

## Research Article

# Medical Image Retrieval Using Empirical Mode Decomposition with Deep Convolutional Neural Network

Shaomin Zhang, Lijia Zhi , and Tao Zhou 

School of Computer Science and Engineering, North Minzu University, Yinchuan 750021, China

Correspondence should be addressed to Lijia Zhi; [ljzhi\\_paper@163.com](mailto:ljzhi_paper@163.com)

Received 8 November 2020; Revised 7 December 2020; Accepted 14 December 2020; Published 28 December 2020

Academic Editor: Lin Gu

Copyright © 2020 Shaomin Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Content-based medical image retrieval (CBMIR) systems attempt to search medical image database to narrow the semantic gap in medical image analysis. The efficacy of high-level medical information representation using features is a major challenge in CBMIR systems. Features play a vital role in the accuracy and speed of the search process. In this paper, we propose a deep convolutional neural network- (CNN-) based framework to learn concise feature vector for medical image retrieval. The medical images are decomposed into five components using empirical mode decomposition (EMD). The deep CNN is trained in a supervised way with multicomponent input, and the learned features are used to retrieve medical images. The IRMA dataset, containing 11,000 X-ray images, 116 classes, is used to validate the proposed method. We achieve a total IRMA error of 43.21 and a mean average precision of 0.86 for retrieval task and IRMA error of 68.48 and F1 measure of 0.66 on classification task, which is the best result compared with existing literature for this dataset.

## 1. Introduction

Imaging through different kinds of medical devices plays a fundamental role in clinical diagnosis [1], treatment planning [2], and treatment response assessing [3] in the process of medical care. In modern hospitals, different modalities and protocols of digital imaging techniques have been used to generate diagnostic images for each patient, including computed tomography (CT), X-ray, ultrasound, hybrid positron emission tomography and computed tomography (PET-CT), and magnetic resonance imaging (MRI). These medical images with multiple dimensions (e.g., 2D, volumetric 3D, and time series) reflect anatomic and functional aspects of organs and tissue types that require domain experts' analysis and interpretation. These volumes are usually formed in the Digital Imaging and Communications in Medicine (DICOM) format and stored in picture archiving and communication systems (PACS) [4]. A domain expert can search PACS through patient's ID, study ID, time range, or other textual keywords, which is labor intensive and time consuming. As an important part of computer-aided diagnostics

(CAD), content-based medical image retrieval (CBMIR) [5–8] can retrieve medical images mainly via visual contents (e.g., same modality, same body orientation, same anatomical region, or same disease condition) in an existing dataset for more accurate comparative diagnosis.

In the CBMIR domain, there are two major directions in research works. One kind of methods focuses on automatic retrieving images from PACS-like databases, which search images of the same imaging modality, body orientation, body region, and the like [9–11]. Another kind of methods put their efforts into retrieving images that characterize the similar disease convenient for diagnostic comparing [12, 13]. In this study, we follow the former methods to propose an effective CBMIR system for 2D slice retrieval. That is because volumetric 3D medical images are formed by a series of 2D slices acquired from the target body organ, and physicians mainly rely on these 2D slices when they are analyzing and interpreting images on hand [8].

Unlike similarity defined in generic image retrieval domain, the retrieved medical images by directly comparing features using some similarity measure may not be in

accordance with what a physician would want for diagnosis, which formed a “semantic gap” in medical image retrieval [5]. To reduce this gap, CBMIR systems are generally designed under a classification-driven strategy. That is, a CBMIR system is trained using supervised approaches with labeled images. When a query image is submitted to the CBMIR system, the query image is classified first, and then, some visual features and similarity measures are used for similarity retrieval [9, 10, 14]. Deep learning is a breakthrough in machine learning research. Using artificial neural networks with many hidden layers to represent digital images has been proven to be a very effective method to describe low-level, mid-level, and high-level semantic features of an image for recognition and other purposes [15–17]. Among different deep learning architectures, deep convolutional neural networks (CNNs) have proven to be powerful tools that achieved very high precision results in many natural image classification contests [18–20]. In the medical field, deep CNNs are also quickly applied for different tasks, and promising results are emerging [9, 10, 21–24]. Training deep CNNs need a large number of labeled images to choose the huge number of parameters. Note that in the medical domain, such large image datasets are quite rare, due to the unbearable high cost of domain experts’ manual image labeling and annotations [5, 21, 24]. And in contrast to generic image databases, medical image datasets usually are unbalanced because of uneven incidence rates of different malignancies. Dropout [25], data augmentation, and transfer learning [26] are the most common techniques used to prevent overfitting in the process of training deep CNNs on small and unbalanced image datasets. However, for medical image analysis tasks, these techniques meet various problems [5, 9, 24]; the requirement for a more effective and more robust CBMIR system is still urgent.

In this paper, inspired by pioneering research works [9, 10, 14], we focus on 2D medical image retrieval and put forth an effort to alleviate the two main difficulties in CBMIR (i.e., (1) the labeled medical image datasets are commonly not large enough for training deep CNNs and (2) the imbalance problem is naturally attached to medical image datasets from clinic diagnosis). A new deep CNN-based 2D medical slice retrieval method is proposed, which can be effectively trained on relatively small labeled and unbalanced medical image dataset and promote the retrieval precision. First, in addition to commonly used methods for training deep CNNs on small and unbalanced datasets, e.g., dropout [25] and data augmentation, we supplement nonlinear components by using empirical mode decomposition (EMD) on 2D medical images to enhance effective information and reduce the image noise for training deep CNNs. Second, as for deep CNN architecture in this work, we employ residual network (ResNet) [19] as the backbone network adapted for learning different level features from medical images, which is combined with an attention mechanism to focus on the most relevant features by integrating local and global features in different scales [27]. And center loss function is combined with softmax loss function as a supervision signal in a deep CNN training process to facilitate nearest-neighbor similarity retrieval performance. The contributions of this paper are given as follows:

- (1) Nonlinear empirical mode decomposition on 2D medical images is proposed for supplementing effective information to original 2D medical images for better distinctively expressing 2D medical images
- (2) A residual network-based deep CNN model with attention and center loss modules is employed and trained on publicly available medical image datasets. The learned concise feature vectors are suitable for both classification-based and nearest-neighbor similarity-based medical image retrieval and show the great potential to handle large-scale medical image retrieval

## 2. Related Work

Among CBMIR literatures, there are two crucial factors that determine the performance of systems: (1) *Feature vector construction*: medical image features such as texture, shape, etc., should be extracted and formed into a vector to represent the query image and the images in datasets. (2) *Retrieval strategy*: classification-based retrieval strategy, nearest-neighbor search strategy, or their combination should be carefully chosen for different medical retrieval task.

**2.1. Hand-Crafted Features.** Hand-crafted features including texture features, keypoint-based features, local features, and global features are commonly used in CBMIR systems [5, 6, 8, 28, 29]. Jiang et al. [30] proposed a retrieval strategy that used mammographic region of interest (ROI) as query input, then retrieve breast tumor based on SIFT features. Caicedo et al. [31] used SIFT features to retrieve basal-cell carcinoma. Haas et al. [32] used SURF to capture the local texture of lung CTs for retrieval. Local Binary Patterns (LBPs) as local texture features were successfully used in ImageCLEFmed, 2D-Hela, and brain MRI retrieval tasks [33–35]. Xu et al. [36] proposed a corner-guided partial shape matching method that can dramatically increase the matching speed for spine X-ray image retrieval. Holistic features such as global GIST, global HOG, global color histogram, and moments were also used in medical image retrieval [37–41].

**2.2. Learned Features Using Deep CNNs.** In recent years, using features get through deep CNNs has achieved impressive results in generic image classification, object recognition, detection, retrieval, and other related tasks. But in the medical field, there is not much attention on exploring deep neural networks CBMIR task, partially because the amount of labeled medical images is typically limited. Qayyum et al. [10] proposed a CNN framework and trained the CNN on the medical image set they collected. Khatami et al. [9, 14] tried two retrieval strategies for medical image retrieval: the first method used one CNN model with transferred weights to shrink the search space and then used Radon projection to do similarity search. The second method employed multiple CNN models trained in a parallel way to get the shrunk search space. Bar et al. [42] used a pretrained CNN model from natural images for chest X-ray retrieval. Semedo and Magalhães [43] trained their CNN models on provided medical images in ImageCLEFmed 2016; they employed dropout

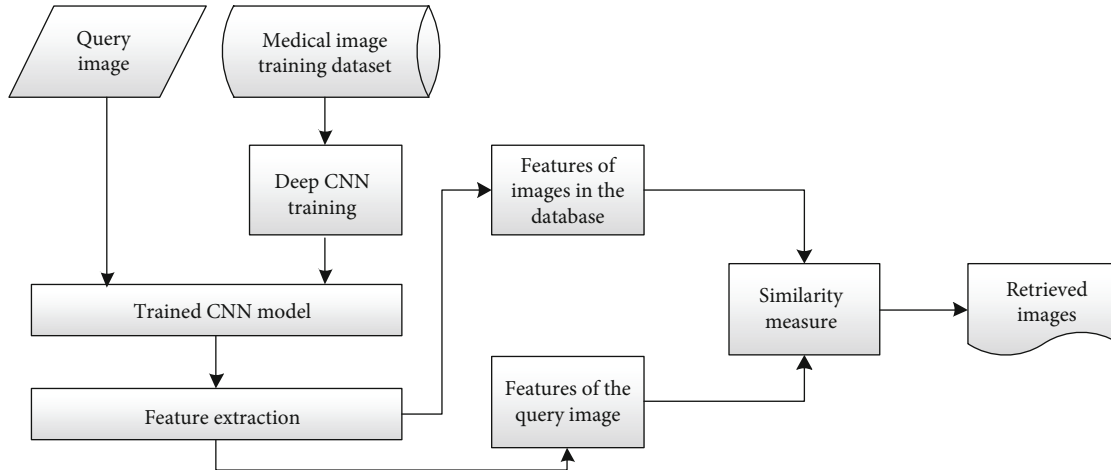


FIGURE 1: The proposed deep CNN-based content-based medical image retrieval flowchart.

and data augmentation to avoid overfitting. Hofmanninger and Langs [44] trained CNN using clinical routine images and radiology reports and carried out fine-tuning on current medical image retrieval task.

### 3. Methodology

There are pioneering studies that have been performed on deep CNNs for medical image retrieval and have shown promising results [9, 10, 14]; the problems of short of labeled images and highly imbalanced data distribution are still two main challenges for applying deep CNNs in medical image retrieval task [5]. There are also needs for more accurate and faster image retrieval methods for CBMIR [5]. To tackle these problems, in this work, we propose a multicomponent combined deep CNN framework for 2D medical image retrieval. The flowchart of content-based medical image retrieval is shown in Figure 1. This deep convolutional neural network is trained by a supervised learning way for classification and gets a concise feature vector for efficient nearest neighbor searching similar medical images. A brief description of the proposed framework is presented in the following sections.

**3.1. Processing 2D Medical Image with Empirical Mode Decomposition (EMD).** Empirical mode decomposition was originally introduced for the adaptive analysis of nonstationary and nonlinear time-domain signals and has become one of the most powerful tools for analyzing time-frequency (T-F) signal [45]. Then, EMD was extended to handle multidimensional data and acquired successful application in image tasks [46–49]. For image analysis, EMD is a fully data-adaptive multiresolution data analysis technique to decompose the multispatial resolution spatial-frequency-amplitude components of the image into a set of intrinsic mode functions (IMFs) [50, 51]. By advantage of the EMD principle, we can get multifrequency components (i.e., IMFs) of 2D medical images, and these frequencies are not pre-designed; these frequencies can self-adapt to different content of an image. Thus, we acquire nonstationary and nonlinear

multiresolution components of 2D medical images, which can provide supplementary information to commonly used spatial filter sets in image processing. EMD is implemented in an iterative process. First, a sifting process is used to find IMFs. Given a signal  $x(t)$ , Equation (1) is the process to get one IMF.

$$x(t) - \sum_{i=1}^k m_i = h_k \Rightarrow h_k = c_1, \quad (1)$$

where  $m_i$  is the local mean of the maxima and minima envelopes. These two envelopes are formed by connecting all local maxima or minima with a cubic spline. With the IMFs, the data  $x(t)$  can be decomposed by another sifting process:

$$x(t) = \sum_{j=1}^n c_j + r_n, \quad (2)$$

where  $c_j$  ( $j = 1$  to  $n$ ) is the IMFs and  $r_n$  is the final residual component. Figure 2 shows an example of a 2D X-ray image decomposed using EMD.

**3.2. The Proposed Medical Image Retrieval Method.** In this section, we introduce a deep CNN framework for medical image retrieval on a rather small dataset and with highly imbalanced data distribution. First, we discuss the network architecture employed in this work. Second, the supervision signal combining softmax loss function with center loss function to train deep CNN is discussed. Third, the training process is detailed. The proposed deep CNN framework is illustrated in Figure 3. For the input of the network, we employ original image and its IMF2, IMF3, and IMF4 components, because IMF1 contains mainly noise with quite high spatial frequency, and IMF5 contains the overall image intensity trend with very low spatial frequency. For medical image classification, IMF1 and IMF5 cannot provide useful structure information.

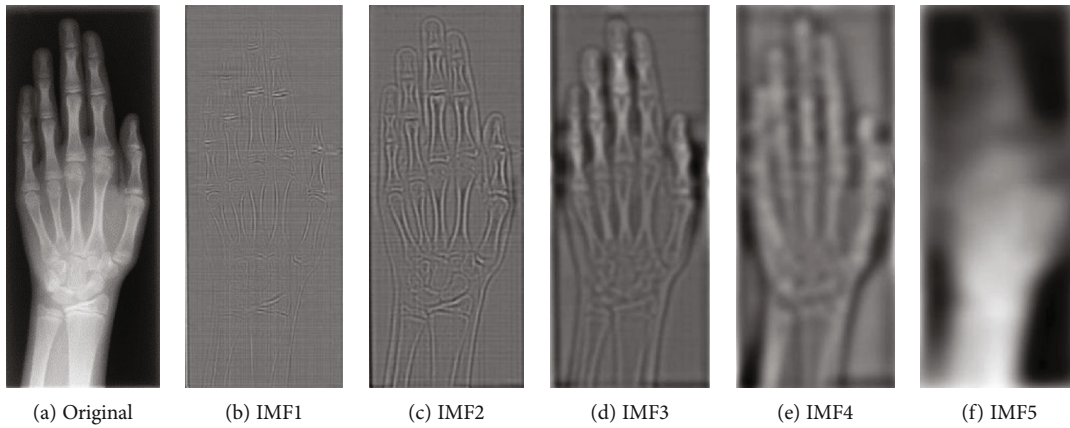


FIGURE 2: An example of a medical X-ray image is decomposed into five IMFs using empirical mode decomposition.

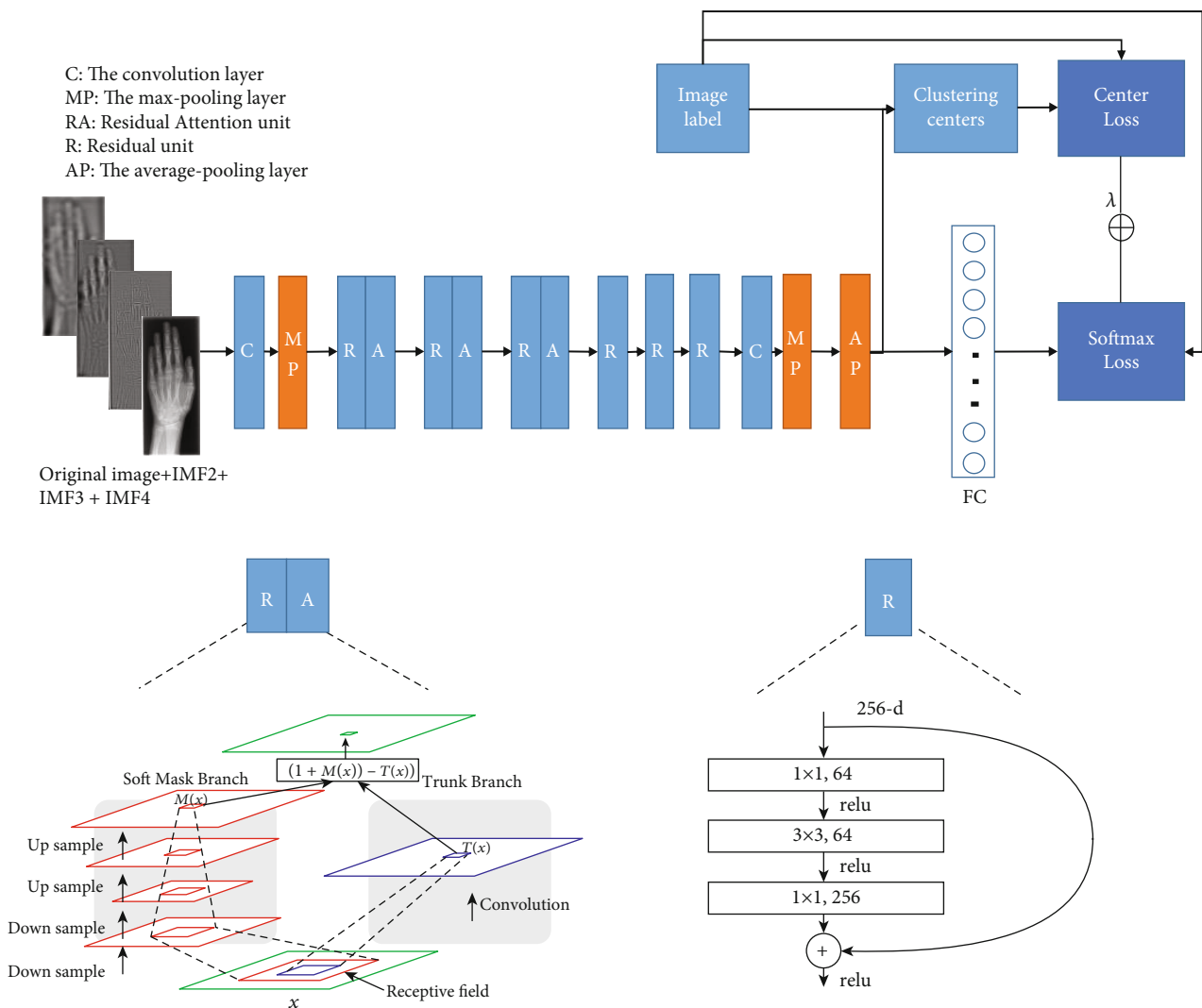


FIGURE 3: The proposed deep CNN framework for content-based medical image retrieval.

3.2.1. *The Network Architecture.* The proposed deep CNN architecture employs Residual Attention Network (RAN) [27] as the backbone network. In RAN, mixed attention activation function is used for both spatial and channel attention.

The attention mechanism was implemented as multiple attention modules, and each module consisted of a mask branch and a trunk branch, in which the mask branch was used to select good properties of original features and

TABLE 1: The details of deep CNN used for medical image retrieval.

Layer name	Output size	Layer
Conv1	$128 \times 128$	$7 \times 7, 64, \text{stride } 2$
Max pooling	$64 \times 64$	$3 \times 3 \text{ stride } 2$
Residual attention unit	$64 \times 64$	$\begin{pmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{pmatrix} \times 1$ Attention mask block $\times 1$
Residual attention unit	$32 \times 32$	$\begin{pmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 1512 \end{pmatrix} \times 1$ Attention mask block $\times 1$
Residual attention unit	$16 \times 16$	$\begin{pmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{pmatrix} \times 1$ Attention mask block $\times 1$
Residual unit	$8 \times 8$	$\begin{pmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{pmatrix} \times 3$
Conv2	$8 \times 8$	$3 \times 3, 32, \text{stride } 1, \text{padding: same}$
Max pooling	$4 \times 4$	$2 \times 2 \text{ stride } 1$
Average pooling	$1 \times 1$	$4 \times 4 \text{ stride } 1$
FC, softmax, Center loss		116
Trunk depth		57

suppress noises from trunk features. Residual learning was introduced in the learning process of RAN; the mask branch was constructed as identical mapping. With the residual learning, Residual Attention Network can go very deep, and the training process was much efficient. For medical image retrieval task, the nearest-neighbor similarity search is the most common way used to rank retrieved images. If the length of vector used to compute the similarity between two compared medical images is too long, the retrieval process will be very time consuming and cannot be used in practice. Thus, a dimensionality reduction model is added to get concise while strong distinguishing features. Table 1 details the CNN structure used in this work.

**3.2.2. Joint Loss Function.** Wen et al. [52] firstly introduced center loss function in deep CNN for face recognition task. In their work, center loss function was linearly jointed with softmax loss function to form a mixture supervision signal to train deep CNN. These two loss functions that were used in conjunction with each other can achieve discriminative feature learning, that is, the deeply learned features contained intraclass compactness and interclass dispersion. Discriminative features are very suitable for medical image classification and retrieval task in which nearest-neighbor similarity

search is most commonly used to accomplish the retrieve. Equation (3) formulates this joint loss function.

$$\text{Loss}_{\text{mixture}} = \text{Loss}_{\text{softmax}} + \lambda \text{Loss}_{\text{center}} - \sum_{i=1}^m \log \frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^n e^{W_{y_j}^T x_i + b_{y_j}}} + \frac{\lambda}{2} \sum_{i=1}^m \|x_i - c_{y_i}\|_2^2, \quad (3)$$

where the left part is the original softmax loss and the right part is the center loss. The  $c_{y_i}$  denotes the  $y_i$ th class center in the form of a feature vector. The parameter  $\lambda$  is empirically set as 0.002 in this paper's experiments.

**3.2.3. Network Training Setting.** As shown in Figure 3, the input of the network is the original medical image with its EMD components that contain IMF2, IMF3, and IMF4 got from EMD. The network training is developed and trained by using Keras on TensorFlow. The training processes are performed on a workstation with Ubuntu 18.04, having Intel(R) Xeon(R) Gold 6154 CPU with 256G RAM, and NVIDIA TITAN V graphic card with 12G RAM. Data



IRMA Code	1121-240-422-700
Technical code	X-ray, plain radiography, analog, overview image
Directional code	Sagittal, lateromedial
Anatomical code	Upper extremity (arm), radio carpal joint, right carpal joint
Biological code	Musculoskeletal system

FIGURE 4: A sample image (arm) with the corresponding IRMA code.

argumentation and dropout are employed in the training process. The number of epochs is 500, the batch size is 16, the initial learning rate is 0.0001, and early stopping is on. When the network accuracy is not improved within 20 training iterations, the early stopping mechanism will be triggered. The 500 epoch setting is to make sure that in most cases, the network training is stopped by the early stopping mechanism.

## 4. Experimental Results

In this paper, the very challenging IRMA dataset is chosen to evaluate the proposed framework and compare with other methods reported in the literature. The proposed CNN model is evaluated in terms of classification performance and retrieval performance, respectively.

**4.1. Database Description.** IRMA (Image Retrieval in Medical Applications) database is a well-known medical image dataset for content-based medical image retrieval research, which was made by Aachen University of Technology (RWTH) [53]. This dataset was arbitrarily selected from a routine at the Department of Diagnostic Radiology, Aachen University of Technology. IRMA code is used to specify each image's class along four independent hierarchical axes: TTTT-DDD-AAA-BBB. In this code, T represents the technical code (imaging modality), D represents the directional code (body orientations), A represents the anatomical code (the body region examined), and B represents the biological code (the biological system examined). This dataset contains a total of 12,000 images divided into 116 classes, 11,000 image radiographs with known categories for training, and the rest 1000 radiographs as test. Figure 4 illustrates a sample image with the corresponding IRMA code.

### 4.2. Classification Performance

**4.2.1. IRMA Error.** ImageCLEF07 proposed the error evaluation procedure for IRMA Medical Image Annotation to cal-

culate the retrieval error [54, 55]. The total IRMA error can be computed by the following formula:

$$\sum_{i=1}^I \frac{1}{b_i} \frac{1}{i} \delta(l_i, \hat{l}_i) \text{ with } \delta(l_i, \hat{l}_i) = \begin{cases} 0, & \text{if } l_j = \hat{l}_j, \forall j \leq i, \\ 0.5, & \text{if } l_j \neq \hat{l}_j, \exists j \leq i, \\ 1, & \text{if } l_j \neq \hat{l}_j, \exists j \leq i. \end{cases} \quad (4)$$

Here,  $l_1^I = l_1, l_2, \dots, l_i, \dots, l_I$  is the *correct* code (for one axis) of an image, and  $\hat{l}_1^I = \hat{l}_1, \hat{l}_2, \dots, \hat{l}_i, \dots, \hat{l}_I$  is the *classified* code (for one axis) of an image.  $I$  is the depth of the tree to which the classification is specified. If there is an incorrect classification at position  $\hat{l}_i$ , all succeeding decisions will be considered as wrong decisions.

**4.2.2. Commonly Used Classification Performance Measure.** To evaluate the performance of different methods for classification task, commonly used performance evaluation indicators include average precision (AP), average recall (AR), and F1 measure. These indicators are calculated as the following:

$$\begin{aligned} \text{AP} &= \frac{1}{M} \sum_{i=1}^M \frac{\text{TP}_i}{\text{TP}_i + \text{FP}_i}, \\ \text{AR} &= \frac{1}{M} \sum_{i=1}^M \frac{\text{TP}_i}{\text{TP}_i + \text{TN}_i}, \\ \text{F1 measure} &= 2 \times \frac{\text{AP} \times \text{AR}}{\text{AP} + \text{AR}}, \end{aligned} \quad (5)$$

where TP is true positive, indicating the number of images correctly classified as class  $k$ ; FP is false positive, indicating the number of images misclassified as class  $k$ ; TN is true negative, indicating the number of images correctly classified as not class  $k$ ; FN is false negative, indicating the number of images misclassified as not class  $k$ ; and  $M$  means the total

TABLE 2: Comparison of our classification performance to other CNN reported in the literature for IRMA images. The IRMA error with \* was the best test value selected from the literature.

Methods	IRMA error
Proposed method	68.48
Parallel shrink CNN+Radon [9]	165.55*
Sequential shrink CNN+LBP [14]	168.05*
CNN+Radon [56]	210.35
CNNC+RBC [57]	224.13

TABLE 3: Comparison of classification performance of the proposed framework with other deep models and state-of-the-art on IRMA images.

Methods	IRMA error	AP	AR	F1 measure
VGG16 [18]	115.08	0.56	0.56	0.53
ResNet50 [19]	80.80	0.65	0.64	0.63
AttentionResNet56 [27]	76.83	0.65	0.66	0.64
AttentionResNet56+center loss [52]	73.85	0.68	0.66	0.66
Proposed method	68.48	0.67	0.67	0.66

number of classes that is 116 IRMA classes in this paper. As the F1 measure is more sensitive to data distribution, it is a suitable measure for classification problems on imbalanced datasets [10].

**4.2.3. Classification Performance and Comparison.** The performance of the proposed single-model framework for medical image classification is evaluated by the IRMA error and commonly used measures for image classification methods, which are detailed in Sections 4.2.1 and 4.2.2. Table 2 compares the IRMA error got by the proposed framework and several deep CNN-based methods reported in the literature [9, 14]. Table 2 shows that with the fast development of the deep CNN technique, much better classification accuracy (i.e., lower IRMA error score) can be gotten by employing a more powerful CNN model as a backbone network. In terms of the IRMA error, our proposed framework gets a much lower score than referenced deep CNN-based methods reported in the literature.

Considering the relative lag of the technology applied on IRMA dataset and the rapid development of the deep CNNs in computer vision area, Table 3 compares the classification accuracy measures on the IRMA dataset including IRMA error, AP, AR, and F1 measure of the proposed method with various state-of-the-art deep CNNs including VGG [18], ResNet [19], and AttentionResNet [27] that have achieved a very high recognition score on large image dataset challenges (such as ImageNet [58] and CoCo [59]). Table 3 shows that the proposed framework performs better in classifying IRMA images. The proposed framework and the compared deep CNNs are trained under the same condition, that is, using the same training dataset, same image argumentation strategy, same number of epochs, same learning rate, and so on.

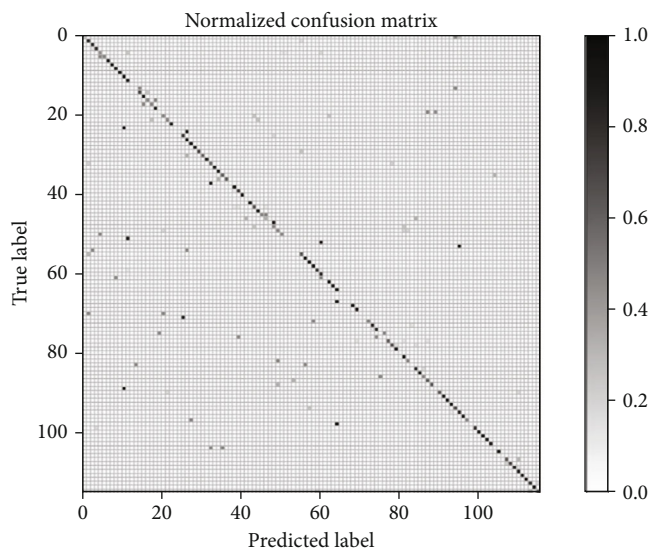


FIGURE 5: Confusion matrix of IRMA image classification with 116 classes using the proposed deep CNN.

TABLE 4: Comparison of our retrieval performance to other CNN reported in the literature for IRMA images. The proposed method used Cosine similarity measure to get the IRMA error in this table.

Methods	IRMA error
<i>Proposed method</i>	43.21
SVM+multiscale LBP [61]	146.55
Parallel shrink CNN+LBP, HOG, Radon [9]	165.55
Sequential shrink CNN+LBP [14]	168.05
TAUbiomed [60]	169.5
Diap [60]	178.93
CNN+Radon [56]	210.35
CNNC+RBC [57]	224.13
FEITIJS [60]	242.46
SuperPixel [57]	249.34
VPA [60]	261.16
SP-R [57]	311.8
MedGIFT [60]	317.53
SP-RBC [57]	356.57
IRMA [60]	359.29
MedGIFT [60]	420.91

For classification-based medical image retrieval, the retrieval performance depends entirely on the accuracy of classification, the higher classification accuracy means the better retrieval performance. As in Table 3, our proposed framework achieved the lowest IRMA error and the best F1 measure.

The confusion matrix is shown in Figure 5, where most classes can be classified rightly. There are 38.2% classes with accuracy better than 90%, 51.2% classes with accuracy better than 80%, and 59.1% classes with accuracy better than 70%.

TABLE 5: Comparison of retrieval performance of the proposed framework with other deep models and state-of-the-art on IRMA images. The IRMA error with \* was the best test value selected from the literature.

Methods	Vector dimension for similarity retrieval	IRMA error on Cosine similarity	Euclidean distance	Manhattan distance	Cosine similarity
Sequential shrink CNN+LBP [14]	8496	168.05*	—	—	—
Parallel shrink CNN+LBP, HOG, Radon [9]	8496 (LBP)	165.55*	—	—	—
	3528 (HOG)				
SVM+multiscale LBP [61]	1800 (Radon)	146.55*	—	—	—
	$N$ subblocks $\times 4 \times 4 \times 1062$ on SVM				
VGG16 [18]	512	65.53	0.53	0.53	0.53
AttentionResNet56 [27]	2048	49.66	0.45	0.45	0.62
AttentionResNet56+center loss [52]	2048	47.76	0.35	0.35	0.53
ResNet50 [19]	2048	47.53	0.72	0.72	0.73
Proposed method	32	43.21	0.84	0.85	0.86



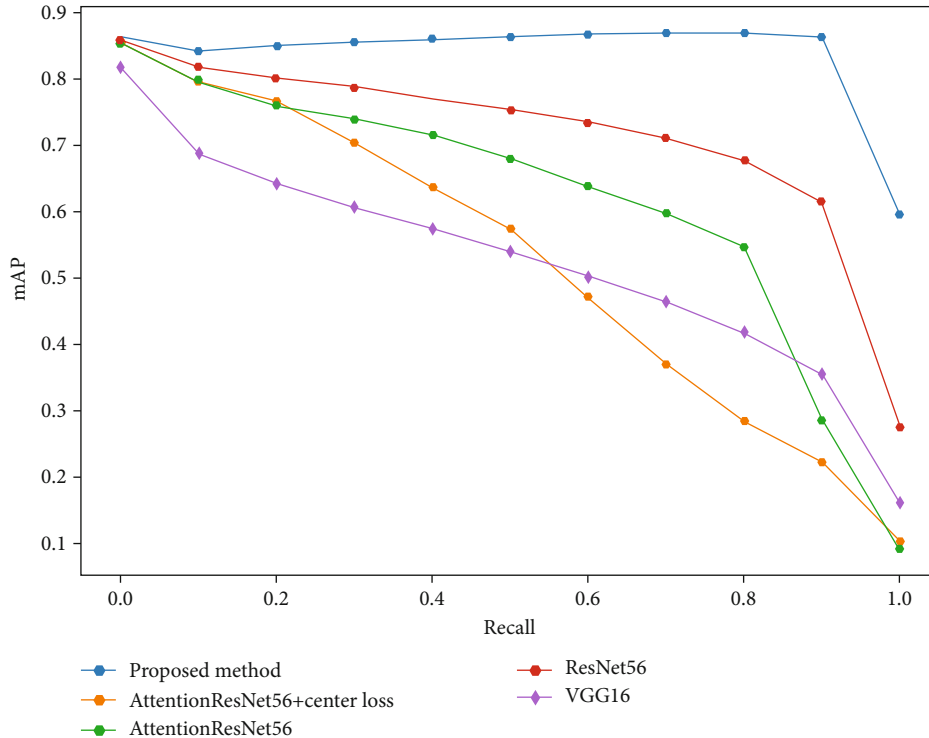


FIGURE 6: mAP vs. recall for medical retrieval on IRMA images.

4.3. Retrieval Performance

4.3.1. Retrieval Performance Measure. Precision and recall are two measures commonly used as retrieval performance evaluation measures [5].

$$\begin{aligned}
 P &= \frac{\text{Number of relevant images retrieved}}{\text{Total number of images retrieved}}, \\
 R &= \frac{\text{Number of relevant images retrieved}}{\text{Total number of relevant images}}.
 \end{aligned}
 \tag{6}$$

Besides precision and recall, mean average precision (MAP) is a very popular evaluation metric for algorithms that do search in medical image sets [5]. MAP combines precision and recall in one number. It is defined as the mean of average precision (AP) metric over all queries that can alleviate the bias during precision evaluation. The AP and mAP can be formulated as the following:

$$AP(q) = \frac{1}{N_R} \sum_{n=1}^{N_R} P_q(R_n),
 \tag{7}$$

where  $P_q(R_n)$  is the precision value when the recall value is  $R_n$  and  $N_R$  indicates the top  $N_R$ -ranked relevant images for the query image  $q$ .

$$mAP = \frac{1}{|Q|} \sum_{q \in Q} AP(q),
 \tag{8}$$

where  $Q$  is the query image set and  $|Q|$  is the number of the query image set.

4.3.2. Retrieval Performance and Comparison. In the proposed deep CNN framework, the feature vector for nearest-neighbor similarity searching of medical images is gotten from the last fully connected layer. For comparison, the proposed framework retrieval performance on the IRMA dataset is evaluated using both the IRMA error and the mean average precision (mAP). The calculation of the IRMA error in image retrieval follows the nearest-neighbor rule, that is, the query image’s class label is determined by the most similar image returned in the retrieval process. Table 4 compares the retrieval performance achieved by the proposed framework with the other methods reported in the literature [9, 14, 56, 57, 60, 61] on the IRMA dataset with the IRMA error. The proposed deep CNN framework gets the lowest IRMA error in nearest-neighbor similarity retrieval. Table 5 compares the proposed framework with state-of-the-art deep CNNs on the IRMA error and mAP. For mAP, we test three usually used distance/similarity measures in image retrieval: Euclidean distance, Manhattan distance, and Cosine similarity, and the IRMA error is evaluated by using the best distance/similarity measure: Cosine similarity. Table 5 shows that the proposed deep CNN framework gets the best mAP and the lowest IRMA error on these three distance/similarity measures and gets the highest score on Cosine similarity. In Table 5, we also list the vector length used for similarity retrieval. The feature vector for retrieval gotten from the proposed framework is just 32 dimensions that are much shorter than output vectors reported in literatures and state-of-the-

TABLE 6: Comparison of classification and retrieval performance of the proposed framework with other deep models and state-of-the-art on IRMA images with and without EMD components.

Methods	Classification				Retrieval (on Cosine similarity)			
	IRMA error		F1 measure		IRMA error		mAP	
	With EMD	Without EMD	With EMD	Without EMD	With EMD	Without EMD	With EMD	Without EMD
VGG16 [18]	115.08	98.29	0.53	0.52	65.53	57.18	0.53	0.54
ResNet50 [19]	80.8	92.62	0.63	0.55	47.53	62.2	0.73	0.70
AttentionResNet56 [27]	76.83	81.54	0.64	0.59	49.66	55.03	0.62	0.61
AttentionResNet56+center loss [52]	73.85	74.81	0.66	0.62	47.76	47.13	0.54	0.54
Proposed method	68.48	77.45	0.66	0.59	43.21	46.81	0.86	0.85

art deep CNNs, which illustrate the great potential of our method to implement large-scale medical retrieval. Suppose  $\mathbf{a}$  and  $\mathbf{b}$  are two feature vectors representing two medical images, the three distance/similarity measures are formulated as the following:

$$\begin{aligned} \text{Euclidean distance}(\mathbf{a}, \mathbf{b}) &= \sqrt{\sum_{i=1}^n (a_i - b_i)^2}, \\ \text{Manhattan distance}(\mathbf{a}, \mathbf{b}) &= \sum_{i=1}^n |a_i - b_i|, \\ \text{Cosine similarity}(\mathbf{a}, \mathbf{b}) &= \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \cdot \|\mathbf{b}\|} \end{aligned} \quad (9)$$

Figure 6 summarizes the retrieval performance of the proposed framework and state-of-the-art deep CNNs by the mAP-recall curve. And all these curves are calculated using the Cosine similarity measure.

**4.4. Performance Comparison with and without EMD Components.** To illustrate the effect of EMD components, Table 6 details the classification and retrieval measures between the proposed framework and the state-of-the-art deep CNNs with and without using EMD components. The results show that with EMD components, we can get higher performance in both classification and retrieval applications. With EMD components, deep CNNs can consistently achieve better classification and retrieval performance than without EMD components except for VGG16 on the IRMA error. This may be because the ResNet backbone is deeper than VGG16, so the CNNs based on the ResNet backbone can effectively handle more image information.

## 5. Conclusions

This paper has proposed a deep convolutional neural network for medical image retrieval task. By training deep CNN with input medical image and its multifrequency components (i.e., IMFs get from empirical mode decomposition (EMD)) in a supervised classification way, we have got a scheme that is very suitable for similarity-based medical image retrieval. Using an imbalanced IRMA medical image dataset, the proposed framework has surpassed existing algo-

rithms with the highest classification accuracy and lowest retrieval error. The concise and distinguishable feature vector output from the proposed deep CNN has also shown great potential to handle large-scale medical image retrieval. We intend to further examine CBMIR on other medical datasets, different modalities, and 3D volumetric applications.

## Data Availability

The IRMA2007 data used to support the findings of this study are available in a public repository at [http://publications.rwth-aachen.de/search?ln=en&cc=Dataset&sc=1&p=IRMA&f=&action\\_search=Search](http://publications.rwth-aachen.de/search?ln=en&cc=Dataset&sc=1&p=IRMA&f=&action_search=Search).

## Conflicts of Interest

The authors declared that they have no conflicts of interest in this work.

## Acknowledgments

The authors would like to thank Bo Wen and Guihui Liu for their help for some experiments. This research is sponsored by the National Natural Science Foundation of China (Nos. 61561002 and 62062003), Natural Science Foundation of Ningxia (No. 2020AAC03213), Ningxia Medical Imaging Clinical Research Center Innovation Platform Construction Project (No. 2018DPG05006), "Image and Intelligent Information Processing Innovation Team" the State Ethnic Affairs Commission Innovation Team (Nos. PY1905 and PY1606), Ningxia Key Research and Development Project (special projects for talents) (2020BEB04022), North Minzu University Research Project of Talent Introduction under Grant 2020KYQD08, and General Research Project of North Minzu University (No. 2021XYZJK04).

## References

- [1] K. Doi, "Computer-aided diagnosis in medical imaging: historical review, current status and future potential," *Computerized Medical Imaging and Graphics*, vol. 31, no. 4-5, pp. 198-211, 2007.
- [2] H. Zaidi, H. R. Veves, and M. Wissmeyer, "Molecular PET/CT imaging-guided radiation therapy treatment planning," *Academic Radiology*, vol. 16, no. 9, pp. 1108-1133, 2009.

- [3] C. D. Marcus, V. Ladam-Marcus, C. Cucu, O. Bouché, L. Lucas, and C. Hoeffel, "Imaging techniques to evaluate the response to treatment in oncology: current standards and perspectives," *critical reviews in oncology hematology*, vol. 72, no. 3, pp. 217–238, 2009.
- [4] C. E. Kahn, J. A. Carrino, M. J. Flynn, D. J. Peck, and S. C. Horii, "DICOM and radiology: past, present, and future," *Journal of the American College of Radiology*, vol. 4, no. 9, pp. 652–657, 2007.
- [5] Z. Li, X. Zhang, H. Müller, and S. Zhang, "Large-scale retrieval for medical image analytics: a comprehensive review," *Medical Image Analysis*, vol. 43, pp. 66–84, 2018.
- [6] H. Müller, N. Michoux, D. Bandon, and A. Geissbuhler, "A review of content-based image retrieval systems in medical applications—clinical benefits and future directions," *International Journal of Medical Informatics*, vol. 73, no. 1, pp. 1–23, 2004.
- [7] T. M. Lehmann, M. O. Güld, C. Thies et al., "Content-based image retrieval in medical applications," *Methods of Information in Medicine*, vol. 43, no. 4, pp. 354–361, 2004.
- [8] C. B. Akgül, D. L. Rubin, S. Napel, C. F. Beaulieu, H. Greenspan, and B. Acar, "Content-based image retrieval in radiology: current status and future directions," *Journal of digital imaging*, vol. 24, no. 2, pp. 208–222, 2011.
- [9] A. Khatami, M. Babaie, A. Khosravi, H. R. Tizhoosh, and S. Nahavandi, "Parallel deep solutions for image retrieval from imbalanced medical imaging archives," *Applied Soft Computing*, vol. 63, pp. 197–205, 2018.
- [10] A. Qayyum, S. M. Anwar, M. Awais, and M. Majid, "Medical image retrieval using deep convolutional neural network," *Neurocomputing*, vol. 266, pp. 8–20, 2017.
- [11] M. Srinivas, R. R. Naidu, C. S. Sastry, and C. K. Mohan, "Content based medical image retrieval using dictionary learning," *Neurocomputing*, vol. 168, pp. 880–895, 2015.
- [12] L. Pan, Y. Qiang, J. Yuan, and L. Wu, "Rapid retrieval of lung nodule CT images based on hashing and pruning methods," *Biomed Research International*, vol. 2016, 10 pages, 2016.
- [13] H.-C. Shin, K. Roberts, D. D.-F. Le Lu, J. Yao, and R. M. Summers, "Learning to read chest X-rays: recurrent neural cascade model for automated image annotation," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2497–2506, Las Vegas, NV, USA, 2016.
- [14] A. Khatami, M. Babaie, H. R. Tizhoosh, A. Khosravi, T. Nguyen, and S. Nahavandi, "A sequential search-space shrinking using CNN transfer learning and a Radon projection pool for medical image retrieval," *expert systems with applications*, vol. 100, pp. 224–233, 2018.
- [15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.
- [16] A. Khatami, Y. Tai, A. Khosravi et al., "A deep learning-based model for tactile understanding on haptic data percutaneous needle treatment," in *International Conference on Neural Information Processing*, pp. 317–325, Springer, 2017.
- [17] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, The MIT Press, 2016.
- [18] K. Simonyan and A. Zisserman, *Very deep convolutional networks for large-scale image recognition*, International Conference on Learning Representations, 2015.
- [19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
- [20] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.
- [21] H. Greenspan, B. Van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1153–1159, 2016.
- [22] A. A. A. Setio, F. Ciompi, G. Litjens et al., "Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1160–1169, 2016.
- [23] H. R. Roth, L. Lu, J. Liu et al., "Improving computer-aided detection using convolutional neural networks and random view aggregation," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1170–1181, 2016.
- [24] H. C. Shin, M. R. Orton, D. J. Collins, S. J. Doran, and M. O. Leach, "Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4D patient data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1930–1943, 2013.
- [25] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [26] H. C. Shin, H. R. Roth, M. Gao et al., "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.
- [27] F. Wang, M. Jiang, C. Qian et al., *Residual attention network for image classification* 6450–6458.
- [28] H. K. Hoon, L. Haejun, and C. Duckjoo, "Medical image retrieval: past and present," *Healthcare Informatics Research*, vol. 18, no. 1, pp. 3–9, 2012.
- [29] A. Kumar, J. Kim, W. Cai, M. Fulham, and D. Feng, "Content-based medical image retrieval: a survey of applications to multidimensional and multimodality data," *Journal of Digital Imaging*, vol. 26, no. 6, pp. 1025–1039, 2013.
- [30] M. Jiang, S. Zhang, H. Li, and D. N. Metaxas, "Computer-aided diagnosis of mammographic masses using scalable image retrieval," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 2, pp. 783–792, 2015.
- [31] J. C. Caicedo, A. Cruz, and F. A. Gonzalez, "Histopathology image classification using bag of features and kernel functions," *Artificial Intelligence in Medicine in Europe*, pp. 126–135, 2009.
- [32] S. Haas, R. Donner, A. Burner, M. Holzer, and G. Langs, "Superpixel-based interest points for effective bags of visual words medical image retrieval," in *MICCAI International Workshop on Medical Content-Based Retrieval for Clinical Decision Support*, pp. 58–68, Springer, Berlin, Heidelberg, 2011.
- [33] H. Müller, A. G. de Herrera, J. Kalpathy-Cramer, D. Demner-Fushman, S. K. Antani, and I. Eggel, *Overview of the ImageCLEF 2012 medical image retrieval and classification tasks*, Cross-Language Evaluation Forum, 2012.
- [34] L. Nanni, A. Lumini, and S. Brahmam, "Local binary patterns variants as texture descriptors for medical image analysis," *Artificial intelligence in medicine*, vol. 49, no. 2, pp. 117–125, 2010.
- [35] S. Murala, R. P. Maheshwari, and R. Balasubramanian, "Directional binary wavelet patterns for biomedical image indexing

- and retrieval,” *Journal of Medical Systems*, vol. 36, no. 5, pp. 2865–2879, 2012.
- [36] X. Xu, D. Lee, S. Antani, and L. R. Long, “A spine X-ray image retrieval system using partial shape matching,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 1, pp. 100–108, 2008.
- [37] J. Liu, S. Zhang, W. Liu, X. Zhang, and D. N. Metaxas, “Scalable mammogram retrieval using anchor graph hashing,” in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, pp. 898–901, Beijing, China, 2014.
- [38] M. Jiang, S. Zhang, J. Huang, L. Yang, and D. N. Metaxas, “Scalable histopathological image analysis via supervised hashing with multiple features,” *Medical image analysis*, vol. 34, pp. 3–12, 2016.
- [39] L. Song, X. Liu, L. Ma, C. Zhou, X. Zhao, and Y. Zhao, “Using HOG-LBP features and MMP learning to recognize imaging signs of lung lesions,” in *2012 25th IEEE International Symposium on Computer-Based Medical Systems (CBMS)*, pp. 1–4, Rome, Italy, 2012.
- [40] K. Bunte, M. Biehl, M. F. Jonkman, and N. Petkov, “Learning effective color features for content based image retrieval in dermatology,” *Pattern Recognition*, vol. 44, no. 9, pp. 1892–1902, 2011.
- [41] M. M. Rahman, P. Bhattacharya, and B. C. Desai, “A framework for medical image retrieval using machine learning and statistical similarity matching techniques with relevance feedback,” *International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 11, no. 1, pp. 58–69, 2007.
- [42] Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen, and H. Greenspan, “Chest pathology detection using deep learning with non-medical training,” in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pp. 294–297, New York, NY, USA, 2015.
- [43] D. Semedo and J. Magalhães, *NovaSearch at ImageCLEFmed 2016 subfigure classification task*, pp. 386–398, 2016.
- [44] J. Hofmanninger and G. Langs, “Mapping visual features to semantic profiles for retrieval in medical imaging,” *Computer Vision and Pattern Recognition*, pp. 457–465, 2015.
- [45] N. E. Huang, Z. Shen, S. R. Long et al., “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis,” *Proceedings of the royal society a mathematical physical and engineering sciences*, vol. 454, no. 1971, pp. 903–995, 1998.
- [46] J. C. Nunes, Y. Bouaoune, E. Delechelle, O. Niang, and P. Bunel, “Image analysis by bidimensional empirical mode decomposition,” *Image and vision computing*, vol. 21, no. 12, pp. 1019–1026, 2003.
- [47] K. Oberleithner, M. Sieber, C. N. Nayeri et al., “Three-dimensional coherent structures in a swirling jet undergoing vortex breakdown: stability analysis and empirical mode construction,” *Journal of fluid mechanics*, vol. 679, pp. 383–414, 2011.
- [48] S. Al-Baddai, P. Marti-Puig, E. Gallego-Jutglà et al., “A recognition-verification system for noisy faces based on an empirical mode decomposition with Green’s functions,” vol. 24, no. 5, pp. 3809–3827, 2020.
- [49] H.-S. Oh and D. Kim, “Image decomposition by bidimensional ensemble patch transform,” vol. 135, pp. 173–179, 2020.
- [50] Z. Wu, N. E. Huang, and X. Chen, “The multi-dimensional ensemble empirical mode decomposition method,” *Advances in Adaptive Data Analysis*, vol. 1, no. 3, pp. 339–372, 2011.
- [51] Y. H. Wang, C. H. Yeh, H. W. Young, K. Hu, and M. T. Lo, “On the computational complexity of the empirical mode decomposition algorithm,” *Physica a statistical mechanics and its applications*, vol. 400, pp. 159–167, 2014.
- [52] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, “A discriminative feature learning approach for deep face recognition,” in *European Conference on Computer Vision*, pp. 499–515, Cham, 2016.
- [53] T. M. Lehmann, M. O. Güld, T. Deselaers et al., “Automatic categorization of medical images for content-based retrieval and data mining,” *Computerized Medical Imaging and Graphics*, vol. 29, no. 2-3, pp. 143–155, 2005.
- [54] T. M. Lehmann, M. O. Güld, C. Thies et al., “IRMA—content-based image retrieval in medical applications,” *Studies in health technology and informatics*, vol. 107, pp. 842–846, 2004.
- [55] T. M. Lehmann, H. Schubert, D. Keysers, M. Kohnen, and B. B. Wein, “The IRMA code for unique classification of medical images,” *Medical Imaging 2003: PACS and Integrated Medical Information Systems: Design and Evaluation*, vol. 5033, pp. 440–451, 2003.
- [56] A. Khatami, M. Babaie, A. Khosravi, H. R. Tizhoosh, S. M. Salaken, and S. Nahavandi, “A deep-structural medical image classification for a Radon-based image retrieval,” in *Canadian Conference on Electrical and Computer Engineering*, pp. 1–4, Windsor, ON, Canada, 2017.
- [57] X. Liu, H. R. Tizhoosh, and J. Kofman, “Generating binary tags for fast medical image retrieval based on convolutional nets and Radon transform,” in *International Joint Conference on Neural Network*, pp. 2872–2878, Vancouver, BC, Canada, 2016.
- [58] O. Russakovsky, J. Deng, H. Su et al., “ImageNet large scale visual recognition challenge,” *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.
- [59] T. Y. Lin, M. Maire, S. Belongie et al., “Microsoft COCO: common objects in context,” in *European Conference on Computer Vision*, pp. 740–755, Cham, 2014.
- [60] T. Tommasi, B. Caputo, P. Welter, M. O. Güld, and T. M. Deserno, “Overview of the CLEF 2009 medical image annotation track,” in *Workshop of the Cross-Language Evaluation Forum for European Languages*, pp. 85–93, Springer, Berlin, Heidelberg, 2009.
- [61] Z. Çamlica, H. R. Tizhoosh, and F. Khalvati, “Medical image classification via SVM using LBP features from saliency-based folded data,” in *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, pp. 128–132, Miami, FL, USA, 2015.