Convolutional Neural Network for Medical Image Classification using Wavelet Features

Amin Khatami*, Asef Nazari†, Amin Beheshti‡, Thanh Thi Nguyen§, Saeid Nahavandi¶, Jerzy Zieba‖
*¶IISRI Deakin University, Geelong, Australia
†§School of Information Technology, Deakin University, Geelong, Australia
‡ Department of Computing, Macquarie University, Australia
‖School of Medical Sciences,University of New South Wales, Sydney, Australia
Email: {*amin.khatami,†asef.nazari,§thanh.nguyen, ¶saeid.nahavandi}@deakin.edu.au,
‡amin.beheshti@mq.edu.au, ‖jerzy.zieba@unsw.edu.au

***Abstract*—Automatic classification algorithms are an important component of expert decision support systems that are used in a number of medical applications including diagnostic radiology and disease detection. This study proposes a deep learning-based framework for medical image classification using wavelet features. Convolutional neural networks are incorporated to discover informative latent patterns and features from a set of X-ray images pertaining to human body parts. The features are then passed to a classifier for labelling the respective X-ray images. The experimental results show that the low-pass filter wavelet-based convolutional model outperforms the original convolutional network and some models for classifying X-ray images. The performance of the proposed method implies that it can be implemented effectively in practice for disease detection using radiological images.**

***Index Terms*—Convolutional neural network, deep learning, classification, wavelet, medical imaging**

I. Introduction

Medical image analysis deals with the extraction of the most important features in health-related images to improve clinical diagnosis. Correctly inferring positive and negative conclusions based on medical image classification models leads to a reliable content-based image retrieval (CBIR) system [14], [52]. A wealth of studies have been conducted to examine different classification algorithms that result in an appropriate and effective retrieval scheme [3], [4], [12], [37], [42], [45], [47].

In this regard, a hierarchical scheme for the characterisation of human perceptual similarity in a CBIR system was presented in [10]. Likewise, Ko et al. [21] studied a CBIR system using the random forest classifier by utilising the features extracted from wavelet-based symmetric binary patterns. Additionally, Qasem et al. [40] introduced a radial basis function network that evolves its centres and weights simultaneously using a multi-objective optimisation method to increase classification performance for medical disease diagnosis. The method was applied to retrieve clinical mammograms. Krawczyk et al. [23] developed a method to analyse breast thermograms based on image features. They proposed a hybrid multiple classifier technique using neural networks and support vector machines (SVM) and devised a fuzzy measure to assess the diversity of the ensemble. Also, Tsochatzidis et al. [49] incorporated CBIR into computer-aided diagnosis system to help radiologists in the mammographic masses characterisation process. In this work, an ensemble of SVMs was used during the retrieval stage to exploit the margin and pathology type likelihood in input samples.

Among a variety of public domain image databases, ImageCLEFmed 2005 dataset [53] is a popular benchmark collection for categorization and retrieval of automated medical images [54]. As illustrated in Fig. 1, it is extremely imbalanced considering the distribution of train and test set sizes. This database consists of 10,000 X-ray images including 57 categories captured by plain radiographies and classified by the Image Retrieval in Medical Application (IRMA) group of the University Hospital, Aachen, Germany [27].

The ImageCLEFmed 2005 dataset has been widely used in literature for examining the performance of different classification methods. Particularly, a two-level hierarchical merging model utilising different shape and texture features was introduced in [39] for classifying images in this dataset. Nine invariant types of features were extracted individually, and the best one for each class was identified through a weighted scheme. Alternatively, Rahman et al. [41] introduced a method to retrieve medical images using image filtering, similarity fusion, and a relevance feedback procedure. They also considered supervised and unsupervised methods

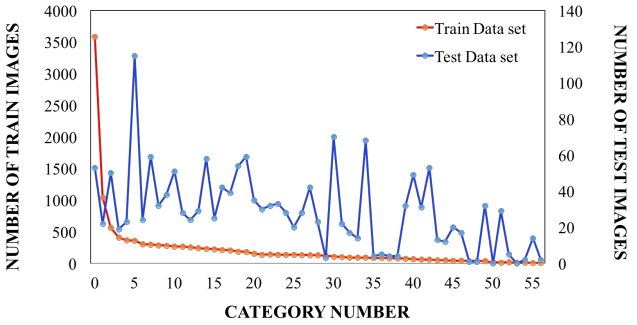

Fig. 1: Imbalanced training and testing sets of ImageCLEFmed 2005 dataset

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

to associate low-level and high-level features in the image retrieval process. On the other hand, Mohammadi et al. [33] proposed a shape-texture feature extraction method to classify radiological images. Different features and classifiers were examined based on 4,402 X-ray images in 21 out of 57 classes of the ImageCLEFmed 2005 database.

Designing a deep convolutional neural network for a specific problem is a nontrivial task because of a large number of hyperparameters and algorithmic choices [2], [18], [43], [44]. In addition, the lack of interpretability of the CNNs brings the opportunity to combine them with some well-studied techniques from signal processing. The discrete wavelet transform (DWT) can decompose an image into different frequency levels which is useful in improving the feature extraction of a CNN. Theoretical properties of wavelets in multiresolution signal processing are well considered in CNNs [6], [13], [17], [19], [20], [28], [46]. For example, Williams et al. [50] fed the wavelet sub-bands of the original images as a new input to CNNs. In addition, Oyallon et al. [36] proposed a wavelet scattering network instead of the first layer of ResNet. This hybrid network has the advantage of showing comparable performance with ResNet with a considerably smaller number of hyperparameters. In another work, Lu et al. [29] studied an augmented CNN a dual-tree wavelet transform to solve the organ tissue segmentation problem. Furthermore, Williams et al. [51] used a second-level wavelet decomposition to subsample features as a wavelet pooling algorithm. [25] proposed a combination of CNN, DWT, and long short term memory (LSTM) techniques in liver and brain tumor detection. Also, [46] investigated a deep neural network by integrating multiresolution analysis originating from utilising DWT. The main motivation of this study is to investigate an effective preprocessing stage to extract meaningful information for CNNs.

In this work, we use wavelet decomposition to deep structural networks due to its ability to remove noise and outliers from images. This combination proposes a robust classifier for a CBIR system, classifying 35 categories of the ImageCLEFmed 2005 benchmark. Therefore the main contribution of this study is a CNN-based classification model which is fed by approximate features extracted by a wavelet transform. The rest of the article is organized as follows. Descriptions of the technical and fundamental aspects of the proposed model and other competing methods are presented in Section 2. Section 3 reports and discusses the experimental results. Performance comparison between the proposed method with some existing methods in the literature is presented in Section 4, followed by conclusions and future work in Section 5.

II. The Proposed Medical Image Classification Approach

This study introduces a system to categorise 35 classes of images corresponding to 35 different parts of human body in the ImageCLEFmed 2005 dataset. The system includes three fundamental stages. The first stage investigates the effect of

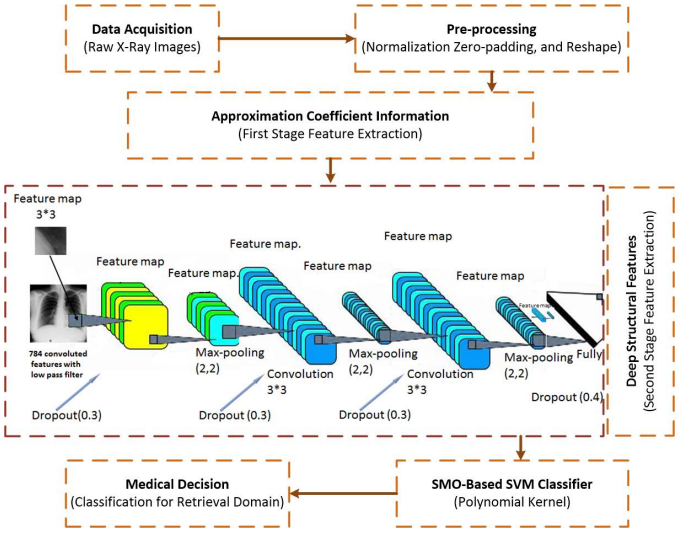

Fig. 2: The proposed method: Wavelet filter and CNN.

information extracted by a wavelet decomposition. A low-pass wavelet filter is applied to obtain approximate coefficient matrices. This stage reduces the number of dimensions of raw images and thus increases the efficiency of the deep learning methods. Dimension reduction helps computationally expensive methods in dealing with high-dimensional image data. As X-ray images encompass texture information, it is possible using a high-pass filter may remove some texture information [1]. Therefore, the coefficients matrices (horizontal, vertical, and diagonal) achieved by a high-pass wavelet filter are not considered. In the second stage, a LeNet structure [26] for CNNs is utilised for further extraction of the deep structural features of the information fed by the low-pass filter. Lastly, the wavelet-based features extracted from the previous stages serve as inputs into a SVM classifier, which is implemented based on the sequential minimal optimization (SMO) kernel.

The key challenge of the proposed approach stems from the large number of parameters that need to be learned in a CNN. We aim to prune a deep CNN using the information obtained from wavelet coefficients to reduce the load of finding all the parameters. The proposed method creates a robust deep wavelet-based learning model, which is able to extract and learn effective features of medical images. As demonstrated in Fig. 2, in the first stage, wavelet coefficient matrices are extracted from the images. This procedure is performed by a low-pass Haar wavelet filter, which is applied to the re-shaped and normalised raw data. The orthogonality property of Haar wavelet transforms (WT) enables an efficient representation of features with limited number of wavelet coefficients [31], [34], [35]. Accordingly, the proposed method extracts more informative compact features to be convolved with deep neural networks in the second stage. The CNN is used due to its ability to tolerate translations over raw images. In addition, the polynomial kernel of SMO-based SVM [38] is used to classify the deep wavelet-based features extracted from the previous stages. SVM is a classifiers aiming to split the data samples

into two or more categories. The SVM uses a kernel function to transform the input data points into a higher dimensional space and to classify them linearly. We examine two different kernels here, i.e. the popular RBF kernel in SMO-type SVM decomposition method developed by Fan et al. [11] and the polynomial Kernel in SMO-based SVM as proposed in [38]. The SMO decomposes the optimisation problem into quadratic programming sub-problems and solves the smallest possible optimisation problem, involving two Lagrange multipliers, at each step.

A. Haar Wavelet Transform

A 2-D multiresolutional decomposition of an image [30] enables us to have a scale-invariant interpretation of the image to extract its information. In particular, the details of an image could be extracted by comparing information difference between its approximations at defferent resolution levels. The difference of information between two approximations at the reolutions 2^{j+1} and 2^j can be obtained by decomposing a signal in a wavelet orthonormal basis which defines a wavelet representation. For a 1-D signal, let $\phi(x)$ be the mother Haar wavelet function, whose dilation and translation $\sqrt{s}\phi(sx-t)$ for $(s,t) \in \mathbb{R}^+ \times \mathbb{R}$ can be used to approximate functions in $\mathbf{L}^2(\mathbb{R})$, the vector space of measurable, squared-integrable functions. It is shown that there exist wavelets $\phi(x)$ such that their translation and dilation

$$\sqrt{2^j}\phi(2^j x - k), (j,k) \in \mathbb{Z}^2 \tag{1}$$

provides an orthonormal basis of $\mathbf{L}^2(\mathbb{R})$. In addition, this model can be considered for higher dimensions, in particular, two-dimensional image signals. A multiresolution approximations of an image at resolution 2^j is constructed using a two-dimensional wavelet $\phi(x,y)$ whose dilation and translation gives an orthonormal basis of $\mathbf{L}^2(\mathbb{R}^2)$. The wavelet decomposition can be considered as a signal decomposition in a set of independent and spatially oriented frequency channels and is used as a feature reduction technique in highlighting low frequency texture-level information [35].

B. The Components of a Convolutional Neural Network

A convolutional neural network has one or more blocks of convolution, which consist of convolutional and pooling layers followed by one or more fully connected multi-layer perceptrons (MLP). The CNN networks have fewer connections and parameters compared to the standard feedforward networks, resulting in an easier training process [24]. The convolution layers convert the extracted information at a higher resolution from an image to more complex features at a coarser level. The feature map, i.e. the output of the convolution layers, encompasses the convoluted information of the previous layers. As a result, each feature map can be created by applying an individual kernel into each patch of an image.

A rectified linear unit (ReLU), as an activation function, is a threshold operation, calculated by

$$\text{ReLU}(x) = \max\{0, x\} \tag{2}$$

where x is the input to a neuron. The output of this activation function is zero if the input value is less than zero, otherwise it returns the non-negative input value. Equation (3) shows the feature map of the k^{th} convolutional layer by using the non-linear ReLU activation function

$$h_{ij}^k = \text{ReLU}((W^k x)_{ij} + b_k), \tag{3}$$

where W^k and b_k are, respectively, the weights and biases of the k^{th} layer.

Max-pooling is a non-linear down-sampling technique, dividing the convolved information into $m \times n$ disjoint parts. This layer is always followed by a non-linear activation function, and is used to compute the final feature vector. Equation (4) shows the output of layer k, i.e., the feature map derived by the convolutional filter

$$h_{ij}^k = \text{ReLU}(\text{pool}(x_{ij}^k)), \tag{4}$$

where h_{ij}^k is known as the k^{th} feature map of size $(N-m+1) \times (N-m+1)$ of a given convolutional filter with a pixel at coordinates (i,j), $i,j \in 1,2,...,(N-m+1)$.

Drop-out is a regularisation technique introduced to avoid over-fitting [48]. At each block of the convolutional layer, a percentage of nodes are dropped to obtain a reduced network. This technique can be applied during either the training or testing phase, or both, in the convolutional layer.

C. Competing Methods

The performance of six learning models, including the probabilistic neural network (PNN), a logistic regression, the MLP, the deep belief network (DBN), the original CNN, and the proposed method, are compared in classifying images in ImageCLEFmed 2005 dataset. A description of the six models is as follows.

Model 1: The PNN is a feed-forward neural network using the radial basis function (RBF) as a non-linearity element in its learning procedure. The Haar wavelet-based features, extracted from the raw data, are fed to the PNN. This strategy has been conducted in a variety of studies in medicine [31], [34].

Model 2: Logistic regression is considered in this study as it is a widely used probabilistic technique for categorising medical data [8].

Model 3: The MLP is also another commonly used technique in the medical domain [32].

Model 4: The DBN is an unsupervised greedy layer-wise learning method where the output of each individual Restricted Boltzmann Machine (RBM) serves as an input to the next RBM layer [15], [16].

Model 5: The CNN is a hierarchical learning procedure sharing the feature detectors over the entire region of the image [13].

Model 6: This is our proposed method, which employs the Haar wavelet transformation on the raw medical image data. The similar architecture of the CNN described in model 5 is utilised to extract deep structural features from the wavelet filters. The CNN structure, as depicted in Fig. 2, includes four convolutional layers, consisting of different filters and pooling

TABLE I: The deep CNN details

Name	Filter size	Filter dimension	Stride	Padding
Conv1	3	10	1	2
ReLU1	1		1	0
Pooling1	2		2	0
Conv2	3	32	1	0
ReLU2	1		1	0
Pooling2	2		2	0
Conv3	3	64	1	0
ReLU3	1		1	0
Pooling3	2		2	0
Fc4	1	625	1	0
Classifier5	1	35	0	0

sizes, which is followed by a fully-connected layer to classify the X-ray images from ImageCLEFmed 2005 (see parameter details in Table I).

In addition, we used the Kruskal-Wallis H-test [9] to compare performance of six competing models in terms of the accuracy rates. The H-test is a non-parametric test, in which the null hypothesis evaluates whether the probability of a random sample from a group is equally likely to be above or below a random sample from another group. In other words, the test can be used to determine whether two groups obtained form an independent variable on a continuous or ordinal dependent variable are significantly different from a statistical perspective.

III. Experimental Results

The experiments are performed using the IRMA's ImageCLEFmed 2005 dataset. The dataset consists of 10,000 radiographs. The imagery represents different ages, genders, view positions, and pathologies. The quality of images varies significantly. The images have been categorised into very imbalanced 57 classes by reference coding using a mono-hierarchical coding scheme [27]. We choose 35 out of 57 categories for experiments in this study.

The images are split following the 10-fold cross validation method to evaluate the performance of all models. The training, testing, and validation sets consist of 80%, 10%, and 10% of the data respectively. In this experiment, all models are ran 200 times and their accuracy are recorded. As the images are in different shapes and sizes, a gray-scale model with two dimensions is considered to reduce the computational load. Therefore, in the pre-processing stage, all images are represented in two dimensions, re-sized to 56×56, and re-shaped into row images with 3136 features.

Fig. 3 compares the performance of all competing methods with the proposed method. The figure shows the box plots, representing distributions of accuracy results obtained by 200 runs of each individual classification model. As shown in the figure, the original CNN and the proposed model are more accurate and robust than the other models considering the median and the length of interquartile range. In addition, the H-test rejects the null hypothesis, since the p-values are smaller than 0.05 (95% significance level) in favour of the proposed model.

The PNN achieved about 79% accuracy after adjusting 400 latent nodes in one neural network along with 784 extracted

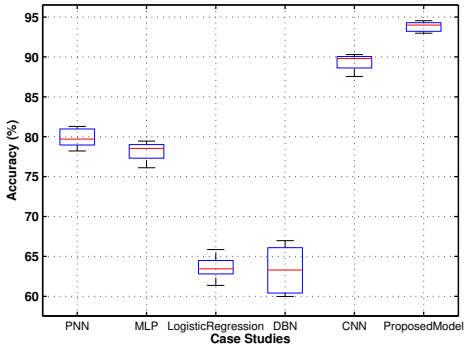

Fig. 3: Accuracy comparison among different models.

features obtained from WT. A range of hidden neurons between 50 to 450 is examined to obtain the best performance of the PNN that uses the discriminative coefficients obtained by WT. The average accuracy yielded by logistic regression is 76.9%.

Deep neural networks without and with wavelet features achieve 89% and 94.2% accuracy rates, respectively. The former feeds the raw data to the CNN, while the latter applies the low pass filter to the convolution layer of the CNN. The proposed method increases the performance by 5.2%. The experiment reveals that the approximated coefficients matrix, derived by a low pass filter, conveys more useful features than the original information. Note that the 94.2% accuracy rate is achieved by feeding the approximated coefficients matrix to the CNN. However, the accuracy rate decreases by 13.3% once the approximation matrix is merged with the detail information. The main reason for the loss of accuracy is that sharp features and edges are key information recognised by deep CNNs.

To achieve a robust model, three different parameters, i.e. (1) the depth in terms of the number of layers and filters, (2) the number of convoluted features and drop-out probabilities, and (3) different classifiers are analysed statistically.

A. The Depth Analysis

The 56×56 image patches are selected to be the input to the low pass filter, which is followed by a series of 4 convolutional layers, one fully connected layer, and a classifier at the top of the proposed architecture. This structure is conducted based on the layer patterns of AlexNet [24] and LeNet [26]. It is found that LeNet performs better than AlexNet on the experimental dataset. Note that the architecture of AlexNet is similar to the architecture of LeNet, but it is deeper and larger in size. In LeNet, a pooling layer is constructed immediately after each convolutional layer. Different kernel sizes of 2×2, 3×3, and 4×4 are considered to choose the best kernel size. Table II shows some of the experiments with different architectures. The accuracy results reported are based on the median of the prediction rates from ten-fold cross validation. As can be seen in Table II, by increasing the kernel size to 4×4, a reduction

TABLE II: The prediction rates of different models on test set.

Architecture	Input Size	Description	Kernel size	FC	Acc.%
AlexNet(5layers)	56×56	Conv-Conv-Pool	3×3	2 FC layers	91.75
AlexNet(6layers)	112×112	Conv-Conv-Pool	3×3	2 FC layers	92.5
AlexNet(6layers)	56×56	Conv-Conv-Pool	2×2	2 FC layers	93
LeNet(6layers)	112×112	Conv-Pool	3×3	1 FC layer	92
LeNet(4layers)	56×56	Conv-Pool	4×4	1 FC layer	93.3
LeNet(5layers)	56×56	Conv-Pool	3×3	1 FC layer	**94.2**

TABLE III: CCD for 3D-variables

Trail no.	Coded values of the variables		
	Feature no.	Drop-Out Conv.	Drop-Out Pool.
1	-1	-1	-1
2	1	-1	-1
3	-1	1	-1
4	1	1	-1
5	-1	-1	1
6	1	-1	1
7	-1	1	1
8	1	1	1
9	-1.682	0	0
10	1.682	0	0
11	0	-1.682	0
12	0	1.682	0
13	0	0	-1.682
14	0	0	1.682
15	0	0	0

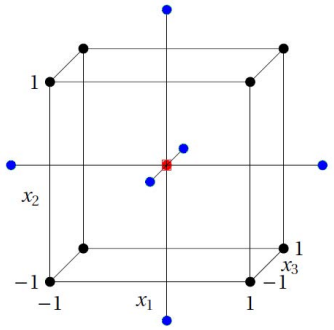

Fig. 4: Central composite design for the optimization of three variables where points of factorial design are in black, star points are in blue and central point is in red.

of about 1% accuracy is observed. Decreasing the kernel size to less than 3×3 results in a deeper structure and a reduction of 1.2% accuracy is observed.

B. Drop-Out and Wavelet-Based Feature Selection Analysis

The model has three main hyper parameters, including the number of features used for classification, the drop-out convolutional parameter and the drop-out pooling parameter. As one does not know which combination of these parameter values would result in the best result, the central composite design (CCD) is used for this problem. A CCD for examining f factors (variables) comprises a full factorial design (2^f experiments), a star design ($2f$ experiments), and a centre point [22]. Fig. 4 shows the rotatable circumscribed CCD used in this study where the distance α from centre point to star points is determined by $|\alpha| = (2^f)^{1/4} = 1.682$ with $f = 3$ variables. The total of $N = 15$ experiments are therefore needed to examine 3 variables. Table III shows the coordinates of the points in Fig. 4 where the experiments are run. To implement CCD in our study, we need to assign actual values of the variables to the coded values shown in Table III. We allow the number of features used for classification to vary in the range $[525, 725]$ whilst the drop-out probabilities for the convolutional and pooling layers both to change in the range $[0.0, 0.6]$. The experiment achieves the best result when the number of features is at 625 and 30% and 40% drop-out rates are applied for the convolutional and pooling layers, respectively.

TABLE IV: Parameter settings of the classifiers

Classifiers	**Classifier Settings**
RBF kernel SVM [11]	C-SVC type SVM with a RBF Kernel, trained by SMO, Gamma=0.28, cost of C=1, and data is subject to Normalisation
Polynomial Kernel SVM [38]	Trained by SMO, polynomial kernel with exponent exp=2.5, the complexity of parameter of C=1, and data is subject to Normalisation
Naive Bayes	Using kernel estimator function
Random Forest	The number of trees to be generated = 10
Logistic Regression	Softmax function with cross entropy loss function

TABLE V: A comparison among different classifiers.

	Precision	Recall	F-measure
Naive Bayes	0.914	0.9100	0.91
RBF kernel in SMO-based SVM	0.933	0.924	0.924
RBF Network	0.907	0.901	0.902
Random Forest	0.914	0.91	0.908
Logistic Regression	0.925	0.927	0.921
Polynomial kernel in SMO-based SVM	**0.935**	**0.932**	**0.931**

C. Classification Performance

Several classification models are considered as the last stage of our system's architecture. The classification performances of naive Bayes, RBF, random forest, logistic regression, and the two SVM kernels are analysed and compared. Table IV shows the parameter setting of the classifiers. These results are obtained by passing the raw data through the low pass filter of a single-level two-dimensional transformation. The Haar wavelets is utilized to select the best information of the pre-processed raw data. Then, LeNet is utilised by using those classifiers with 35 nodes for classification. The experimental results show that it is useful to utilise the polynomial kernel of the SMO-based SVM. This kernel outperforms the RBF kernel-based SVM in terms of both accuracy and computational cost.

All the experiments are implemented within the Theano library [5], [7]. It is a framework based on Python developed by the LISA group at University of Montreal. To conduct the experiments, all datasets are employed to obtain the performance of the classification model implemented using Theano built on Linux OS on an NVIDIA Titan Black GPU.

D. Performance Comparisons of Competing Methods

Table V presents the results of the classification models on the ImageCLEFmed 2005 dataset. The features extracted using the CNN model with Haar wavelet are fed into all the models. Polynomial kernel in SMO-based SVM achieves the best performance among all classifiers on the same input features. In addition, Table VI shows the performance of each category in terms of prediction rate, recall, and precision for

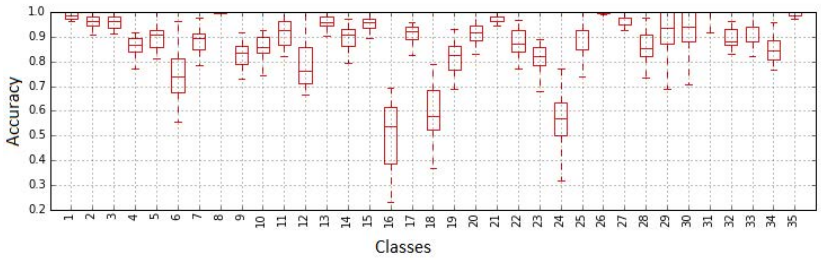

Fig. 5: Boxplot analysis per class of the proposed method (ten-fold cross-validation).

the best classifier. More specifically, the box plot per class, in terms of variation of prediction rates on a ten-fold basis is also depicted in Fig. 5. As can be seen in the figure, 30 out of 35 classes have achieved an accuracy score of more than 80%.

From the computational time viewpoint, applying a low-pass filter to the deep CNN significantly reduces the computational cost, as compared with the original CNN. The computational cost for extracting the test features takes 0.29 seconds, which is less than 0.63 seconds of the original CNN. The shorter execution time originates from feeding the most relevant features and discriminative coefficients, captured by the Haar wavelets. This contribution also reduces the data dimension significantly, from $56 \times 56 = 3,136$ to $28 \times 28 = 784$.

IV. Comparisons with Literature Results and Discussions

A major challenge faced in this experimental study is to handle the imbalanced data. As detailed in Table VI, the number of samples for each class varies, for instance, there are only 61 samples for class 16, while there are 3587 samples for class 8, which affects the overall accuracy. In addition, different versions of the dataset with different image numbers have been used in the literature, which makes a direct comparison difficult. However, the proposed model outperforms some reported results on the ImageCLEFmed 2005 dataset. Table VII shows examples of some studies using almost the same dataset, as presented in [39] and [21]. A merged-based scheme used in [39] to cope with the low accuracy rate, i.e., 64% for classifying 40 classes. In that experiment, a total of 27 out of 40 classes exhibited an accuracy lower than 60%. As such, some classes were merged based on some similarities, which produced 90.83% accuracy for 25 merged classes, but with the cost of losing critically useful and important information. In their work, the low prediction accuracy for the three categories of "Lower leg", "Forearm", and "Upper Leg" are owing to insufficient number of images for a deep structural network to learn a model. Moreover, the other reason is that, as shown in Fig. 6, these categories semantically and visually have more similarity with some other categories. Therefore, the authors in [39] merged all images of "Lower leg" category with images in the "Knee" category, in order to achieve a better performance. However, it might not be useful in the real medical domain.

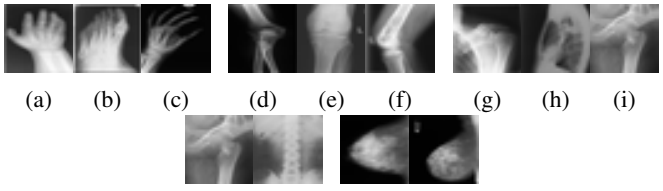

Fig. 6: Classes merged by [39]

Fig. 6 shows a sample of the classes merged in [39]. Fig. 6a, 6b, and 6c are from three different classes 4, 14, and 34 respectively are considered in the same group in [39]. Similarly, Fig. 6d, 6e, and 6f are from three different classes 6, 17, and 28, and are collected in the same class. In addition, same story happened to Figs. (6g, 6h, 6i), (6j, 6k), and (6l, 6m). However, our proposed model achieve an overall median accuracy rates above 82% for those categories without merging. Our results confirms that there is no need to merge several categories to improve the accuracy of classification models.

V. Conclusions and Future Work

This study addresses the issue of the curse of dimensionality by introducing a hybrid method combining Haar wavelet features along with the deep CNN structure. While the CNN model helps handle complex and noisy images, it imposes high computational burdens for training and testing. Therefore, the Haar wavelet features provide a set of discriminative features for accelerating the learning process of the CNN model. Two main motivations behind this study are (1) applying the low pass filter of a single-level two-dimensional transformation utilizing the Haar wavelet to extract the discriminative and compact features leading to a significant reduction in the computational burden of the CNN model; (2) utilising deep neural network model to capture the highly convoluted features from 35 categories of the benchmark, which is challenging with respect to the nature of the X-ray images.

By considering the existence of a semantic gap and strongly imbalanced categories, the investigation on a large number of categories is difficult. However, the techniques of augmentation could be a solution, and a proper investigation is required. The difficulty stems from a high level of similarity among many classes. Hence, as a future work, with respect to the

TABLE VI: Performance of the proposed model on the categories.

Class Name	Anatomic	Direction	Train #	Test #	Median	Precision	Recall
1	Cranium	Coronal	328	83	98.80%	98.80%	98.80%
2	Cervical	Coronal	224	56	96.43%	97.28%	96.43%
3	Lumbar spine	Coronal	230	58	96.55%	94.96%	96.55%
4	Hand	Coronal	450	113	86.73%	86.63%	86.73%
5	Radio carpal joint	Coronal	108	27	90.74%	91.83%	92.59%
6	Elbow	Coronal	106	27	74.07%	80.71%	74.07%
7	Shoulder	Coronal	184	47	89.36%	82.65%	89.36%
8	Chest	Coronal	2869	718	99.86%	99.51%	99.86%
9	Chest bones	Coronal	145	37	83.78%	84.62%	83.78%
10	Abdomen	Gastrointestinal system	170	43	86.05%	95.06%	87.21%
11	Thoracic spine	Coronal	110	28	92.86%	90.00%	92.86%
12	Shoulder	Sagittal	82	21	76.19%	85.36%	78.57%
13	Pelvis	Coronal	212	53	96.23%	97.16%	96.23%
14	Foot	Coronal	289	73	91.10%	90.86%	91.10%
15	Ankle joint	Coronal	193	49	95.92%	90.18%	95.92%
16	Lower leg	Coronal	48	13	53.85%	66.67%	50.00%
17	Knee	Coronal	205	52	92.31%	88.57%	92.31%
18	Upper leg	Coronal	72	19	57.89%	72.38%	57.89%
19	Hip	Coronal	112	29	82.76%	80.95%	81.03%
20	Left breast	Other orientation	68	18	91.67%	91.61%	88.89%
21	Cervical spine	Sagittal	215	54	98.15%	93.04%	98.15%
22	Thoracic spine	Sagittal	123	31	87.10%	90.16%	87.10%
23	Radio carpal joint	Sagittal	110	28	82.14%	77.03%	82.14%
24	Forearm	Sagittal	85	22	56.82%	63.30%	56.82%
25	Elbow	Sagittal	106	27	92.59%	92.86%	90.74%
26	Chest	Sagittal	833	209	**100.00%**	**99.05%**	**100.00%**
27	Ankle joint	Sagittal	166	42	97.62%	95.24%	97.62%
28	Knee	Sagittal	180	45	85.56%	87.93%	85.56%
29	Right breast	Axial	64	16	93.75%	88.89%	93.75%
30	Left breast	Axial	68	17	94.12%	89.47%	94.12%
31	Knee	Axial	98	25	100.00%	92.59%	100.00%
32	Facial cranium	Other orientation	240	60	88.33%	93.15%	88.33%
33	Right breast	Other orientation	67	17	88.24%	88.56%	88.24%
34	Hand	Other orientation	103	26	84.62%	91.67%	84.62%
35	Neuro cranium	Sagittal	292	73	100.00%	97.98%	100.00%

TABLE VII: Examples of studies using the ImageCLEFmed 2005 database

Approach	#Images	#Classes	ImageCLEFmed	Acc.%
[39]	9100	40	2005	64.00%
[39]	9100	25	2005	90.83%
[21]	2400	30	2007	93.10% precision, 89.43% recall
[33]	4402	21	2005	**94.2%**
[41]	5000	20	2004	81.96%

nature of the data, proper augmentation techniques would be utilized to allow incorporate all categories separately.

REFERENCES

[1] Tinku Acharya and Ajoy K Ray. *Image Processing: Principles and Applications*. John Wiley & Sons, 2005.

[2] Munif Alotaibi and Ausif Mahmood. Improved gait recognition based on specialized deep convolutional neural network. *Computer Vision and Image Understanding*, 164:103–110, 2017.

[3] Mutasem K Alsmadi. Content-based image retrieval using color, shape and texture descriptors and features. *Arabian Journal for Science and Engineering*, pages 1–14, 2020.

[4] Syed Muhammad Anwar, Muhammad Majid, Adnan Qayyum, Muhammad Awais, Majdi Alnowami, and Muhammad Khurram Khan. Medical image analysis using convolutional neural networks: a review. *Journal of medical systems*, 42(11):226, 2018.

[5] Frédéric Bastien, Pascal Lamblin, Razvan Pascanu, James Bergstra, Ian Goodfellow, Arnaud Bergeron, Nicolas Bouchard, David Warde-Farley, and Yoshua Bengio. Theano: new features and speed improvements. *arXiv preprint arXiv:1211.5590*, 2012.

[6] Amin Beheshti, Vahid Moraveji-Hashemi, Shahpar Yakhchi, Hamid Reza Motahari-Nezhad, Seyed Mohssen Ghafari, and Jian Yang. personality2vec: enabling the analysis of behavioral disorders in social networks. In *Proceedings of the 13th International Conference on Web Search and Data Mining*, pages 825–828, 2020.

[7] James Bergstra, Olivier Breuleux, Frédéric Bastien, Pascal Lamblin, Razvan Pascanu, Guillaume Desjardins, Joseph Turian, David Warde-Farley, and Yoshua Bengio. Theano: a CPU and GPU math expression compiler. In *Proceedings of the Python for Scientific Computing Conference (SciPy)*, June 2010.

[8] Corina Botoca, Razvan Bardan, Mircea Botoca, and Florin Alexa. Prostate cancer prognosis evaluation assisted by neural networks. *WSEAS Transactions on Computers*, 9(2):164–73, 2010.

[9] Norman Breslow. A generalized kruskal-wallis test for comparing k samples subject to unequal patterns of censorship. *Biometrika*, pages 579–594, 1970.

[10] Issam El-Naqa, Yongyi Yang, Nikolas P Galatsanos, Robert M Nishikawa, and Miles N Wernick. A similarity learning approach to content-based image retrieval: application to digital mammography. *IEEE Transactions on Medical Imaging*, 23(10):1233–1244, 2004.

[11] Rong-En Fan, Pai-Hsuen Chen, and Chih-Jen Lin. Working set selection using second order information for training support vector machines. *The Journal of Machine Learning Research*, 6:1889–1918, 2005.

[12] Mohammed Suliman Haji, Mohammed Hazim Alkawaz, Amjad Rehman, and Tanzila Saba. Content-based image retrieval: a deep look at features prospectus. *International Journal of Computational Vision and Robotics*, 9(1):14–38, 2019.

[13] Amin Khatami, Morteza Babaie, Abbas Khosravi, HR Tizhoosh, Syed Moshfeq Salaken, and Saeid Nahavandi. A deep-structural medical image classification for a radon-based image retrieval. In *IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)*, pages 1–4. IEEE, 2017.

[14] Amin Khatami, Morteza Babaie, HR Tizhoosh, Abbas Khosravi, Thanh Nguyen, and Saeid Nahavandi. A sequential search-space shrinking using cnn transfer learning and a radon projection pool for medical image retrieval. *Expert Systems with Applications*, 100:224–233, 2018.

[15] Amin Khatami, Abbas Khosravi, Chee Peng Lim, and Saeid Nahavandi. A wavelet deep belief network-based classifier for medical images. In *International Conference on Neural Information Processing*, pages 467–474. Springer, 2016.

[16] Amin Khatami, Abbas Khosravi, Thanh Nguyen, Chee Peng Lim, and Saeid Nahavandi. Medical image analysis using wavelet transform and deep belief networks. *Expert Systems with Applications*, 86:190–198,

2017.

[17] Amin Khatami, Saeed Mirghasemi, Abbas Khosravi, Chee Peng Lim, Houshyar Asadi, and Saeid Nahavandi. A swarm optimization-based kmedoids clustering technique for extracting melanoma cancer features. In *International Conference on Neural Information Processing*, pages 307–316. Springer, 2017.

[18] Amin Khatami, Asef Nazari, Abbas Khosravi, Chee Peng Lim, and Saeid Nahavandi. A weight perturbation-based regularisation technique for convolutional neural networks and the application in medical imaging. *Expert Systems with Applications*, page 113196, 2020.

[19] Amin Khatami, Yonghang Tai, Abbas Khosravi, Lei Wei, Mohsen Moradi Dalvand, Jun Peng, and Saeid Nahavandi. A haptics feedback based-lstm predictive model for pericardiocentesis therapy using public introperative data. In *International Conference on Neural Information Processing*, pages 810–818. Springer, 2017.

[20] Ivan Kiskin, Davide Zilli, Yunpeng Li, Marianne Sinka, Kathy Willis, and Stephen Roberts. Bioacoustic detection with wavelet-conditioned convolutional neural networks. *Neural Computing and Applications*, 32(4):915–927, 2020.

[21] Byoung Chul Ko, Seong Hoon Kim, and Jae-Yeal Nam. X-ray image classification using random forests with local wavelet-based cs-local binary patterns. *Journal of Digital Imaging*, 24(6):1141–1151, 2011.

[22] Mohamed A Korany, Marwa AA Ragab, Rasha M Youssef, and Mostafa A Afify. Experimental design and machine learning strategies for parameters screening and optimization of hantzsch condensation reaction for the assay of sodium alendronate in oral solution. *RSC Advances*, 5(9):6385–6394, 2015.

[23] Bartosz Krawczyk and Gerald Schaefer. A hybrid classifier committee for analysing asymmetry features in breast thermograms. *Applied Soft Computing*, 20:112–118, 2014.

[24] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.

[25] Hüseyin Kutlu and Engin Avcı. A novel method for classifying liver and brain tumors using convolutional neural networks, discrete wavelet transform and long short-term memory networks. *Sensors*, 19(9):1992, 2019.

[26] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[27] Thomas M Lehmann, MO Gold, Christian Thies, Benedikt Fischer, Klaus Spitzer, Daniel Keysers, Hermann Ney, Michael Kohnen, Henning Schubert, and Berthold B Wein. Content-based image retrieval in medical applications. *Methods of Information in Medicine*, 43(4):354–361, 2004.

[28] Bai-Jiong Lin, Xiang-Ru Li, and Wo-Liang Yu. Binary neutron stars gravitational wave detection based on wavelet packet analysis and convolutional neural networks. *Frontiers of Physics*, 15(2):24602, 2020.

[29] Hongya Lu, Haifeng Wang, Qianqian Zhang, Daehan Won, and Sang Won Yoon. A dual-tree complex wavelet transform based convolutional neural network for human thyroid medical image segmentation. In *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, pages 191–198. IEEE, 2018.

[30] Stephane G Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, 1989.

[31] Roshan Joy Martis, U Rajendra Acharya, and Lim Choo Min. ECG beat classification using pca, lda, ica and discrete wavelet transform. *Biomedical Signal Processing and Control*, 8(5):437–448, 2013.

[32] Michael E Mavroforakis, Harris V Georgiou, Nikos Dimitropoulos, Dionisis Cavouras, and Sergios Theodoridis. Mammographic masses characterization based on localized texture and dataset fractal analysis using linear, neural and support vector machine classifiers. *Artificial Intelligence in Medicine*, 37(2):145–162, 2006.

[33] Seyyed Mohammad Mohammadi, Mohammad Sadegh Helfroush, and Kamran Kazemi. Novel shape-texture feature extraction for medical x-ray image classification. *International Journal of Innovative Computing, Information and Control*, 8:659–76, 2012.

[34] A Padma Nanthagopal and R Sukanesh Rajamony. Classification of benign and malignant brain tumor ct images using wavelet texture parameters and neural network classifier. *Journal of Visualization*, 16(1):19–28, 2013.

[35] Thanh Nguyen, Saeid Nahavandi, Douglas Creighton, and Abbas Khosravi. Mass spectrometry cancer data classification using wavelets and genetic algorithm. *FEBS Letters*, 589(24):3879–3886, 2015.

[36] Edouard Oyallon, Eugene Belilovsky, and Sergey Zagoruyko. Scaling the scattering transform: Deep hybrid networks. In *Proceedings of the IEEE international conference on computer vision*, pages 5618–5627, 2017.

[37] Luca Piras and Giorgio Giacinto. Information fusion in content based image retrieval: A comprehensive overview. *Information Fusion*, 37:50–60, 2017.

[38] J C Platt. Fast training of support vector machines using sequential minimal optimization. In *Advances in Kernel Methods*, chapter 12, page 185–208. MIT Press, Cambridge, MA, 1999.

[39] Hossein Pourghassem and Hassan Ghassemian. Content-based medical image classification using a new hierarchical merging scheme. *Computerized Medical Imaging and Graphics*, 32(8):651–661, 2008.

[40] Sultan Noman Qasem and Siti Mariyam Shamsuddin. Radial basis function network based on time variant multi-objective particle swarm optimization for medical diseases diagnosis. *Applied Soft Computing*, 11(1):1427–1438, 2011.

[41] Md Mahmudur Rahman, Prabir Bhattacharya, and Bipin C Desai. A framework for medical image retrieval using machine learning and statistical similarity matching techniques with relevance feedback. *IEEE Transactions on Information Technology in Biomedicine*, 11(1):58–69, 2007.

[42] Imran Razzak, Michael Blumenstein, and Guandong Xu. Multiclass support matrix machines by maximizing the inter-class margin for single trial eeg classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(6):1117–1127, 2019.

[43] Muhammad Imran Razzak, Saeeda Naz, and Ahmad Zaib. Deep learning for medical image processing: Overview, challenges and the future. In *Classification in BioApps*, pages 323–350. Springer, 2018.

[44] Antonio Robles-Kelly and Asef Nazari. Incorporating the barzilai-borwein adaptive step size into sugradient methods for deep network training. In *2019 Digital Image Computing: Techniques and Applications (DICTA)*, pages 1–6. IEEE, 2019.

[45] Josiane Rodrigues, Marco Cristo, and Juan G Colonna. Deep hashing for multi-label image retrieval: a survey. *Artificial Intelligence Review*, pages 1–47, 2020.

[46] Maria Ximena Bastidas Rodriguez, Adrien Gruson, Luisa Polania, Shin Fujieda, Flavio Prieto, Kohei Takayama, and Toshiya Hachisuka. Deep adaptive wavelet network. In *The IEEE Winter Conference on Applications of Computer Vision*, pages 3111–3119, 2020.

[47] Narjis Mezaal Shati, Noor khalid Ibrahim, and Taha Mohammed Hasan. A review of image retrieval based on ontology model. *Journal of Al-Qadisiyah for computer science and mathematics*, 12(1):Page–10, 2020.

[48] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958, 2014.

[49] Lazaros Tsochatzidis, Konstantinos Zagoris, Nikolaos Arikidis, Anna Karahaliou, Lena Costaridou, and Ioannis Pratikakis. Computer-aided diagnosis of mammographic masses based on a supervised content-based image retrieval approach. *Pattern Recognition*, 71:106–117, 2017.

[50] Travis Williams and Robert Li. Advanced image classification using wavelets and convolutional neural networks. In *2016 15th IEEE international conference on machine learning and applications (ICMLA)*, pages 233–239. IEEE, 2016.

[51] Travis Williams and Robert Li. Wavelet pooling for convolutional neural networks. 2018.

[52] Zhihua Xia, Neal N Xiong, Athanasios V Vasilakos, and Xingming Sun. Epcbir: An efficient and privacy-preserving content-based image retrieval scheme in cloud computing. *Information Sciences*, 387:195–204, 2017.

[53] Jianpeng Zhang, Yutong Xie, Qi Wu, and Yong Xia. Medical image classification using synergic deep learning. *Medical Image Analysis*, 54:10–19, 2019.

[54] Qiang Zheng, Honglun Li, Baode Fan, Shuanhu Wu, and Jindong Xu. Integrating support vector machine and graph cuts for medical image segmentation. *Journal of Visual Communication and Image Representation*, 55:157–165, 2018.