

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

DeTraC: Transfer Learning of Class Decomposed Medical Images in Convolutional Neural Networks

ASMAA ABBAS¹, MOHAMMED M. ABDELSAMEA^{1,2,*}, AND MOHAMED MEDHAT GABER²

¹Mathematics department, faculty of Science, Assiut University, Assiut, Egypt

²School of Computing and Digital Technology, Birmingham City university, Birmingham, UK

*Corresponding author: Mohammed M. Abdelsamea (e-mail: mohammed.abdelsamea@bcu.ac.uk).

ABSTRACT

Due to the high availability of large-scale annotated image datasets, paramount progress has been made in deep convolutional neural networks (CNNs) for image classification tasks. CNNs enable learning highly representative and hierarchical local image features directly from data. However, the availability of annotated data, especially in the medical imaging domain, remains the biggest challenge in the field. Transfer learning can provide a promising and effective solution by transferring knowledge from generic image recognition tasks to the medical image classification. However, due to irregularities in the dataset distribution, transfer learning usually fails to provide a robust solution. Class decomposition facilitates easier to learn class boundaries of a dataset, and consequently can deal with any irregularities in the data distribution. Motivated by this challenging problem, the paper presents Decompose, Transfer, and Compose (DeTraC) approach, a novel CNN architecture based on class decomposition to improve the performance of medical image classification using transfer learning and class decomposition approach. DeTraC enables learning at the subclass level that can be more separable with a prospect to faster convergence. We validated our proposed approach with three different cohorts of chest X-ray images, histological images of human colorectal cancer, and digital mammograms. We compared DeTraC with the state-of-the-art CNN models to demonstrate its high performance in terms of accuracy, sensitivity, and specificity.

INDEX TERMS Convolution neural networks; class decomposition; data irregularity; medical image classification; transfer learning.

I. INTRODUCTION

CLASSIFICATION of chest X-ray (CXR) images into normal or having Tuberculosis (TB) is an essential component in computer-aided diagnosis (CAD) of lung health-care [1]–[4]. Fig. 1 shows a negative example of normal CXR without any signs of TB and a positive one with manifestations.

Statistical/classical machine learning algorithms have been extensively used for lung classification [5]–[8] and nodule diagnosis from computed tomography (CT) images [9]. For instance, in [10], three statistical features were calculated from lung texture to discriminate between malignant and benign lung nodules using support vector machines (SVM) classifier. A grey-level co-occurrence matrix (GLCM) method was used along with back-propagation network (BPN) [11] to classify computed tomography (CT) images from being normal or

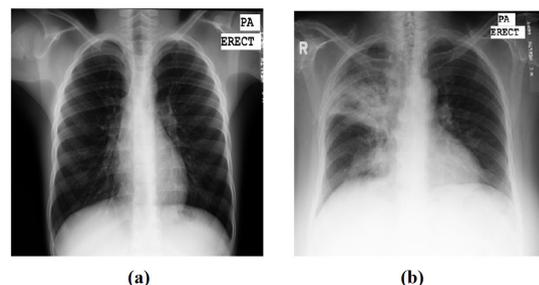


FIGURE 1. Example of a) normal and b) abnormal CXR.

cancerous. For emphysema diseases of lung images, different approaches were used to describe the texture features from lung region, such as in [12] Local Binary Pattern (LBP)

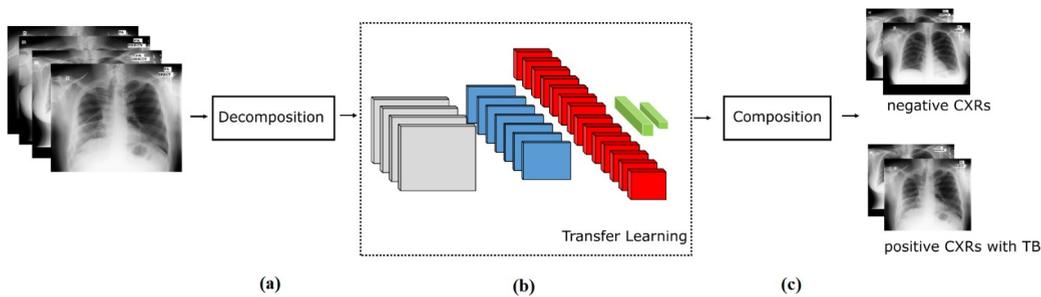


FIGURE 2. Decompose, Transfer, and Compose (DeTraC) model: (a) decomposition of the original classes into sub-classes, (b) transfer learning using ImageNet pre-trained CNN-based model, and (c) recombination back to the final problem.

was used with k nearest neighbour (kNN) as a classifier to classify the lung region into three classes, and in [13] SVM also used with radial basis function (RBF) kernel. Other statistical machine learning methods have also demonstrated good classification performance in other domains, with different data types [14]–[16]. Deep learning algorithms have also demonstrated their great success in different domains including bioinformatics [17]–[19], plant identification [20], medical image analysis [3], [21], [22] and wind power prediction [23]. With the availability of enough annotated images, deep learning approaches [2], [24]–[30] have demonstrated their superiority over the statistical machine learning approaches [9]. In [31], two deep learning approaches, deep belief network (DBN) and restricted Boltzmann machine (RBM) were used to classify the lung nodules from CT into malignant or benign based on three feature extraction methods (GLCM, histogram features, and wavelet transformation). A deep residual network (ResNet) with curriculum learning and transfer learning was applied in [32] to classify lung nodules, and in [33] the stacked denoising auto-encoder (SDAE) was used for classification of the nodules in CT images into two classes: benign and pulmonary and from CXR in [34]. Convolutional networks (CNN) architecture is one of the most popular deep learning approaches with superior achievements in the medical imaging domain [35], [36]. The primary success of the ConvNet is due to its ability to learn features automatically from domain-specific images [37], unlike the statistical machine learning methods. The popular strategy for training a CNN architecture is to transfer learned knowledge from a pre-trained network that fulfilled one task into a new task [20], [38], [39]. This solution is efficient and easy to apply without the need for a huge annotated dataset for training; therefore many researchers tend to apply this strategy especially with medical imaging [19], [40]. Transfer knowledge from pre-trained models via fine-tuning showed outstanding performance in X-ray image classification [25], [36], [41] and recently for the classification of 14 diseases as a multi-task learning [42]. Transfer learning can be accomplished with three major scenarios [43]: a) “shallow tuning”, which adopts only the last classification layer to cope with the new task, and freezes the parameters of the remaining layers without training; b) “deep tuning” which

aims to retrain all the parameters of the pre-trained network from end-to-end manner; and (c) “fine-tuning” that aims to gradually train more layers by tuning the learning parameters until a significant performance boost is achieved.

However, building a robust image classification model for datasets with imbalanced classes can be a very challenging task, especially in the medical imaging domain [44]–[47]. Class decomposition aims to the simplification of the local structure of a dataset, by learning the boundary between certain characters within each class, in a way to cope with any irregularities or imbalances in the data distribution. Class decomposition has been previously used in various automatic learning workbooks as a preprocessing step to improve the performance of different classification models and in the medical diagnostic domain, it has been applied to significantly enhance the classification performance of models such as Random Forests, Naive Bayes, C4.5, and SVM [48]–[52].

In this paper, we propose a novel convolutional neural network architecture based on class decomposition, which we term Decompose, Transfer, and Compose (DeTraC) model, to improve the performance of pre-trained models on the classification of X-ray images. This is done by adding class decomposition and composition components, respectively, before and after transferring knowledge using an ImageNet pre-trained CNN model. The proposed workflow aims to partition each class within the image dataset into k subsets and then assign new labels to the new set, where each subset is treated as an independent class, then those subsets are assembled back to produce the final predictions (a classification of a subclass is mapped to its parent class), as illustrated by Fig. 2. For the classification performance evaluation, we used images of CXR as an exemplar to distinguish between normal and abnormal cases. The dataset possesses complicated computer vision challenging problems due to the intensity inhomogeneity and the huge intensity-overlap between the classes. Moreover, we evaluated our framework on two other different cohorts of digital mammograms and images of human colorectal cancer to demonstrate the robustness of our solution to data irregularities.

The paper is organised as follows. Section II provides an overview of related work and motivates our work. In Section III, we discuss, in details, the main components

of our proposed framework. Section IV illustrates detailed experiments on three different image data sets. Section V highlights the key findings of our work. Finally, Section VI concludes the work with a summary and pointers to possible future work.

II. CONTRIBUTION

With the availability of large annotated datasets, the chance for the different classes to be well-represented is high. As a consequence, the learned in-between class-boundaries are most likely to be generic enough to new samples. On the other hand, with the limited availability of annotated data, especially when some classes are suffering more compared to others in terms of the size and representation, the generalisation error might increase. This is due to the fact that there might be a miscalibration between the minority and majority classes.

Class decomposition has its roots in [51], when it was first proposed to enhance low variance classifiers by increasing the flexibility of the decision boundary. A similar method has been applied in enhancing k -Nearest Neighbours classifiers using clusters (resulted from class decomposition) instead of individual instances as the nearest neighbours, applying gravity, density, and distance, as proximity measures [53]. Class decomposition found its way to enhance more classification methods like decision tree classifier [54], SVM [55] and Random Forests [56], [57] which providing even more flexibility to their respective decision boundaries. A special focus on applying the method in the medical domain can be traced in these papers. The reason behind this special attention to medical data is the inherent complexity of the decision boundaries in this domain. Classes can greatly overlap, resulting in harder to detect decision boundaries.

Although class decomposition has been successfully applied to medical datasets, it is yet to be tested in the area of medical imaging. Thus, the work reported in this paper investigates the adoption of the method to classification in medical imaging. In this paper, we propose a novel image classification framework inspired by the capability of class-decomposition in dealing with data irregularities to improve the performance of ImageNet pre-trained CNN models. To the best of our knowledge, this is the first attempt to employ class decomposition within the CNN framework for image classification. In this work, to demonstrate the effectiveness and importance of class decomposition mechanism in CNN, we compared the performance of our proposed approach before and after the class decomposition process. We used the Wilcoxon signed-rank test [58] to show that applying class decomposition yields statistically significantly more accurate classification results than not applying it. Moreover, to demonstrate the robustness of our solution to data irregularities, we evaluated and compared the performance of our framework with state-of-the-art CNN models on three different datasets of real chest X-ray images (with binary class structure), histological images of human colorectal cancer (with multi-class balance structure) and digital mammograms

(with multi-class imbalance structure).

III. THE DETRAC MODEL

In this section, we describe in detail our proposed Decompose, Transfer, and Compose (DeTraC) model. The proposed approach has three phases. In the first phase, we use ImageNet pre-trained CNN as an off-the-shelf feature extractor to extract a set of local features from the individual images. Accordingly, a class-decomposition method was used to simplify the complexity of the local structure of our image dataset. In the second phase, we apply a sophisticated gradient descent optimisation method to fine-tune the same CNN model (of the first phase) pre-trained from ImageNet to our image classification task. Finally, we adapt and refine the final classification of the input images using error-correction criteria applied to a softmax layer. Fig. 3 shows an overview of the proposed network. We adopt an ImageNet pre-trained CNN model to the classification of our images into the original classes. This step aims to a) extract a set of representative features for the differentiation between the different classes, and b) fine-tune the CNN feature maps accordingly. The high-dimension feature space is reduced using principal component analysis (PCA) (black arrow) and a class decomposition method (yellow arrow) is used to simplify the complexity of the data, which results in more classes. Then the converged CNN weights can be transferred to the same network structure (blue arrow) but adopted specifically to cope with the new classes. The benefit of this transformation is to speed up the convergence and to cope with the limited availability of training samples and irregularities. Finally, a class relabelling (red arrow) is used to remap the classification back to the original problem using a simple error correction criterion.

A. CLASS DECOMPOSITION

A shallow-tuning mode was used during the adaptation, weight initialisation, and training of an ImageNet pre-trained CNN model using our image dataset. The representation was considered from the last fully-connected layer. In other words, we used the off-the-shelf CNN features of pre-trained models on ImageNet (where the training is accomplished only on the final classification layer) to construct the image feature space. However, due to the high dimensionality associated with the images, we applied PCA to project the high-dimension feature space into a lower-dimension, such that the highly correlated features were ignored. This step is important for the class decomposition to produce more homogeneous classes, reduce the memory requirements, and improve the efficiency of the framework.

Now assume that our feature space (PCA's output) is represented by a 2-D matrix (denoted as dataset α), and \mathbf{L} is a class category. α and \mathbf{L} can be written as

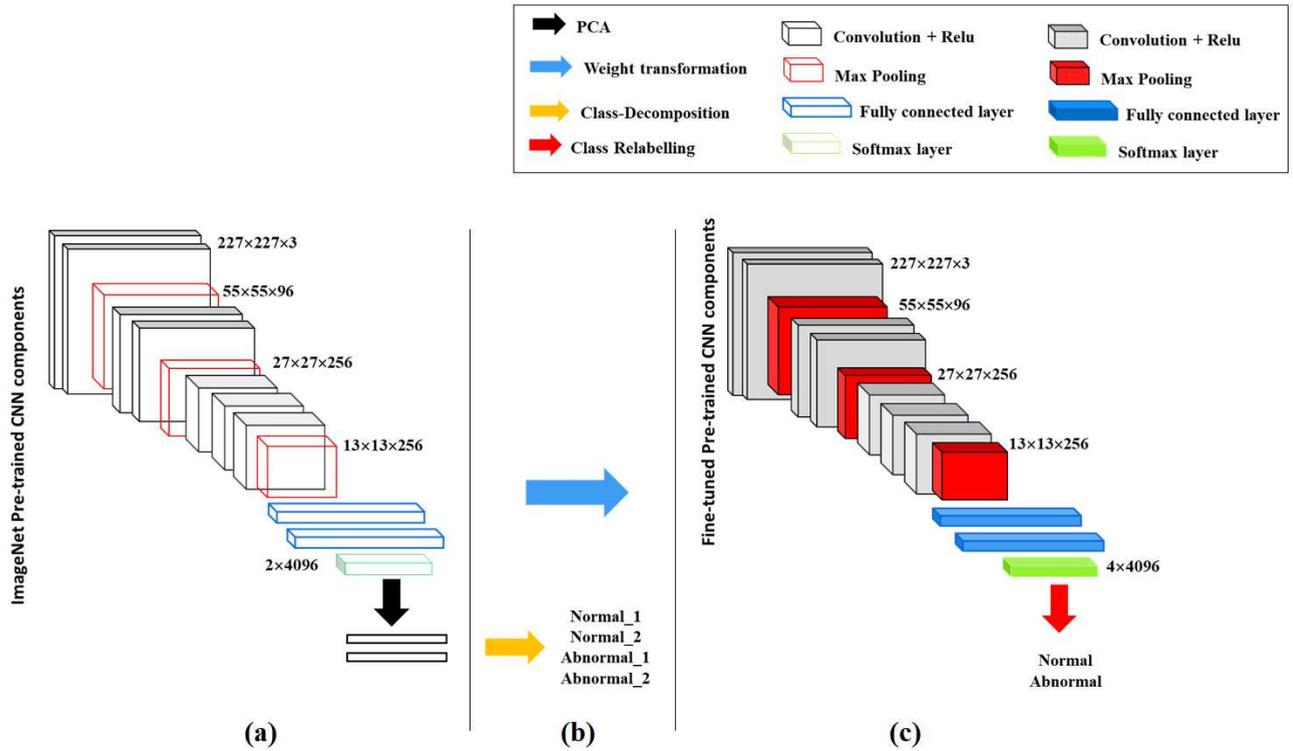


FIGURE 3. The generic architecture of *DeTraC* model: (a) feature extraction using ImageNet pre-trained *CNN* model and principle component analysis (black arrow), (b) class decomposition process with weight transformation (blue arrow), and (c) fine-tuned ImageNet pre-trained *CNN* model with error correction using class composition (red arrow) for output prediction.

$$\alpha = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix}, \mathbf{L} = \{l_1, l_2, \dots, l_c\}, \quad (1)$$

where c is the number of classes and m is the number of features. For class decomposition, we used k -means clustering [59] to further divide each class into homogeneous subclasses, where each pattern in the original class \mathbf{L} is assigned to a class label associated with the nearest centroid based on the squared euclidean distance (SED):

$$SED = \sum_{j=1}^k \sum_{i=1}^n \| a_i^{(j)} - \mu_j \|^2, \quad (2)$$

where centroids are denoted as μ_j .

Now, if $k = 2$ then each class in \mathbf{L} is divided into two clusters, resulting in a new dataset (denoted as dataset B) with 4 sub-classes denoted as ($normal_1$, $normal_2$, $abnormal_1$, $abnormal_2$), see Fig. 4.

Let $C = \{l_{11}, l_{12}, \dots, l_{1k}, l_{21}, l_{22}, \dots, l_{2k}, \dots, l_{ck}\}$, represents the set of subclasses/clusters (i.e. $|C| = c \times k$) and $|\cdot|$ denotes the number of elements in the set. Note that this definition applies for homogeneous decomposition, where each class is decomposed to the same number of clusters/subclasses (e.g. $k = 2$).

Also, the feature space of both dataset A and B can be illustrated as:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} & l_1 \\ a_{21} & a_{22} & \dots & a_{2m} & l_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & l_2 \\ a_{n1} & a_{n2} & \dots & a_{nm} & l_2 \end{bmatrix}, \quad (3)$$

$$B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1m} & l_{11} \\ b_{21} & b_{22} & \dots & b_{2m} & l_{1c} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & l_{21} \\ b_{n1} & b_{n2} & \dots & b_{nm} & l_{2c} \end{bmatrix}.$$

where $a_{ij} = b_{ij}, \forall i \in \{1, 2, \dots, n\}$ and $j \in \{1, 2, \dots, m\}$.

Table 2 shows the new distribution of the dataset when AlexNet was used as the pre-trained model with our class decomposition approach.

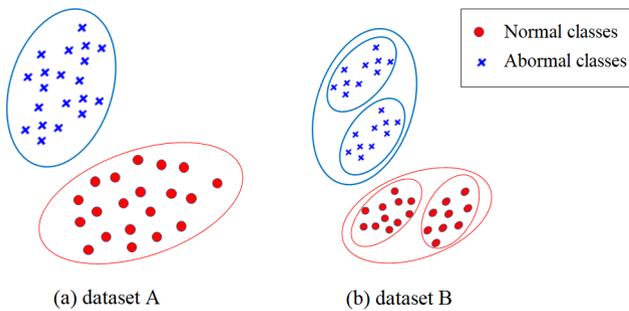


FIGURE 4. Class decomposition process: (a) original class distribution (e.g. CXR dataset A), and (b) new class distribution after class decomposition (e.g. CXR dataset B).

B. TRANSFER LEARNING

In this phase, we used several ImageNet pre-trained CNN models before and after class decomposition including: a) AlexNet [60] which is composed of 5 convolutional layers and 3 fully connected layers. AlexNet uses 3x3 max-pooling layers with ReLU activation and different kernel filters (e.g. 11 x 11, 5 x 5 and 3 x 3); b) VGG16 and VGG19 [61] with 13 and 16 convolution layers, respectively, 3x3 sized filters, and 2x2 max-pooling; c) GoogleNet [62], which has a deeper architecture with 22 layers with a smaller number of parameters compared to AlexNet and VGGNet; and d) ResNet [63], which showed excellent performance (with 3.57% of the error rate) on ImageNet with only 18 layers. Here we consider different training scenarios. We consider fine-tuning parameters of all (except the last fully-connected layer) or a part of CNN layers (when the rest are randomly initialised and the last layer is adapted according to our specific application). We also conduct an off-the-shelf CNN features of pre-trained models on ImageNet and train only the final classification layer to cope with our task. All those scenarios have been investigated before and after class decomposition to highlight the robustness and effectiveness of our solution.

With the limited availability of training data, stochastic gradient descent (SGD) can heavily be fluctuating the objective/loss function and hence overfitting can occur. To improve convergence and overcome overfitting, the mini-batch of stochastic gradient descent (mSGD) was used to minimise the objective function, $E(\cdot)$, with cross-entropy loss

$$E(y^j, z(x^j)) = -\frac{1}{n} \sum_{j=0}^n [y^j \ln z(x^j) + (1 - y^j) \ln (1 - z(x^j))], \quad (4)$$

where x^j is the set of input images in the training, y^j is the ground truth labels while $z(\cdot)$ is the predicted output from a softmax function.

Due to the onerousness in collecting and annotating a huge amount of medical images, large scale annotated image datasets (such as ImageNet) provide effective solutions to such a challenge via transfer learning where tens of mil-

lions parameters (of CNN architectures) are required to be trained. In addition to the fact that DeTraC can deal with data irregularities by class decomposition, DeTraC can also provide an efficient solution when a limited number of training images are available. This is by transferring knowledge from a generic object recognition task (i.e. large-scale image classification) to our specific-domain tasks (i.e. medical image classification) using ImageNet pre-trained models. A potential limitation of DeTraC is that since it is an image-wise CNN, it might confuse when coping with complex large images. This is due to the down-sampling process of input images that might result in losing some spatial information during the training.

C. COMPOSITION AND EVALUATION

For performance evaluation, we adopted Accuracy (ACC), Specificity (SP) and Sensitivity (SN) metrics from the confusion matrix, see Fig. 5. They are defined as:

$$Accuracy(ACC) = \frac{TP + TN}{n}, \quad (5)$$

$$Sensitivity(SN) = \frac{TP}{TP + FN}, \quad (6)$$

$$Specificity(SP) = \frac{TN}{TN + FP}, \quad (7)$$

where, TP is the true positive in case of abnormal and TN is the true negative in case of normal, while FP and FN are the incorrect model predictions for abnormal and normal cases, which are indicated with a red and a blue triangle respectively for dataset A, and the summation values of all the red and blue triangles respectively after error correction for dataset B. More precisely, with class decomposition, the TP is the summation of all the values in $abnormal_1$ and $abnormal_2$ which is represented with the red square while TN is the summation of all the values in $normal_1$ and $normal_2$, which is represented by the blue square.

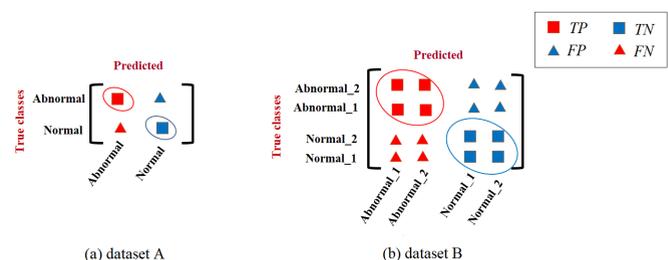


FIGURE 5. Confusion matrix tabular: (a) Binary classes within dataset A, and (b) Multi-classes after applying class decomposition process (dataset B).

In the case of a multi-classification problem, the model has been evaluated using a multi-class confusion matrix of [64]. Before error correction, the input image can be classified into one of (c) non-overlapping classes. As a consequence, the confusion matrix would be a $(N_c \times N_c)$ matrix, and TP , TN , FP and FN for a specific class i are defined as:

$$TP_i = \sum_{i=1}^n x_{ii} \quad (8)$$

$$TN_i = \sum_{j=1}^c \sum_{k=1}^c x_{jk}, j \neq i, k \neq i \quad (9)$$

$$FP_i = \sum_{j=1}^c x_{ji}, j \neq i \quad (10)$$

$$FN_i = \sum_{j=1}^c x_{ij}, j \neq i \quad (11)$$

where x_{ii} is an element in the diagonal of the matrix. We applied the same scenario after error correction to evaluate our framework for colorectal cancer and digital mammograms datasets.

D. ALGORITHMIC DESCRIPTION

Having discussed the formulations of DeTraC model, in the following, the procedural steps of DeTraC model are summarised in Algorithm 1.

Algorithm 1 DeTraC

1: procedure

- Input:
 - Input images.
 - Image-wise labels.
- Output:
 - Classification scores.

Stage I. Class Decomposition:

- 2: Load and initialise weights of an ImageNet pre-trained CNN model.
- 3: Extract off-the shelf CNN deep features using input images.
- 4: Reduce the dimensionality of the feature space using PCA.
- 5: Decompose original classes into a number of sub-classes.

Stage II. Transfer Learning:

- 6: Adapt the final classification layer of the selected pre-trained CNN model to the new classes.
- 7: Apply the mini-batch of sophisticated gradient descent optimisation method to fine-tune the adapted CNN model.

Stage III. Class Composition:

- 8: Calculate the classification scores for images associated to the decomposed classes.
- 9: Refine the final classification using error-correction criteria.

IV. EXPERIMENTAL RESULTS

In this section, we validate the performance of DeTraC with three different datasets:

- CXR image dataset from the Japanese Society of Radiological Technology (JSRT) [65], [66], which consists of 138 images (with 4020×4892 pixels) divided into 58 abnormal cases and 80 normal cases.
- Histological images of human colorectal cancer [67] of 5000 histological images (with 150×150 pixels), divided into three classes (see Table 8 for the number of samples in each class): tumour epithelium, stroma and mucosal glands.
- Digital mammograms from [68], which contains 322 mammograms images (with 1024×1024 pixels) and divided into seven classes (see Table 10 for the number

TABLE 1. The sensitivity to the parameter k in each dataset.

dataset	ACC (%)		
	$k = 1$	$k = 2$	$k = 3$
chest X-ray	69.1	71.2	65.8
digital mammograms	70	73.2	69.4
colorectal cancer	71.6	73	77.5

of samples in each class): normal, calcification (CALC), well-defined/circumscribed masses (CIRC), spiculated masses (SPIC), ill-defined masses (MISC), Architectural distortion (ARCH) and asymmetry (ASYM).

Each dataset was randomly divided into two groups; 70% as training/validation set, and 30% as testing set. We used 10-fold cross-validation technique to fine-tune the parameters of DeTraC while the performance of DeTraC has been evaluated on the testing set. For a fair comparison, all the experiments in our work have been carried out in MATLAB 2019a on a PC with the following configuration: 3.70 GHz Intel(R) Core(TM) i3-6100 Duo, NVIDIA Corporation with the donation of the Quadra P5000GPU, and 8.00 GB RAM.

A. PARAMETER SENSITIVITY

To demonstrate the sensitivity to changes in the parameter k (the number of classes in class decomposition component) with the three datasets, we evaluated the performance of our framework (based on 10-fold cross-validation) when different k values were used. As illustrated by Table 1, the highest accuracy of 71.2% and 73.2% were obtained when $k = 2$ for chest X-rays and digital mammograms datasets, respectively, while 77.5% was obtained when $k = 3$ for colorectal cancer dataset. Hence, the value of k was fixed to 2 for both chest X-ray and digital mammogram datasets and 3 for the colorectal cancer dataset. We also tested the variability between features of PCA to reduce the dimension of feature space in each dataset. We found that the first 5, 127 and 77 components for chest X-ray, digital mammograms, and colorectal cancer datasets, respectively, explain more than 95% of all variability.

B. CLASSIFICATION PERFORMANCE ON CHEST X-RAY DATASET

We first validated DeTraC on the chest X-ray dataset (Table 2) based on AlexNet with a fine-tuning strategy by training specific layers of its architecture, see Table 3. We have stopped the training of the layers at the layer named Conv3 to avoid overfitting due to the limited availability of the data and the deep architecture of the network. Note that we have followed the conventional naming of layers in the pre-trained CNN architectures [60], [61]. Likewise, we used the pre-trained network VGG16 by fine-tuning its parameters starting from the last fully connected layer (fc8) and incrementally learn layers until approached Conv4-1, see Table 4. As illustrated by Tables 3 and 4, DeTraC shows better performance in terms of accuracy, sensitivity, and specificity when compared to the classical models (AlexNet and VGG16) in both shal-

low and fine-tuned modes, in a layer-wise fashion. DeTraC achieved the highest accuracy of 74.7% with VGG16 with a sensitivity of 67.3% and a specificity of 85.0%.

TABLE 2. The number of instances in dataset *A* and *B*

Dataset <i>A</i>	# Instances	Dataset <i>B</i>	# Instances
normal	80	<i>normal</i> ₁	42
		<i>normal</i> ₂	38
abnormal	58	<i>abnormal</i> ₁	28
		<i>abnormal</i> ₂	30

Moreover, three different pre-trained CNN models (VGG19, GoogleNet, and ResNet) were also tested based on a shallow tuning mode before and after class decomposition, where the comparison is reported in Table 5. Note that the classification performance has been slightly improved in all cases with the highest accuracy of 75.4% with a sensitivity of 82% and specificity of 83% when ResNet was used. The standard deviation (*std*) was calculated and used as a robustness measure (when a 10-fold cross-validation technique was used for the accuracy, sensitivity, and specificity metrics), see Fig. 6. As illustrated by Fig. 6, applying class decomposition yields more robust results than not applying it with all the used pre-trained models.

To demonstrate the effectiveness of DeTraC classification framework, we train DeTraC on the previously used pre-trained models (i.e. AlexNet, VGG, GoogleNet, and ResNet) with a deep-tuning strategy. First, a data augmentation with multiple pre-processing techniques was used to increase the number of samples in each class. We used a variety of pre-processing methods such as flipping, transformation (translating, scaling, and rotation with various angles), colour processing and small random noise perturbation. This process resulted in a total of 40,000 lung fields. Second, we investigated the effect of various training/testing sizes on the classification performance, see Fig. 7. As illustrated in Fig. 7, the under-fitting issue is occurred more clearly by gradually decreasing the size of the training set.

The base learning rate was set to 0.0001 for all the CNN layers of all pre-trained models and 0.01 for the last fully connected layer, and was decreased by value 0.9 after every 3 epochs. L2-weight decay was applied with value 0.001 and the momentum of 0.9 was used. As illustrated by Table 7, DeTraC has achieved excellent performance in the classification of chest X-ray images, when compared to the related state-of-the-art statistical machine learning and deep learning methods. The highest accuracy of 99.8% was obtained with ResNet and outperformed 8 previously proposed approaches.

The previous experiments demonstrated the effectiveness of class decomposition. We also measured the statistical significance using the Wilcoxon signed-rank test [58] with continuity correction to establish the statistical significance of the results. At 0.05 significance level, the *p*-value = 0.0008506. This confirms the alternative hypothesis: true location shift is not equal to 0 (i.e. classification result obtained by applying class decomposition is statistically significantly more accurate than not applying it).

Fig. 8 shows the learning curves of the accuracy and loss in case of deep tuning scenario with AlexNet, without under-fitting or overfitting problems.

Fig. 9 illustrates the receiver operating characteristics (ROC) curves between the true positive rate and false-positive rate. Note that the ROC curve for the ResNet (red) has an area under the ROC curve (AUC) that is significantly greater than the other ROC curves obtained by other models.

To demonstrate the direct effect of knowledge transformation between the original model and the fine-tuned one, we compare the classification performance among the different versions of the original model (with and without augmentation) and DeTraC, see Table 6. Note how robust our model to the knowledge transferred from the different pre-trained models. As shown in Table 6, the best accuracy obtained by the original model (using ResNet with ImageNet) was 82.24% while DeTraC has achieved an accuracy of 99.8% with the same pre-trained model.

C. ROBUSTNESS TO DATA IRREGULARITY

1) Classification performance on histological images

To demonstrate the robustness of DeTraC model in coping with data irregularity, we used 5000 histological images [67] with three classes: tumour epithelium, stroma and mucosal glands (see Fig 10).

The distribution of the original data (Dataset *A*) and the generated one after class composition (Dataset *B*) is shown in Table 8. DeTraC was trained using a learning rate of 0.0001 for all the layers, the last fully connected layer was 0.01, the learning rate decreased by value 0.95 after every 4 epochs. The model was trained using mini-batch stochastic gradient descent with 128 samples per batch except VGG was 64, L2-weight decay was applied with value 0.0001 and the default momentum in the network was 0.9.

To demonstrate the superiority of our model, the model has been evaluated using a multi-classes confusion matrix of [64] before and after class decomposition, see Table 9. As shown by Table 9, DeTraC with ResNet achieved the highest accuracy of 99.1% while DeTraC with VGG16 and GoogleNet was behaving almost the same (with 99.8% and 99.7% accuracies, respectively), in case of the colorectal cancer dataset.

2) Classification performance on mammogram images

We also validated DeTraC on 322 mammogram images that were taken from [68]. The size of all the images is 1024×1024 pixels. The data set has seven types of classes as shown in Fig 11. The data distribution in the different classes before and after the composition is reported in Table 10

All the images resized into 256×256 via the Bicubic Interpolation method [70], which is very effective to produce images that are very similar to the original ones. The network was trained using a learning rate of 0.001 for all the CNN layers, 0.01 for the last fully connected layer, and was decreased by value 0.9 after every 2 epochs. The CNN model is trained using mini-batch stochastic gradient descent with 64 samples

TABLE 3. Classification performance before and after class decomposition, obtained by AlexNet using 10-fold cross-validation.

Learning mode: layer(s)	Pre-trained (AlexNet)			<i>DeTraC</i> _{AlexNet}		
	ACC (%)	SN (%)	SP (%)	ACC (%)	SN (%)	SP (%)
Shallow-tuning: only fc8	67.3	42.0	73.4	69.6	51.3	78.7
Fine-tuning: fc7-fc8	67.2	56.3	75.0	67.5	50.6	80.1
Fine-tuning: fc6-fc8	67.3	48.6	79.7	70.2	55.0	81.3
Fine-tuning: conv5-fc8	70.6	64.2	77.5	72.3	65.3	77.5
Fine-tuning: conv4-fc8	72.4	57.6	72.5	70.16	61.6	76.25
Fine-tuning: conv3-fc8	65.8	38.6	83.4	74.5	60.3	85.0

TABLE 4. Classification performance before and after class decomposition, obtained by VGG16 using 10-fold cross-validation.

Learning mode: layer(s)	Pre-trained (VGG16)			<i>DeTraC</i> _{VGG16}		
	ACC (%)	SN (%)	SP (%)	ACC (%)	SN (%)	SP (%)
Shallow-tuning: only fc8	65.0	68.3	61.6	67.4	60.1	78.7
Fine-tuning: fc7-fc8	65	67.6	65	72.6	69.13	75.0
Fine-tuning: fc6-fc8	70.2	71.3	69.0	69.6	70.2	80.0
Fine-tuning: conv5-3-fc8	65.0	70.3	66.3	68.2	65.33	73.75
Fine-tuning: conv5-2-fc8	61.5	71.6	54.8	70.4	65.6	81.25
Fine-tuning: conv5-1-fc8	61.7	69.3	58.0	72.6	71.3	80.0
Fine-tuning: conv4-3-fc8	62.3	75	62.2	70.3	64.0	75.0
Fine-tuning: conv4-2-fc8	61.5	81	64.5	72.6	70.33	80.0
Fine-tuning: conv4-1-f8	72.2	65.4	70.2	74.7	67.3	85.0

TABLE 5. Classification performance before and after class decomposition, obtained by VGG19, GoogleNet, and ResNet in case of shallow tuning.

	before			after (<i>DeTraC</i>)		
	Acc (%)	SN	SP	Acc (%)	SN	SP
Vgg19	67.4	73	62.6	70	68.3	71.6
GoogleNet	68.3	61.7	74.6	70.4	78.1	68.3
ResNet	72.4	68	75.3	75.4	83.3	82.5

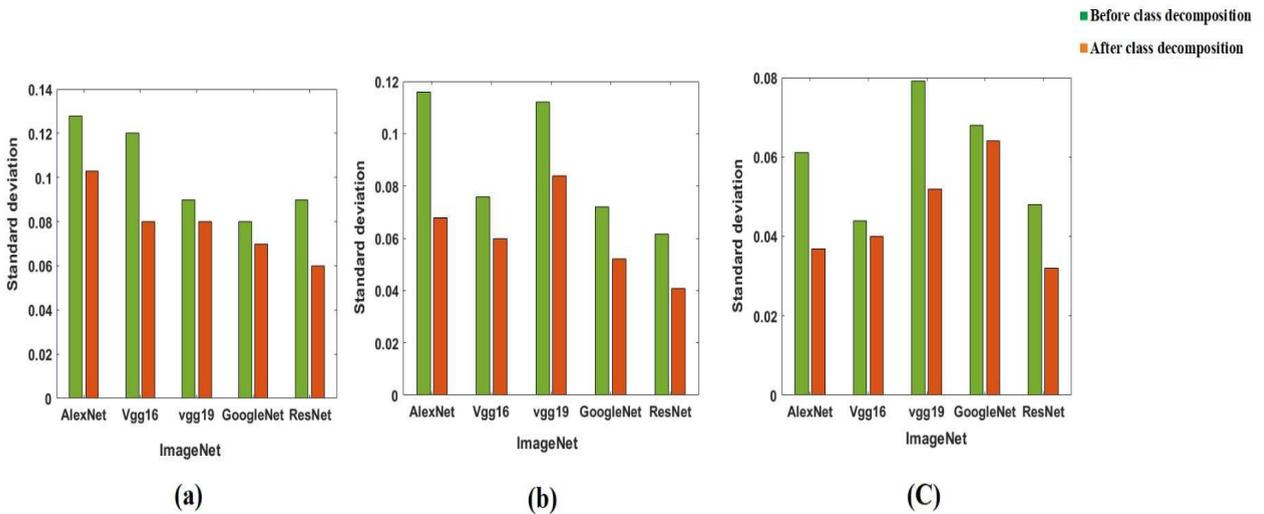


FIGURE 6. The standard deviation for ACC, SN, and SP obtained by *DeTraC* (after class decomposition) and standard pre-trained models (before class decomposition)

TABLE 6. Classification performance on X-ray dataset obtained by original knowledge transformations and shallow-tuned and fine-tuned versions of (*DeTraC*).

	Knowledge Transformation		<i>(DeTraC)</i>	
	ACC (%) w/o Aug.	ACC (%) w/ Aug.	Shallow-tuned ACC (%)	fine-tuned ACC (%)
AlexNet	66.8	80.71	68.2	99.2
Vgg16	65	78.43	67.5	97.6
Vgg19	67.4	77.39	70	98.8
GoogleNet	68.3	75.91	70.4	98.2
ResNet	72.4	82.24	75.4	99.8

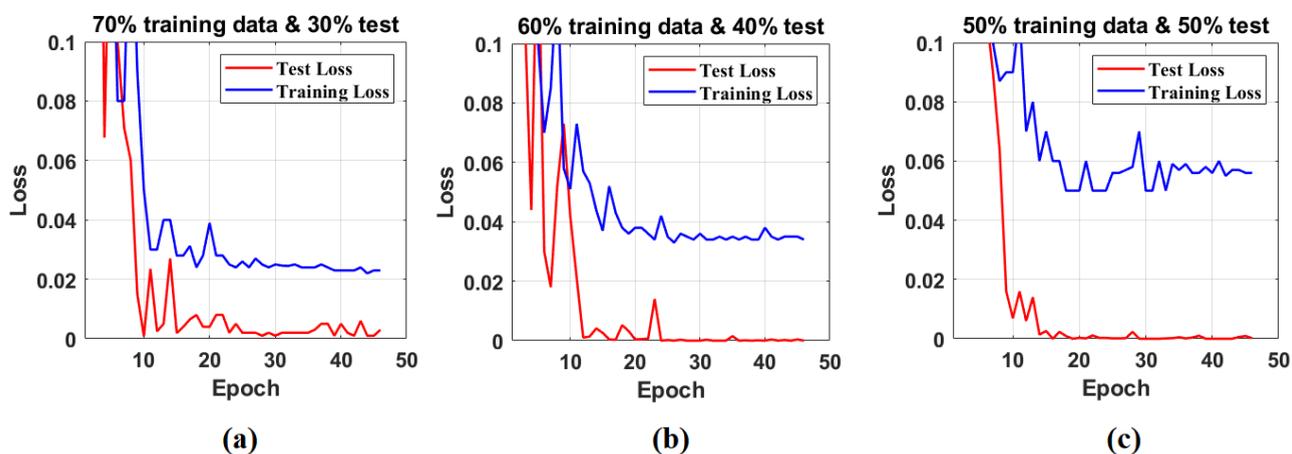


FIGURE 7. Loss error curves obtained by various sizes for the training set (blue curves) and testing set (red curves): (a) 70% Training and 30% test, (b) 60% Training and 40% test, and (c) 50% Training and 50% test.

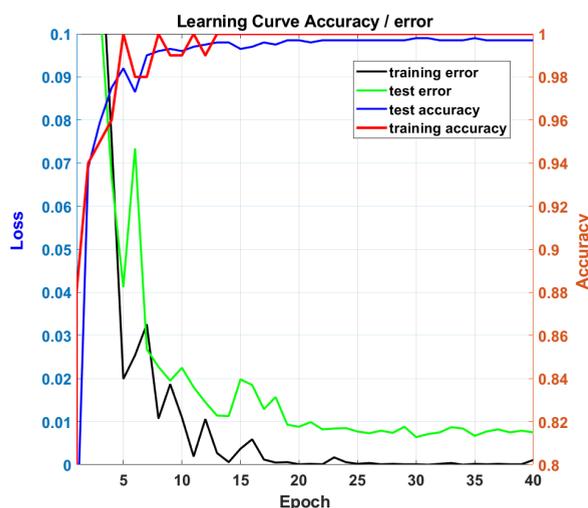


FIGURE 8. The learning curve accuracy and error obtained by AlexNet pre-trained network.

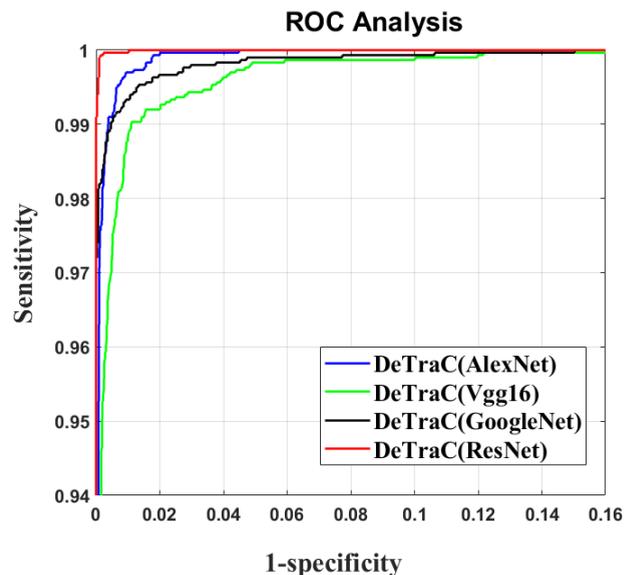


FIGURE 9. Comparison of ROC curves obtained by training *DeTraC* with deep tuning, based on different pre-trained models (e.g. AlexNet, VGG16, GoogleNet, ResNet).

TABLE 7. Classification accuracy obtained by *DeTraC* and other state-of-the-art related models.

Model	ACC (%)	SN (%)	SP (%)
<i>DeTraC_{AlexNet}</i>	99.2	98	98
<i>DeTraC_{Vgg16}</i>	97.6	98	97.4
<i>DeTraC_{GoogleNet}</i>	98.2	97	97.5
<i>DeTraC_{ResNet}</i>	99.8	98	98.5
[12]	95.2	97	93
[13]	96.4	95	98
[6]	93.3	91.4	100
[10]	94	100	93
[33]	94.4	91	98
[32]	89.90	91	89
[69]	93.33	94	916
[31]	94.56	96	94

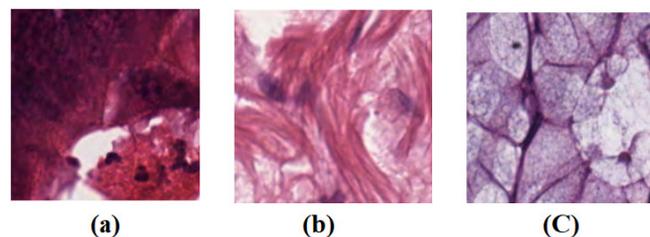
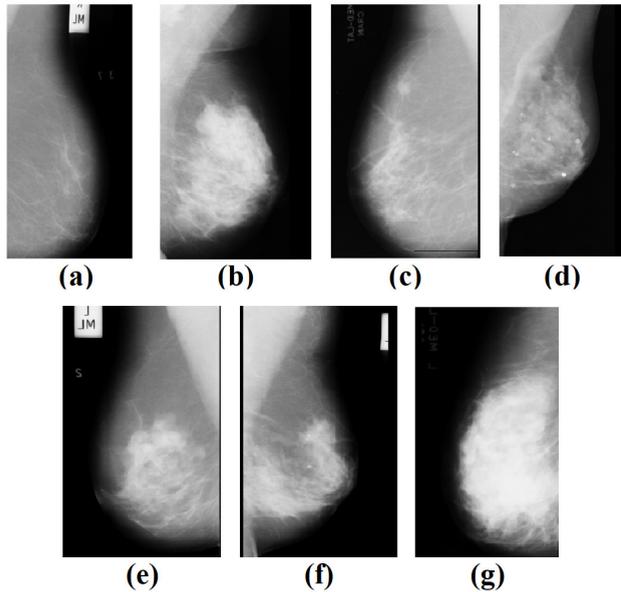


FIGURE 10. Example images from colorectal cancer dataset: (a) tumour epithelium (b) stroma, and (c) mucosal gland.

TABLE 8. The number of instances in original classes and after class decomposition

Dataset A	# Instances	Dataset B	# Instances
Tumor	625	<i>Tumor</i> ₁	272
		<i>Tumor</i> ₂	167
		<i>Tumor</i> ₃	186
Stroma	625	<i>Stroma</i> ₁	242
		<i>Stroma</i> ₂	199
		<i>Stroma</i> ₃	184
Mucosal	625	<i>Mucosal</i> ₁	94
		<i>Mucosal</i> ₂	276
		<i>Mucosal</i> ₃	255

**FIGURE 11.** Example images of digital mammograms: (a) normal, (b) Calcification (*CALC*), (c) Well-defined/circumscribed masses (*CIRC*), (d) Speculated masses (*SPIC*), (e) ill-defined masses (*MISC*), (f) Architectural distortion (*ARCH*) and (g) Asymmetry (*ASYM*).

per batch, L2-weight decay was applied with value 0.00001 and the default momentum in the network 0.9 is used. As demonstrated by Table 11, our model shows the highest accuracy of 99.8% with ResNet and 99.6% with GoogleNet. Moreover, Tables 9 and 11, show the ability of our framework to cope with data irregularities effectively where significant differences, in terms of accuracy and computational time, can be noticed before and after class decomposition with all used pre-trained models.

D. TIME COMPLEXITY

To illustrate the time complexity of DeTraC model, we describe the time complexity for each component: For the PCA, the time complexity is $O(p^3 + p^2n)$, where n is the number of data points, each one represented with p features, $O(p^2n)$ is the covariance matrix, and its eigen-value decomposition is $O(p^3)$. The time complexity for the k -means cluster is $O(t \times k \times n \times d)$, where t is number of iterations, k is the number of clusters or class decomposition component, n is the number of data points to be clustered and d is

the number of dimensions. Finally, the time complexity for the core component of DeTraC network (e.g. convolution operation) is $O\left(\sum_{l=1}^d n_{l-1} \times s_l^2 \times n_l \times m_l^2\right)$, where d is the number of convolution layer, n_l and s_l are the number and the size of filters respectively, n_{l-1} is the number of input channels of the l layer, and m_l is the size of the output feature map. Subsequently, the total time complexity of DeTraC is $O(n_{d-1} \times s_d^2 \times n_d \times m_d^2)$.

V. DISCUSSION

The historical conception of computer diagnosis systems from medical images has been comprehensively explored through several approaches ranging from feature engineering to feature learning. Deep Convolutional Neural Network (DCNN) is one of the most popular and effective approaches in the medical imaging domain, especially for classification problems. The superiority of DCNNs over classical/statistical machine learning approaches comes from their ability to learn different levels of abstraction in a hierarchical fashion directly from the images. Training DCNNs can be accomplished using two different strategies. They can be used as an end-to-end network, where an enormous number of annotated images must be provided (which is impractical in medical imaging). Alternatively, transfer learning usually provides an effective solution with the limited availability of annotated images by transferring knowledge from pre-trained CNNs (that have been learned from a bench-marked natural dataset) to the specific medical imaging task. Transfer learning can be further accomplished by three main scenarios: shallow-tuning, fine-tuning, or deep-tuning. However, data irregularities, especially in medical imaging applications, remain a challenging problem that usually results in miscalibration between the different classes in the dataset. Here, we propose an effective and yet efficient workflow solution, we call DeTraC, to deal with such a challenging problem by exploiting the advantages of class decomposition within the *CNN*s for image classification. DeTraC achieved high accuracies of 99.8% with ResNet on CXR images, 98.5% with VGG16 on digital mammograms, and 99.7% with GoogleNet on CRC images, when class-decomposition was applied with a deep-tuning strategy. The classification performance of DeTraC when the shallow-tuning mode was used, has shown a significant improvement based on several ImageNet pre-trained models before and after class-decomposition. Eventually, a full study has been provided to demonstrate the effect of class decomposition on knowledge transferred from the individual layers of AlexNet and VGG to confirm the robustness and superiority of DeTraC in all cases.

VI. CONCLUSION AND FUTURE WORK

In this paper, we propose a CNN architecture based on a class decomposition approach to improve the performance of ImageNet pre-trained CNN models using transfer learning. Our framework can provide effective and robust solutions for the classification of medical images and cope with data

TABLE 9. Classification performance and Elapsed Time (ET.) on colorectal cancer dataset before and after (*DeTraC*) class decomposition, based on different pre-trained networks.

	before					after (<i>DeTraC</i>)				
	ACC (%)	SN (%)	SP (%)	# Epochs	ET. (~h:m)	ACC (%)	SN (%)	SP (%)	# Epochs	ET. (~h:m)
AlexNet	97.7	97	97	46	2:28	98.3	98	99	47	1:54
Vgg16	97.8	98.2	97	31	16:00	99.8	98	98	32	14:00
GoogleNet	98.1	98	98	49	4:46	99.7	99	99.5	66	6:38
ResNet	98.3	98	98	43	3:38	99.1	99	99.8	39	4:12

TABLE 10. Samples distribution in digital mammograms database before and after class decomposition.

dataset A	ARCH		ASYM		CALC		CIRC		MISC		Norm		SPIC	
# instances	19		15		24		22		14		209		19	
dataset B	ARC_1	ARC_2	ASY_1	ASY_2	CA_1	CA_2	CI_1	CI_2	MI_1	MI_2	Nor_1	Nor_2	SPI_1	SPI_2
# instances	13	6	10	5	9	15	9	13	6	8	137	72	11	8

TABLE 11. Classification performance on digital mammograms dataset before and after (*DeTraC*) class decomposition, based on different pre-trained networks.

	before					after (<i>DeTraC</i>)				
	ACC (%)	SN (%)	SP (%)	# Epochs	ET. (~h:m)	ACC (%)	SN (%)	SP (%)	# Epochs	ET. (~h:m)
AlexNet	95.7	98	97	42	3 : 00	99.5	99	99	34	2 : 15
Vgg16	98.5	98	98	34	32 : 00	99	99.7	98	27	14 : 00
GoogleNet	97.1	97	98	63	9 : 00	99.6	99.7	98	43	3 : 30
ResNet	97.8	98	98.7	38	4 : 30	99.8	98	99	32	2 : 00

irregularity and the limited number of training samples too. The proposed method has been validated with three different datasets of chest X-ray, digital mammograms, and histological sections of human colorectal cancer. Several pre-trained architectures were integrated and tested with class decomposition and the experiments demonstrated the effectiveness with all architectures.

The work reported in this paper opens the door for a number of research directions. First, the adoption of class decomposition to other image classification tasks, especially in the area of medical imaging, where different data irregularities exist, thanks to the underlying complexities of the biological processes, is an interesting direction. Second, the optimisation of different hyperparameters (e.g. number of subclasses per class) required to carry out the class decomposition process has the potential to further enhance the performance of the models in terms of accuracy and convergence time. Finally, the application of class decomposition to other deep learning architectures like Recurrent Neural Networks (RNNs) and Long Short Term Memory (LSTM) for sequence models can be explored.

REFERENCES

- [1] C. Qin, D. Yao, Y. Shi, and Z. Song, "Computer-aided detection in chest radiography based on artificial intelligence: a survey," *Biomedical engineering online*, vol. 17, no. 1, pp. 1–23, 2018.
- [2] K.-L. Hua, C.-H. Hsu, S. C. Hidayati, W.-H. Cheng, and Y.-J. Chen, "Computer-aided classification of lung nodules on computed tomography images via deep learning technique," *OncoTargets and therapy*, vol. 8, 2015.
- [3] W. Sun, B. Zheng, and W. Qian, "Computer aided lung cancer diagnosis with deep learning algorithms," in *Medical imaging 2016: computer-aided diagnosis*, vol. 9785. International Society for Optics and Photonics, 2016, p. 97850Z.
- [4] S. A. El-Regaily, M. A. Salem, M. H. Abdel Aziz, and M. I. Roushdy, "Survey of computer aided detection systems for lung cancer in computed tomography," *Current Medical Imaging Reviews*, vol. 14, no. 1, pp. 3–18, 2018.
- [5] E. Dandil, M. Çakiroğlu, Z. Ekşi, M. Özkan, Ö. K. Kurt, and A. Canan, "Artificial neural network-based classification system for lung nodules on computed tomography scans," in *2014 6th International conference of soft computing and pattern recognition (SoCPAR)*. IEEE, 2014, pp. 382–386.
- [6] J. Kuruville and K. Gunavathi, "Lung cancer classification using neural networks for ct images," *Computer methods and programs in biomedicine*, vol. 113, no. 1, pp. 202–209, 2014.
- [7] A. Karargyris, J. Siegelman, D. Tzortzis, S. Jaeger, S. Candemir, Z. Xue, K. Santosh, S. Vajda, S. Antani, L. Folio et al., "Combination of texture and shape features to detect pulmonary abnormalities in digital chest x-rays," *International journal of computer assisted radiology and surgery*, vol. 11, no. 1, pp. 99–106, 2016.
- [8] S. Govindarajan and R. Swaminathan, "Analysis of tuberculosis in chest radiographs for computerized diagnosis using bag of keypoint features," *Journal of medical systems*, vol. 43, no. 4, p. 87, 2019.
- [9] G. Zhang, Z. Yang, L. Gong, S. Jiang, L. Wang, X. Cao, L. Wei, H. Zhang, and Z. Liu, "An appraisal of nodule diagnosis for lung cancer in ct images," *Journal of medical systems*, vol. 43, no. 7, p. 181, 2019.
- [10] T. Manikandan and N. Bharathi, "Lung cancer detection using fuzzy auto-seed cluster means morphological segmentation and svm classifier," *Journal of medical systems*, vol. 40, no. 7, p. 181, 2016.
- [11] P. Sangamithraa and S. Govindaraju, "Lung tumour detection and classification using ek-mean clustering," in *2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*. IEEE, 2016, pp. 2201–2206.
- [12] L. Sorensen, S. B. Shaker, and M. De Bruijne, "Quantitative analysis of pulmonary emphysema using local binary patterns," *IEEE transactions on medical imaging*, vol. 29, no. 2, pp. 559–569, 2010.
- [13] M. J. Gangeh, L. Sorensen, S. B. Shaker, M. S. Kamel, M. De Bruijne, and M. Loog, "A texton-based approach for the classification of lung parenchyma in ct images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2010, pp. 595–602.
- [14] J. Bi, T. Feng, and H. Yuan, "Real-time and short-term anomaly detection for gwac light curves," *Computers in Industry*, vol. 97, pp. 76–84, 2018.
- [15] J. Bi, H. Yuan, L. Zhang, and J. Zhang, "Sgw-scnn: An integrated machine learning approach for workload forecasting in geo-distributed cloud data centers," *Information Sciences*, vol. 481, pp. 57–68, 2019.
- [16] J. Bi, H. Yuan, and M. Zhou, "Temporal prediction of multiapplication consolidated workloads in distributed clouds," *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 4, pp. 1763–1773, 2019.

- [17] Y. Li, S. Wang, R. Umarov, B. Xie, M. Fan, L. Li, and X. Gao, "Deepre: sequence-based enzyme ec number prediction by deep learning," *Bioinformatics*, vol. 34, no. 5, pp. 760–769, 2017.
- [18] Z. Zou, S. Tian, X. Gao, and Y. Li, "mldeepr: Multi-functional enzyme function prediction with hierarchical multi-label deep learning," *Frontiers in genetics*, vol. 9, p. 714, 2018.
- [19] Y. Li, C. Huang, L. Ding, Z. Li, Y. Pan, and X. Gao, "Deep learning in bioinformatics: Introduction, application, and perspective in the big data era," *Methods*, 2019.
- [20] M. M. Ghazi, B. Yanikoglu, and E. Aptoula, "Plant identification using deep neural networks via optimization of transfer learning parameters," *Neurocomputing*, vol. 235, pp. 228–235, 2017.
- [21] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis," *Ieee Access*, vol. 6, pp. 9375–9389, 2017.
- [22] R. Tang, F. I. Tushar, S. Han, R. Hou, G. D. Rubin, and J. Y. Lo, "Classification of chest ct using case-level weak supervision," in *Medical Imaging 2019: Computer-Aided Diagnosis*, vol. 10950. International Society for Optics and Photonics, 2019, p. 1095017.
- [23] A. S. Qureshi, A. Khan, A. Zameer, and A. Usman, "Wind power prediction using deep neural network based meta regression and transfer learning," *Applied Soft Computing*, vol. 58, pp. 742–755, 2017.
- [24] P. Monkam, S. Qi, M. Xu, F. Han, X. Zhao, and W. Qian, "Cnn models discriminating between pulmonary micro-nodules and non-nodules from ct images," *Biomedical engineering online*, vol. 17, no. 1, p. 96, 2018.
- [25] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mousiopoulos, "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1207–1216, 2016.
- [26] A. Abbas and M. M. Abdelsamea, "Learning transformations for automated classification of manifestation of tuberculosis using convolutional neural network," in *2018 13th IEEE International Conference on Computer Engineering and Systems (ICCES)*, Dec 2018, pp. 122–126.
- [27] H. Xie, D. Yang, N. Sun, Z. Chen, and Y. Zhang, "Automated pulmonary nodule detection in ct images using deep convolutional neural networks," *Pattern Recognition*, vol. 85, pp. 109–119, 2019.
- [28] M. Negahdar and D. Beymer, "Lung tissue characterization for emphysema differential diagnosis using deep convolutional neural networks," in *Medical Imaging 2019: Computer-Aided Diagnosis*, vol. 10950. International Society for Optics and Photonics, 2019, p. 109503R.
- [29] S. Shen, S. X. Han, D. R. Aberle, A. A. Bui, and W. Hsu, "An interpretable deep hierarchical semantic convolutional neural network for lung nodule malignancy classification," *Expert Systems with Applications*, 2019.
- [30] J. H. Lam, Y. Li, L. Zhu, R. Umarov, H. Jiang, A. Héliou, F. K. Sheong, T. Liu, Y. Long, Y. Li et al., "A deep learning framework to predict binding preference of rna constituents on protein surface," *Nature communications*, vol. 10, no. 1, pp. 1–13, 2019.
- [31] S. Lakshmanaprabu, S. N. Mohanty, K. Shankar, N. Arunkumar, and G. Ramirez, "Optimal deep learning model for classification of lung cancer on ct images," *Future Generation Computer Systems*, vol. 92, pp. 374–382, 2019.
- [32] A. Nibali, Z. He, and D. Wollersheim, "Pulmonary nodule classification with deep residual networks," *International journal of computer assisted radiology and surgery*, vol. 12, no. 10, pp. 1799–1808, 2017.
- [33] J.-Z. Cheng, D. Ni, Y.-H. Chou, J. Qin, C.-M. Tiu, Y.-C. Chang, C.-S. Huang, D. Shen, and C.-M. Chen, "Computer-aided diagnosis with deep learning architecture: applications to breast lesions in us images and pulmonary nodules in ct scans," *Scientific reports*, vol. 6, p. 24454, 2016.
- [34] C. Wang, A. Elazab, F. Jia, J. Wu, and Q. Hu, "Automated chest screening based on a hybrid model of transfer learning and convolutional sparse denoising autoencoder," *Biomedical engineering online*, vol. 17, no. 1, p. 63, 2018.
- [35] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.
- [36] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "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.
- [37] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai et al., "Recent advances in convolutional neural networks," *Pattern Recognition*, vol. 77, pp. 354–377, 2018.
- [38] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.
- [39] G. Wang, J. Qiao, J. Bi, W. Li, and M. Zhou, "Tl-gdbn: Growing deep belief network with transfer learning," *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 2, pp. 874–885, 2018.
- [40] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.
- [41] M. Gao, U. Bagci, L. Lu, A. Wu, M. Buty, H.-C. Shin, H. Roth, G. Z. Papadakis, A. Depeursinge, R. M. Summers et al., "Holistic classification of ct attenuation patterns for interstitial lung diseases via deep convolutional neural networks," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 6, no. 1, pp. 1–6, 2018.
- [42] Anonymous, "Understanding and improving information transfer in multi-task learning," in *Submitted to International Conference on Learning Representations, 2020*, under review. [Online]. Available: <https://openreview.net/forum?id=SylzhkBTDB>
- [43] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, "Medical image classification with convolutional neural network," in *2014 13th International Conference on Control Automation Robotics & Vision (ICARCV)*. IEEE, 2014, pp. 844–848.
- [44] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Transactions on knowledge and data engineering*, vol. 21, no. 9, pp. 1263–1284, 2009.
- [45] N. V. Chawla, N. Japkowicz, and A. Kotcz, "Special issue on learning from imbalanced data sets," *ACM Sigkdd Explorations Newsletter*, vol. 6, no. 1, pp. 1–6, 2004.
- [46] J. Stefanowski, "Overlapping, rare examples and class decomposition in learning classifiers from imbalanced data," in *Emerging paradigms in machine learning*. Springer, 2013, pp. 277–306.
- [47] Y. Sun, M. S. Kamel, A. K. Wong, and Y. Wang, "Cost-sensitive boosting for classification of imbalanced data," *Pattern Recognition*, vol. 40, no. 12, pp. 3358–3378, 2007.
- [48] I. Polaka et al., "Clustering algorithm specifics in class decomposition," *No: Applied Information and Communication Technology*, 2013.
- [49] J. Wu, H. Xiong, and J. Chen, "Cog: local decomposition for rare class analysis," *Data Mining and Knowledge Discovery*, vol. 20, no. 2, pp. 191–220, 2010.
- [50] D. Fradkin, "Clustering inside classes improves performance of linear classifiers," in *2008 20th IEEE International Conference on Tools with Artificial Intelligence*, vol. 2. IEEE, 2008, pp. 439–442.
- [51] R. Vilalta, M.-K. Achari, and C. F. Eick, "Class decomposition via clustering: a new framework for low-variance classifiers," in *Third IEEE International Conference on Data Mining*. IEEE, 2003, pp. 673–676.
- [52] R. Vilalta and I. Rish, "A decomposition of classes via clustering to explain and improve naive bayes," in *European Conference on Machine Learning*. Springer, 2003, pp. 444–455.
- [53] Z. S. Abdallah and M. M. Gaber, "Kb-cb-n classification: Towards unsupervised approach for supervised learning," in *2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*. IEEE, 2011, pp. 283–290.
- [54] I. Polaka and A. Borisov, "Clustering-based decision tree classifier construction," *Technological and Economic Development of Economy*, vol. 16, no. 4, pp. 765–781, 2010.
- [55] R. Akbani, S. Kwek, and N. Japkowicz, "Applying support vector machines to imbalanced datasets," in *European conference on machine learning*. Springer, 2004, pp. 39–50.
- [56] E. Elyan and M. M. Gaber, "A genetic algorithm approach to optimising random forests applied to class engineered data," *Information sciences*, vol. 384, pp. 220–234, 2017.
- [57] —, "A fine-grained random forests using class decomposition: an application to medical diagnosis," *Neural computing and applications*, vol. 27, no. 8, pp. 2279–2288, 2016.
- [58] F. Wilcoxon, "Individual comparisons by ranking methods," in *Break-throughs in statistics*. Springer, 1992, pp. 196–202.
- [59] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, S. Y. Philip et al., "Top 10 algorithms in data mining," *Knowledge and information systems*, vol. 14, no. 1, pp. 1–37, 2008.
- [60] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
- [61] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

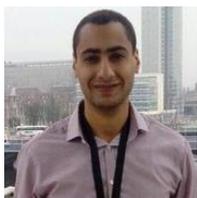
- [62] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1–9.
- [63] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
- [64] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information processing & management*, vol. 45, no. 4, pp. 427–437, 2009.
- [65] S. Candemir, S. Jaeger, K. Palaniappan, J. P. Musco, R. K. Singh, Z. Xue, A. Karargyris, S. Antani, G. Thoma, and C. J. McDonald, "Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration," *IEEE Transactions on Medical Imaging*, vol. 33, no. 2, pp. 577–590, Feb 2014.
- [66] S. Jaeger, A. Karargyris, S. Candemir, L. Folio, J. Siegelman, F. Callaghan, Z. Xue, K. Palaniappan, R. K. Singh, S. Antani, G. Thoma, Y. Wang, P. Lu, and C. J. McDonald, "Automatic tuberculosis screening using chest radiographs," *IEEE Transactions on Medical Imaging*, vol. 33, no. 2, pp. 233–245, Feb 2014.
- [67] J. N. Kather, C.-A. Weis, F. Bianconi, S. M. Melchers, L. R. Schad, T. Gaiser, A. Marx, and F. G. Zöllner, "Multi-class texture analysis in colorectal cancer histology," *Scientific reports*, vol. 6, p. 27988, 2016.
- [68] P. SUCKLING J, "The mammographic image analysis society digital mammogram database," *Digital Mammo*, pp. 375–386, 1994.
- [69] P. Hattikatti, "Texture based interstitial lung disease detection using convolutional neural network," in 2017 International Conference on Big Data, IoT and Data Science (BIG-IoT). IEEE, 2017, pp. 18–22.
- [70] A. Amanatiadis and I. Andreadis, "A survey on evaluation methods for image interpolation," *Measurement Science and Technology*, vol. 20, no. 10, p. 104015, 2009.



MOHAMED MEDHAT GABER. Mohamed is a Professor in Data Analytics at the School of Computing and Digital Technology, Birmingham City University. Mohamed received his PhD from Monash University, Australia. He then held appointments with the University of Sydney, CSIRO, and Monash University, all in Australia. Prior to joining Birmingham City University, Mohamed worked for the Robert Gordon University as a Reader in Computer Science and at the University of Portsmouth as a Senior Lecturer in Computer Science, both in the UK. He has published over 200 papers, co-authored 3 monograph-style books, and edited/co-edited 6 books on data mining and knowledge discovery. His work has attracted well over four thousand citations, with an h-index of 37. Mohamed has served in the program committees of major conferences related to data mining, including ICDM, PAKDD, ECML/PKDD and ICML. He has also co-chaired numerous scientific events on various data mining topics. Professor Gaber is recognised as a Fellow of the British Higher Education Academy (HEA). He is also a member of the International Panel of Expert Advisers for the Australasian Data Mining Conferences. In 2007, he was awarded the CSIRO teamwork award.



ASMAA ABBAS. Asmaa was born in Assiut city, Egypt. She finished her B.Sc. in Computer Science from the Department of Mathematics at Assiut University in 2004. Asmaa has got a Preliminary master degree from the Mathematics Department at the Assiut University in 2016. She is currently working as a Research Assistant at Mathematics Department in Assiut University. Her research interests include deep learning, medical image analysis, and data mining.



MOHAMMED ABDELSAMEA. Abdelsamea is currently a Senior Lecturer in Data and Information Science at the School of Computing and Digital Technology, Birmingham City University. Before joining BCU, he worked for the School of Computer Science at Nottingham University, Mechanochemical Cell Biology at Warwick University, Nottingham Molecular Pathology Node (NMPN), and Division of Cancer and Stem Cells both at Nottingham Medical School, as a Research Fellow. In 2016, he was a Marie Curie Research Fellow at the School of Computer Science at Nottingham University. Before moving to the UK, he worked as a Lecturer of Computer Science for Assiut University, Egypt. He was also a Visiting Researcher at Robert Gordon University, Aberdeen, UK. Abdelsamea has received his Ph.D. degree (with Doctor Europaeus) in Computer Science and Engineering from IMT - Institute for Advanced Studies, Lucca, Italy. His main research interests are in computer vision including: image processing, deep learning, data mining and machine learning, pattern recognition, and image analysis.