20 image crops, 50 color-enhanced images), the DOLL method, the fused features, and the voting strategies together to further improve classification accuracy of breast cancer pathological images. Three E-SVMs are trained by utilizing the fused features of six dual-networks (e.g., D-R and MI-MV, R-MI and D-MV, R-MV and D-MI), as shown in Fig. 4 (c), and the results are shown in Fig. 7 (b), (c), and (d). The E-SVM classifier of using the dual-network features of D-R and MI-MV has the best performance when compared with the other configurations. Moreover, the fused features of different image crop sizes can obtain better accuracy with the single image crop size features. This is because the features fusion cannot only obtain local detailed feature information that utilize small size image crop, but also obtains global feature information from large size image crop.

Furthermore, we compare the experimental results of different feature selection methods (e.g., PCA, LLE, MRR, DALM, and DOLL), which use the features of different cropped image sizes and fused features, and three E-SVM classifiers. The results are summarized in Table 3. The feature

TABLE 3. Experimental results using the voting strategy to train e-svm classifiers: ics is image crop size; fd is feature dimension (± 0.50%).

ICS (FD)	Method	D-R + MI-MV					R-MI + D-MV					R-MV + D-MI				
		ACC	AUC	Pre	Recall	F1	ACC	AUC	Pre	Recall	F1	ACC	AUC	Pre	Recall	F1
400 (200)	PCA	86.90	91.21	87.89	87.00	87.44	89.80	93.20	90.84	89.93	90.38	88.20	92.09	89.14	88.26	88.70
	LLE	93.20	95.51	93.62	93.26	93.44	93.20	95.42	93.73	93.26	93.49	89.50	92.97	90.55	89.69	90.12
	MRR	92.60	95.11	93.21	92.65	92.93	92.85	95.28	93.45	92.86	93.15	94.75	96.38	95.98	94.75	95.36
	DALM	95.65	97.07	96.08	95.68	95.88	96.35	97.56	96.68	96.39	96.53	91.50	94.23	92.42	91.54	91.98
	Ours	**96.20**	**97.50**	**96.55**	**96.27**	**96.41**	93.30	95.54	93.88	93.35	93.61	93.75	95.84	94.25	93.77	94.01
650 (200)	PCA	85.50	90.43	86.67	85.64	86.15	88.85	92.53	89.81	88.99	89.40	87.00	91.42	88.25	87.19	87.72
	LLE	92.90	95.23	93.52	93.00	93.26	91.75	94.56	93.92	91.79	92.84	92.10	94.70	92.91	92.18	92.54
	MRR	94.45	96.31	94.40	94.54	94.47	93.95	95.98	94.53	93.99	94.26	95.25	96.75	97.10	95.32	96.20
	DALM	93.75	95.79	96.00	93.74	94.86	95.50	96.97	96.92	95.52	96.21	94.25	96.20	94.73	94.28	94.50
	Ours	**95.45**	**97.05**	**95.88**	**95.50**	**95.69**	93.20	95.44	93.80	93.27	93.53	95.75	97.16	96.16	95.79	95.97
Fusion (400)	PCA	91.00	93.88	91.84	91.13	91.48	92.95	95.27	93.56	93.01	93.28	91.75	94.54	92.40	91.78	92.09
	LLE	95.95	96.00	96.25	97.32	96.78	94.25	96.18	96.70	94.39	95.53	93.40	95.52	94.10	93.56	93.83
	MRR	94.20	96.13	94.63	94.24	94.43	95.10	96.77	95.58	95.16	95.37	96.25	97.43	97.36	96.27	96.81
	DALM	96.50	97.62	96.83	96.53	96.68	97.50	98.29	97.73	97.51	97.62	94.90	96.61	95.43	94.93	95.18
	Ours	**97.70**	**98.48**	**97.92**	**97.72**	**97.82**	96.00	97.34	96.34	96.04	96.19	96.15	97.43	96.51	96.19	96.35

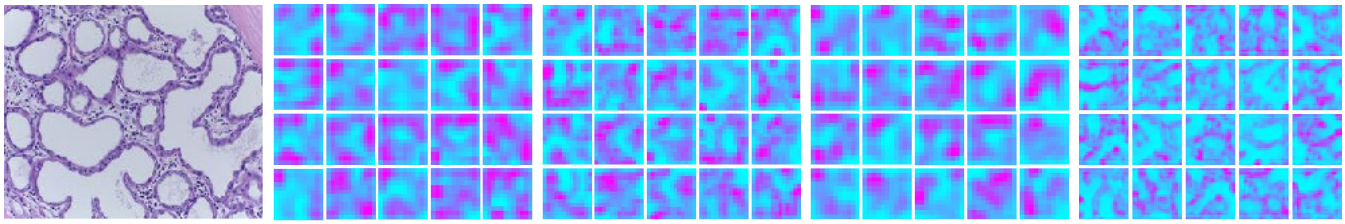


FIGURE 9. Feature maps of the last convolutional layer of different DCNN.

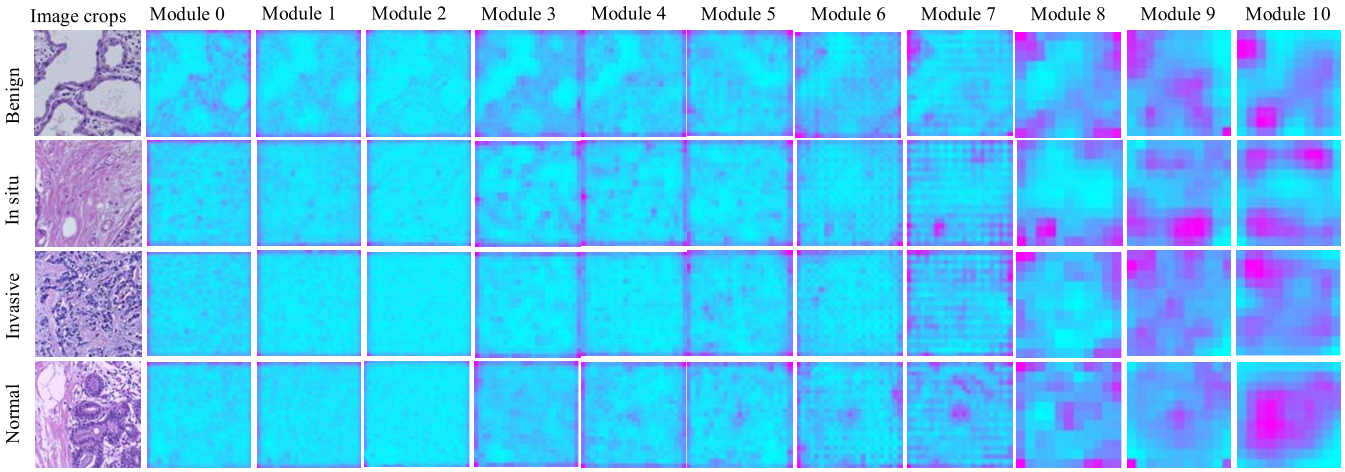


FIGURE 10. The feature maps of the middle module of our proposed ML-InceptionV3 model.

dimensions of each DCNN is reduced to 100 dimensions in this experiment. Our method obtains the best classification accuracy compared with other methods, except the results of DALM method in E-SVM classifier trained with the R-MI and D-MV. In Fig. 8, we plot the ROC curves of the experimental results of the E-SVM classifiers with D-R and MI-MV, which use different feature selection methods. By analyzing the ROC curves of the four classes of breast cancer pathological images, we also find that the classification accuracy of our proposed method is better than the competing methods. Finally, we compare the performance of our proposed model with other state-of-the-art methods, as shown in Table 4. The classification model of our proposed method obtains the best classification accuracy of 97.70%. And the reduction of feature dimensions can reduce computational costs and avoid overfitting, which makes the E-SVM classifier more robust.

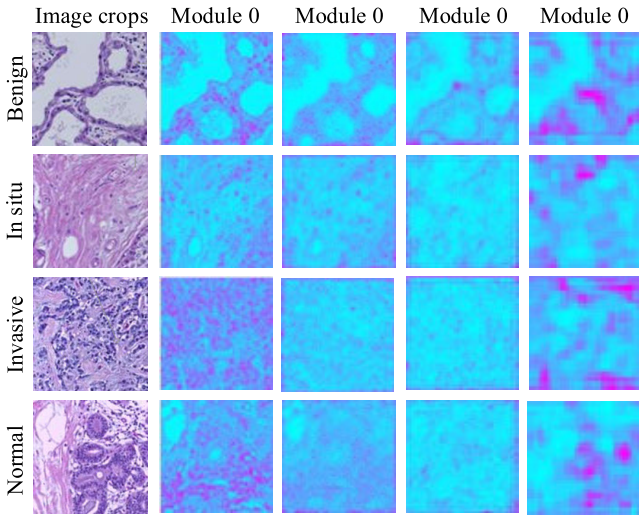


FIGURE 11. Feature maps of the middle module of our proposed ML-VGG-16 model.

TABLE 4. Experimental results of different classification methods (± 0.50%).

Method	Accuracy
Ours	**97.70**
Nadia *et al.* [34]	97.30
Kaushiki *et al.* [33]	90.00
Artem *et.a.l.*[36]	90.00
Alexander *et al.* [12]	87.20
Aditya *et al.* [37]	85.00

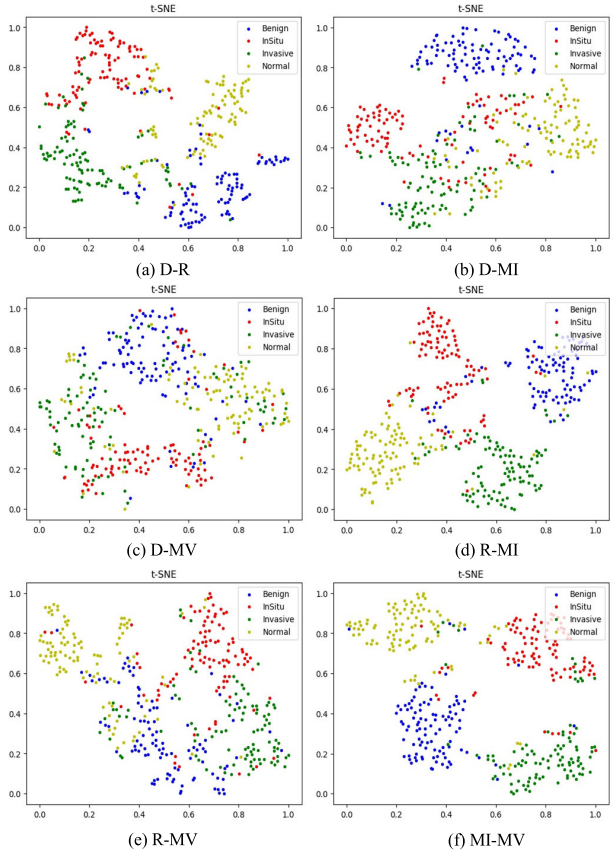


FIGURE 12. t-SNE visualization of fused features of the six dual-networks after dimension reduction by DOLL method.

V. DISCUSSIONS

In this paper, to improve the efficacy and accuracy of the early diagnosis of breast cancer, we propose a classification model based on the multi-network features, the DOLL feature dimension reduction, and the E-SVM classifier. First, to demonstrate the complementarity of the features of different DCNNs, for each input image, we randomly obtain 20 image crops with a size of 400 × 400 and processed them using four DCNNs. We extract the feature maps of these image crops from the last convolution layer and visualize them in Fig. 9. These feature maps have complementarity and redundancy. DenseNet-121 and ResNet-50 have stronger learning ability, which acquire many abstract semantic features. Hence, it is wise and effective to use dual-network fused features to achieve more accurate and robust classification.

Second, to verify the effectiveness of ML-InceptionV3 and ML-VGG-16, we extract and visualize the feature map of the middle module of the networks in Figs. 10 and 11. By increasing the depth of network structure increases gradually, we obtain not only the shallow detail features (e.g., shape, boundary, texture and color), but also the semantic ones. Meanwhile, the features of all modules before *i*-th intermediate module are fused, which are used to train the SVM classifiers. The experimental results are shown in Tables 5 and 6. It is noted that better experimental results are obtained by feature fusion with the increase of blocks in ML-InceptionV3 and ML-VGG-16, respectively. This shows that our improvements for VGG-16 and InceptionV3 are effective.

TABLE 5. Experimental results of the fused features among all the preceding block of each intermediate block in ML-InceptionV3 (± 0.50%).

Module	400 × 400 pixels				
	ACC	AUC	Pre	Recall	F1
0 - 0	69.10	79.34	73.64	69.83	71.68
0 - 1	70.05	80.27	75.69	70.70	73.11
0 - 2	70.05	80.27	75.69	70.70	73.11
0 - 3	74.65	83.36	79.64	75.25	77.38
0 - 4	80.20	86.75	82.49	80.60	81.53
0 - 5	81.65	87.66	83.11	81.80	82.45
0 - 6	81.85	88.52	84.07	82.94	83.50
0 - 7	80.95	87.27	83.31	81.29	82.29
0 - 8	84.25	89.52	85.89	84.39	85.13
0 - 9	85.45	90.27	86.80	85.55	86.17
0 - 10	**83.30**	**88.97**	**84.53**	**83.49**	**84.01**

Third, to illustrate the separability of the features after dimension reduction, the fused features of six dual-networks are visualized using t-SNE visualization method [46]. They all perform feature dimension reduction using the DOLL method, as shown in Fig. 12. It is shown that the fused features of D-R, R-MI and MI-MV have better separability, which are consistent with the experimental results in Table 3.

TABLE 6. Experimental results of the fused features among all the preceding block of each intermediate block in ML-VGG-16 (± 2.00%).

Module	400 × 400 pixels				
	ACC	AUC	Pre	Recall	F1
0 - 0	45.60	63.82	47.25	45.74	46.48
0 - 1	62.95	75.29	64.44	63.09	63.76
0 - 2	73.65	82.45	75.64	73.89	74.75
0 - 3	**78.10**	**85.31**	**80.36**	**78.28**	**79.31**

VI. CONCLUSION

In this paper, we propose an effective model for the classification of H&E stained histological breast cancer images. To increase the accuracy and robustness of the classifier, we extract image features of the multi-network by using four pre-trained DCNNs. Further, based on three relationships, we devise a new feature selection method of DOLL algorithm to enhance the classification results by reducing the feature dimension to alleviate overfitting. Moreover, we train the E-SVM classifiers by fused features and voting strategies to improve the classification accuracy. Our proposed model achieves considerable high accuracy and strong robustness. The proposed method is expected to be clinically useful for doctors to achieve the early diagnosis of breast cancer, which can benefit the survival chances of breast cancer patients. In our future work, we will validate our method on clinical data and investigate the efficacy of our method on different types of cancers.

REFERENCES

[1] C. E. Desantis, J. Ma, A. G. Sauer, L. A. Newman, and A. Jemal, "Breast cancer statistics, 2017, racial disparity in mortality by state," *CA, A Cancer J. Clinicians*, vol. 67, no. 6, pp. 439–448, Nov. 2017.

[2] R. M. Rangayyan and J. S. Suri, *Recent Advances in Breast Imaging, Mammography, and Computer-Aided Diagnosis of Breast Cancer*. Bellingham, WA, USA: SPIE, 2006.

[3] T. Messay, R. C. Hardie, and S. K. Rogers, "A new computationally efficient CAD system for pulmonary nodule detection in CT imagery," *Med. Image Anal.*, vol. 14, no. 3, pp. 390–406, Jun. 2010.

[4] K. Doi, "Computer-aided diagnosis in medical imaging: Historical review, current status and future potential," *Computerized Med. Imag. Graph.*, vol. 31, nos. 4–5, pp. 198–211, Jun. 2007.

[5] D. Gur, J. H. Sumkin, H. E. Rockette, M. Ganott, C. Hakim, L. Hardesty, W. R. Poller, R. Shah, and L. Wallace, "Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system," *J. Nat. Cancer Inst.*, vol. 96, no. 3, pp. 185–190, Feb. 2004.

[6] D. Shen, G. Wu, and H. Suk, "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, Jun. 2017.

[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[8] S. K. Zhou, H. Greenspan, and D. Shen, *Deep Learning for Medical Image Analysis*. New York, NY, USA: Academic, 2017.

[9] A. Madabhushi and G. Lee, "Image analysis and machine learning in digital pathology: Challenges and opportunities," *Med. Image Anal.*, vol. 33, pp. 170–175, Oct. 2016.

[10] S. Chattoraj and K. Vishwakarma, "Classification of histopathological breast cancer images using iterative VMD aided Zernike moments & textural signatures," pp. 1–8, 2018, *arXiv:1801.04880*. [Online]. Available: https://arxiv.org/abs/1801.04880

[11] D. M. Vo, N.-Q. Nguyen, and S.-W. Lee, "Classification of breast cancer histology images using incremental boosting convolution networks," *Inf. Sci.*, vol. 482, pp. 123–138, May 2019.

[12] A. Rakhlin, A. Shvets, V. Iglovikov, and A. A. Kalinin, "Deep convolutional neural networks for breast cancer histology image analysis," in *Proc. Int. Conf. Image Anal. Recognit.*, 2018.

[13] T. Araújo, G. Aresta, E. Castro, J. Rouco, P. Aguiar, C. Eloy, A. Polónia, and A. Campilho, "Classification of breast cancer histology images using convolutional neural networks," *PLoS ONE*, vol. 12, no. 6, Jun. 2017, Art. no. e0177544.

[14] Z. Han, B. Wei, Y. Zheng, Y. Yin, K. Li, and S. Li, "Breast cancer multi-classification from histopathological images with structured deep learning model," *Sci. Rep.*, vol. 7, no. 1, pp. 4172–4182, 2017.

[15] P. Wang, X. Hu, Y. Li, Q. Liu, and X. Zhu, "Automatic cell nuclei segmentation and classification of breast cancer histopathology images," *Signal Process.*, vol. 122, pp. 1–13, May 2016.

[16] D. Komura and S. Ishikawa, "Machine learning methods for histopathological image analysis," *Comput. Struct. Biotechnol. J.*, vol. 16, pp. 34–42, 2018.

[17] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Comput. Grap. Appl.*, vol. 21, no. 4, pp. 34–41, Jul./Aug. 2001.

[18] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Comput. Sci.*, vol. 1409, no. 1556, pp. 1–14, Sep. 2015.

[19] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6848–6856.

[20] G. Aresta *et al.*, "BACH: Grand challenge on breast cancer histology images," *Med. Image Anal.*, vol. 56, pp. 122–139, Aug. 2019.

[21] J. Cheng, D. Ni, Y. Chou, J. Qin, C. Tiu, Y. Chang, C. Huang, D. Shen, and C. Chen, "Computer-aided diagnosis with deep learning architecture: Applications to breast lesions in US images and pulmonary nodules in CT scans," *Sci. Rep.*, vol. 6, Apr. 2016, Art. no. 24454.

[22] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

[23] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.

[24] Y.-L. Boureau, J. Ponce, and Y. LeCun, "A theoretical analysis of feature pooling in visual recognition," in *Proc. Int. Conf. Mach. Learn.*, 2010, pp. 111–118.

[25] F. Nie, W. Zhu, and X. Li, "Unsupervised feature selection with structured graph optimization," in *Proc. AAAI Conf. Artif. Intell.*, 2016, pp. 1302–1308.

[26] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 2, nos. 1–3, pp. 37–52, 1987.

[27] S. T. Roweis, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[28] X. Zhu, H.-I. Suk, and D. Shen, "Low-rank dimensionality reduction for multi-modality neurodegenerative disease identification," *World Wide Web*, vol. 22, no. 2, pp. 907–925, Mar. 2019.

[29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[30] Y. Zheng, Z. Jiang, H. Zhang, F. Xie, Y. Ma, H. Shi, and Y. Zhao, "Histopathological whole slide image analysis using context-based CBIR," *IEEE Trans. Med. Imaging*, vol. 37, no. 7, pp. 1641–1652, Jul. 2018.

[31] P. Sudharshan, C. Petitjean, F. Spanhol, L. E. Oliveira, L. Heutte, and P. Honeine, "Multiple instance learning for histopathological breast cancer image classification," *Expert Syst. Appl.*, vol. 117, pp. 103–111, Mar. 2019.

[32] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[33] K. Roy, D. Banik, D. Bhattacharjee, and M. Nasipuri, "Patch-based system for Classification of Breast Histology images using deep learning," *Computerized Med. Imaging Graph.*, vol. 71, pp. 90–103, Jan. 2019.

[34] N. Brancati, M. Frucci, and D. Riccio, "Multi-classification of breast cancer histology images by using a fine-tuning strategy," in *Proc. Int. Conf. Image Anal. Recognit.*, 2018, pp. 771–778.

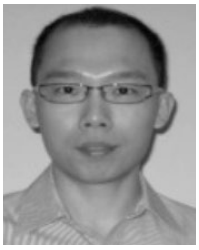
[35] Z. Gandomkar, P. C. Brennan, and C. Mello-Thoms, "MuDeRN: Multi-category classification of breast histopathological image using deep residual networks," *Artif. Intell. Med.*, vol. 88, pp. 14–24, Jun. 2018.

[36] A. Pimkin, G. Makarchuk, V. Kondratenko, M. Pisov, E. Krivov, and M. Belyaev, "Ensembling neural networks for digital pathology images classification and segmentation," in *Proc. Int. Conf. Image Anal. Recognit.*, 2018, pp. 877–886.
[37] A. Golatkar, D. Anand, and A. Sethi, "Classification of breast cancer histology using deep learning," in *Proc. Int. Conf. Image Anal. Recognit.*, 2018, pp. 837–844.
[38] S. Vesal, N. Ravikumar, A. Davari, S. Ellmann, and A. Maier, "Classification of breast cancer histology images using transfer learning," in *Proc. Int. Conf. Image Anal. Recognit.*, 2018, pp. 812–819.
[39] R. Tomioka and M. Sugiyama, "Dual-augmented lagrangian method for efficient sparse reconstruction," *IEEE Signal Process. Lett.*, vol. 16, no. 12, pp. 1067–1070, Dec. 2009.
[40] X. Zhu, H.-I. Suk, L. Wang, S.-W. Lee, and D. Shen, "A novel relational regularization feature selection method for joint regression and classification in AD diagnosis," *Med. Image Anal.*, vol. 38, pp. 205–214, May 2017.
[41] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
[42] Z. Ma and T. Sun, "Adaptive estimation in two-way sparse reduced-rank regression," pp. 1–25, 2014, *arXiv:1403.1922*. [Online]. Available: https://arxiv.org/abs/1403.1922
[43] C. Giraud, "Low rank multivariate regression," *Electron. J. Statist.*, vol. 5, no. 0, pp. 775–799, 2011.
[44] G. Liu and S. Yan, *Low-Rank and Sparse Modeling for Visual Analysis*. Cham, Switzerland: Springer, 2014.
[45] J. C. Gower and G. B. Dijksterhuis, *Procrustes Problems*. Oxford, U.K.: Oxford Univ. Press on Demand, 2004.
[46] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

YONGJUN WANG received the B.S. degree from Nanchang Hangkong University, Jiangxi, China, in 2017. He is currently pursuing the degree with Shenzhen University. His research interests include machine learning, deep learning, artificial intelligence, and image processing.

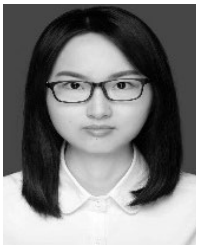
BAIYING LEI received the M.Eng. degree in electronics science and technology from Zhejiang University, China, in 2007, and the Ph.D. degree from Nanyang Technological University (NTU), Singapore, in 2013. She is currently an Associate Professor with the School of Biomedical Engineering, Shenzhen University, China. Her current research interests include medical image analysis, machine learning, digital watermarking, and signal processing.

AHMED ELAZAB received the Ph.D. degree in pattern recognition and intelligent system from the Shenzhen Institutes of Advanced Technology, University of Chinese Academy of Sciences, China, in 2017. He is currently an Assistant Professor with the Computer Science Department, Misr Higher Institute for Commerce and Computers, Mansoura, Egypt. He is also a Postdoctoral Fellow with the School of Biomedical Engineering, Shenzhen University, Shenzhen, China. He has authored or coauthored more than 30 peer-reviewed articles. He received the Best Paper Award of the 7th Cairo International Biomedical Engineering Conference 2014 (IEEE/EMB). He received two outstanding student awards from the Shenzhen Institutes of Advanced Technology. He served as a Reviewer for prestigious peer-reviewed international journals. His current research interests include pattern recognition, medical image analysis, and computer-aided diagnosis.



EE-LENG TAN received the B.Eng. (Hons.) and Ph.D. degrees in electrical and electronic engineering from Nanyang Technological University, in 2003 and 2012, respectively. His research interests include image/audio processing and real-time digital signal processing. His work has been awarded three patents in Japan, Singapore, and USA. He currently holds the position of the Chief Science Officer, leading the research and development of the technological company Beijing Sesame World Company Ltd. He also consults as the Technical Advisor for several start-ups.

WEI WANG received the B.S., M.S., and Ph.D. degrees from Chongqing University, Chongqing, China, in 2012, 2014, and 2017, respectively. From September 2017 to August 2019, he held a postdoctoral position at the College of Information and Engineering, Shenzhen University, where he is currently an Associate Researcher with the School of Biomedical Engineering. His research interests include variational methods, partial differential equations, wavelets, and deep learning for image processing.



FANGLIN HUANG received the B.S. degree from Wenzhou Medical University, Zhejiang, China, in 2017. She is currently pursuing the degree with Shenzhen University. Her research interests include machine learning, artificial intelligence, and image processing.

XUEHAO GONG received the B.S. degree in medical imaging from Former Tongji Medical University, in 1992, and the Ph.D. degree from the Tongji Medical College, Huazhong University of Science and Technology, in 2003. He was a Resident Physician with the Ultrasound Imaging Department, Tongji Hospital, from 1992 to 1998. He is currently a Professor with the First Affiliated Hospital of Shenzhen University, Second People's Hospital of Shenzhen, China.

TIANFU WANG received the Ph.D. degree in biomedical engineering from Sichuan University, in 1997. He is currently a Professor with the School of Biomedical Engineering, Shenzhen University, China, and the Associate Chair of the School of Medicine. His research interests include ultrasound image analysis, medical image processing, pattern recognition, and medical imaging.