This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2020.2971258, IEEE Transactions on Medical Imaging

Unsupervised Domain Adaptation to Classify Medical Images using Zero-bias Convolutional Auto-encoders and Context-based Feature Augmentation

Euijoon Ahn, *Student Member, IEEE*, Ashnil Kumar, *Member, IEEE*, Michael Fulham, Dagan Feng, *Fellow, IEEE*, Jinman Kim, *Member, IEEE*

***Abstract*— The accuracy and robustness of image classification with supervised deep learning are dependent on the availability of large-scale labelled training data. In medical imaging, these large labelled datasets are sparse, mainly related to the complexity in manual annotation. Deep convolutional neural networks (CNNs), with transferable knowledge, have been employed as a solution to limited annotated data through: 1) fine-tuning generic knowledge with a relatively smaller amount of labelled medical imaging data, and 2) learning image representation that is invariant to different domains. These approaches, however, are still reliant on labelled medical image data. Our aim is to use a new hierarchical unsupervised feature extractor to reduce reliance on annotated training data. Our unsupervised approach uses a multi-layer zero-bias convolutional auto-encoder that constrains the transformation of generic features from a pre-trained CNN (for natural images) to non-redundant and locally relevant features for the medical image data. We also propose a context-based feature augmentation scheme to improve the discriminative power of the feature representation. We evaluated our approach on 3 public medical image datasets and compared it to other state-of-the-art supervised CNNs. Our unsupervised approach achieved better accuracy when compared to other conventional unsupervised methods and baseline fine-tuned CNNs.**

***Index Terms*— Convolutional auto-encoders, convolutional neural networks, unsupervised domain adaptation, unsupervised feature learning.**

I. Introduction

SUPERVISED learning is the state-of-the-art in situations where large-scale labelled data are available [1-4]. In medical imaging, however, there is limited availability of large labelled datasets due to the complexity (time, cost, workforce) of annotation and inter- and intra-observer variability among clinicians [5]. In this setting, the concept of 'transfer learning' was introduced, whereby image-related features from different domains, e.g., image features learned from natural photographic images, were adopted [6]. These image features, however, are often generic and do not perform well in specific medical image analysis problems [6]. An alternative approach is to optimise these features by retraining the generic features using a relatively small set of labelled medical images. This 'fine-tuning' approach, however, is not able to match the overall accuracy of learning image features directly from specific, large labelled data. Another approach is to increase the number of training samples through data augmentation. In this context, existing labelled data are transformed to new visual data through numerous image manipulation operations that include shifting, scaling, cropping, affine transformation, and rotation [7]. A major drawback of data augmentation is the reliance on the optimal design of the transformations where domain-specific knowledge is needed [8]. Alternatively, domain adaptation methods (DAs) learn representations that are invariant to domains with different data distributions. DAs use all available unlabelled data from the target domain to learn image features to better represent them rather than using labelled data from the source domain alone. Successful DAs would reduce reliance on labelled data [9, 10]. Although DAs have been applied to object recognition, image segmentation and classification [11, 12], using DAs in medical image analysis is non-trivial because of the scarcity of labelled source data.

II. Related Work

A. Transfer Learning

Deep learning methods, in particular convolutional neural networks (CNNs), are playing an increasingly important role in medical image analysis, due to their ability to learn discriminative features directly from the image data. In early work, pre-trained CNNs were used as generic image feature extractors in medical image classification problems [6, 13-16].

E. Ahn, A. Kumar, D. Feng and J. Kim are with the School of Computer Science, University of Sydney, NSW, Australia. E. Ahn (E-mail: eahn4614@uni.sydney.edu.au); A. Kumar (E-mail: ashnil.kumar@sydney.edu.au); J. Kim (E-mail: jinman.kim@sydney.edu.au).

M. Fulham is with the Department of Molecular Imaging, Royal Prince Alfred Hospital, Camperdown, NSW, Australia, and also with the Sydney Medical School, University of Sydney, Camperdown, NSW, Australia (E-mail: michael.fulham@sydney.edu.au).

D. Feng is also with Med-X Research Institute, Shanghai Jiao Tong University, China (E-mail: dagan.feng@sydney.edu.au).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2020.2971258, IEEE Transactions on Medical Imaging

Anavi et al. [17] reported that generic (natural image) CNN features achieved higher accuracy in chest X-ray classification compared to local texture and binary patterns. Similarly, Phan et al. [13] used generic CNN features to classify HeP-2 cells and Ahn et al. [6] to categorise X-ray images. Other investigators fine-tuned these generic features via supervised training (using a smaller set of labelled medical image data) and showed they could perform as well as features obtained by training a CNN from scratch and outperformed CNNs trained with limited labelled training data [18-20]. Litjens et al. [21] recently published a comprehensive review of transfer learning approaches in medical image analysis.

B. *Domain Adaptation Methods (DAs)*

DAs explore scenarios where the labelled data are shared across the source and target domain or where labelled data are not available in target domain [22-24]. Bermudez-Chacon et al. [22] leveraged the labelled source data (annotated organelles from different parts of a mouse brain) from 3D Electron Microscopy (EM) to learn a classifier for segmenting organelles from other target EM data that were unlabelled. Adversarial networks (ANs) can also be used to learn domain invariant image representation for object localisation and lesion segmentation [23, 24]. Heimann et al. [23] derived an ANs to increase the size of training data and showed that it improved the localisation of transducer in X-ray fluoroscopy images. Similarly, Kamnitsas et al. [24] exploited ANs to improve the lesion segmentation accuracies in magnetic resonance imaging (MRI). The key concept with these approaches is to learn features that are as discriminative as possible in the source domain and as indistinguishable as possible between the source and target domains. However, data acquisition in the medical domain is difficult (especially for the source domain) and quality labelling is time consuming.

C. *Data Augmentation*

Data augmentation is a well-established technique for supervised deep learning approaches, especially CNNs, because it allows the generation of new labelled data from existing labelled data through label preserving image manipulations. A 10-fold augmentation scheme involving cropping and flipping is a common approach to reduce overfitting in supervised CNNs [25]. Similar approaches have been used in medical image modality classification [20] and to improve lesion segmentation accuracy in dermoscopic images [26]. These methods augment data by creating new images based on a single source.

D. *Unsupervised Feature Learning*

Unsupervised feature learning algorithms [27-33] have been used to learn features from unlabelled large-scale image datasets in recent years. The image features are generally learned using unsupervised algorithms such as sparse coding (SC) [27], sparse auto-encoder (AE) [28] and Restricted Boltzmann Machines (RBMs) [29]. Many unsupervised algorithms, however, have often been limited to learning low-level features such as lines and edges. This is mainly attributed to using decomposed image patches that cannot effectively learn global structures and local connections between image content [34, 35]. Convolutional sparse coding and convolutional auto-encoders (CAEs) extend the original patch-based approaches to cope with multidimensional and large-sized images. The CAE, in particular, can efficiently learn global image features using multidimensional filters. Unlike patch-based methods, it is capable of preserving the relationships between neighbouring contents and spatial information.

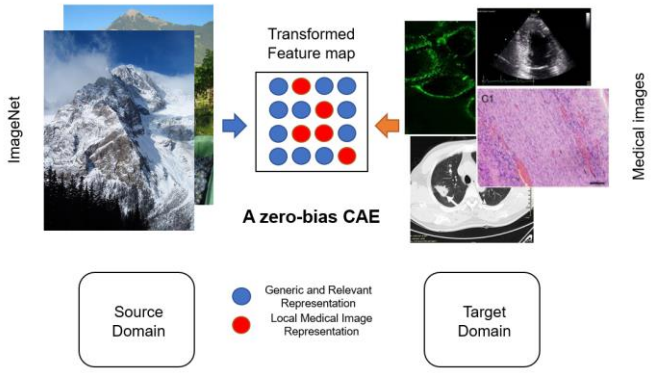

Fig. 1. Concept of our unsupervised domain adaptation approach.

E. *Contribution*

We hypothesise that one way to learn image representation is by maximising relevant and generic shared representations between two domains together with local image characteristics from the target domains. Our approach is to use a deep unsupervised feature extractor to transform the feature maps from the pre-trained network to a set of non-redundant and relevant medical image features. Our concept is outlined in Fig. 1. Our feature extractor incorporates a zero-bias CAE that extends a CNN pre-trained on natural images. Our approach preserves meaningful generic features from the pre-trained domain and learns specific local features that are more representative of the characteristics of the medical image data. We also introduce a context-based feature augmentation that uses the distribution of features in similar images to improve the discriminative power of the feature representation.

In previous preliminary work that was presented at International Symposium on Biomedical Imaging (ISBI) 2019 as a conference paper we proposed using a single-layer zero-bias CAE for unsupervised feature learning and showed it was useful in medical image modality classification with a single dataset (Ahn et al. [36]). In this paper, we present a major extension of the previous work and now introduce: a) a new context-based feature augmentation scheme which proportionally combines the feature representation of other similar images within the feature space; b) a multi-layer zero-bias CAE which provides additional learning capacity to characterise variations between image features through a deeper network and consequently, enhances the discriminative power of the learned feature representation; and c) a comprehensive evaluation of our approach against state-of-the-art unsupervised and supervised methods across different public datasets and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2020.2971258, IEEE Transactions on Medical Imaging

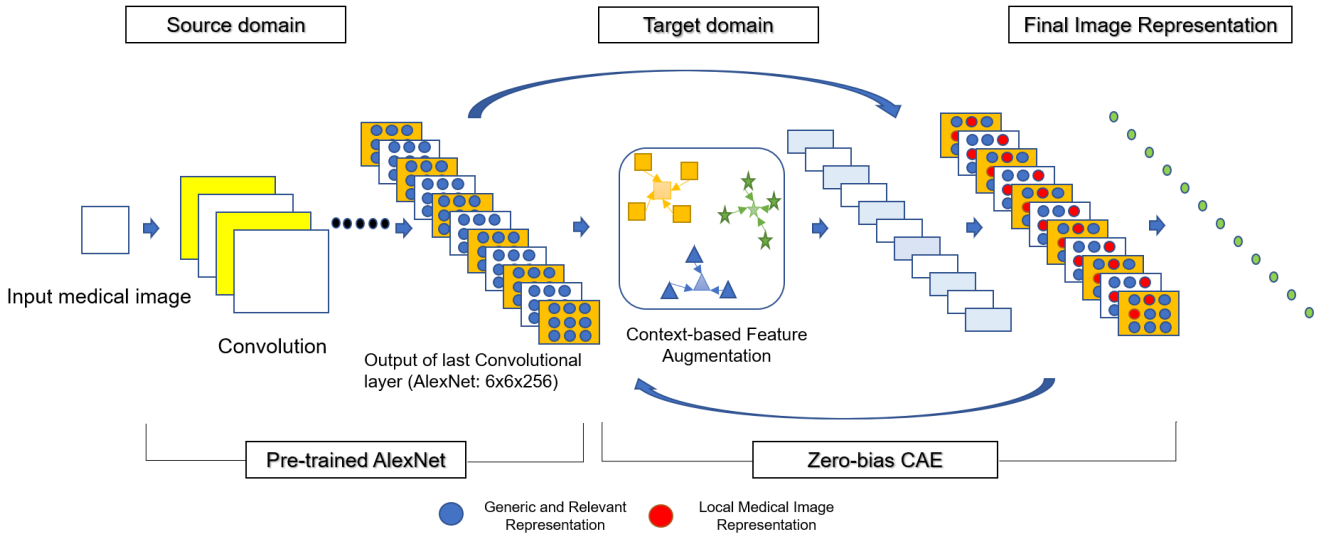

Fig. 2. A schematic of our single-layer zero-bias CAE.

problems: medical image modality classification, skin disease classification and the detection of multi-drug-resistant (MDR) tuberculosis (TB).

III. Materials and Methods

A. Materials

We used 3 public datasets for our experimental analysis. Each dataset was used to assess the performance of our method on 3 different problems:

1) Medical Image Modality Classification

We used the Subfigure Classification dataset used in the Image Conference and Labs of the Evaluation Forum (ImageCLEF) 2016 competition [37, 38]. The dataset has 6776 training images, 4166 test images from 30 different classes comprising of diagnostic images (18 categories) and generic biomedical illustrations (12 categories). Ground truth annotations are included with the dataset. It is used as a benchmark task because of the challenging nature of the diverse dataset: the content of an image can vary greatly depending on the imaging modality used to generate images and on an individual disease that changes the normal appearance / morphology of the involved organs / structures. Hence, we used this task with a public modality classification dataset to enable comparison with existing classification algorithms [20, 30, 33, 39]. Further, research suggests that the ability to automatically search for medical images (e.g., through content-based image retrieval) from wide spectrum of imaging modalities could be used for evidence-based diagnosis, teaching and biomedical research [40-42].

2) Skin Disease Classification

We used the skin disease classification dataset used in International Skin Imaging Collaboration (ISIC) 2017 competition [43]. The dataset contains 2000 training images and 600 test images with 3 different skin conditions - benign nevi, seborrheic keratoses and melanoma. Ground truth annotations and pathology reports were available [43]. The dataset provides various dermoscopic images with complex skin conditions.

3) MDR TB Detection

TB continues to be a serious lung disease that has significant morbidity and mortality worldwide, especially in undeveloped countries [44]. TB causes destruction of the lung parenchyma and these changes are best depicted on anatomical imaging with chest computed tomography (CT). Although clinical assessment is fundamental to patient management in TB, chest CT is the imaging modality of choice to also assess the extent of the lung involvement and also the response to therapy. TB can be cured using anti-TB drugs if treated early, so-called drug-sensitive (DS) TB. However, there has been the emergence of TB that is resistant to the usual multiple drugs (MDR TB) and it is important to recognize these patients as a different approach to treatment is required. Further, because these cases occur in undeveloped countries where although imaging may be available at a regional centre there is often not the imaging specialist expertise to reliably interpret the imaging data. Hence, an accurate and reliable computer aided diagnosis (CAD) system in chest CTs of TB patients and, in particular, with MDR TB, offers the potential for automated detection of failed conventional therapy. We used the MDR detection dataset from the ImageCLEF 2017 competition [45]. The 3D CT images from patients with MDR TB are complex, and heterogeneous and vary from patient to patient. The dataset has 230 patients in the training dataset and 214 patients in the test dataset. The TB patients included are those with DS and MDR disease. We used the annotated data where the ground truth class, DS or MDR, was provided for each patient. Ground truth annotations were only available for the training dataset. We randomly divided the training data into 5 groups, learning the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2020.2971258, IEEE Transactions on Medical Imaging

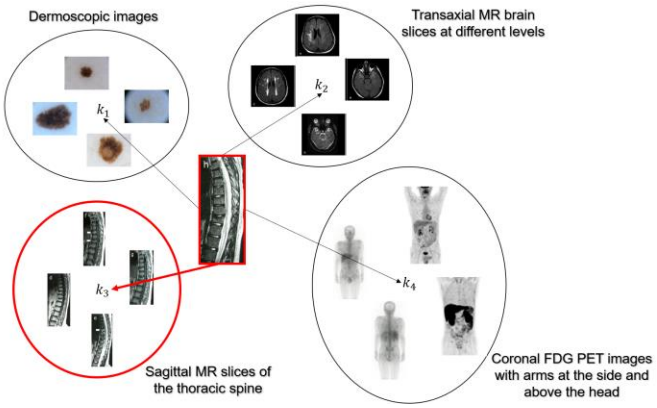

Fig. 3. An illustration of our context learning for feature augmentation; the image feature is augmented by proportionally combining the feature representation of other similar features. For the purpose of visual clarity, we removed an arrow, labels, lines at the top and side of some PET images that relate to the way the images have been cropped for the competition. The unedited version of this figure is in the Supplementary Materials.

model on 80% and using the model to predict MDR in the remaining 20%. We repeated this process 5 times using each set (5-fold cross-validation).

B. Overview of our Unsupervised Approach

A schematic of our approach is shown in Fig. 2. In brief, we used a pre-trained CNN from a different domain and fed the output of the last convolutional layer into a new convolutional layer where the weights of the layer are learned using a zero-bias CAE. The generic image features extracted from the pre-trained CNN were augmented in the feature space by proportionally combining the feature representation of other similar images (neighbours in the feature space). The final feature representation was then extracted in a feed-forward and hierarchal manner.

C. A zero-bias CAE

The CAE has an encoder and a decoder that takes an input $x_{1..,p} \in \mathbb{R}^m$, the output of the n-th feature map can be expressed as:

$$z^n = \sigma(x \quad * W_e^n + b_e^n) \tag{1}$$

where σ is the ReLU activation function, $*$ denotes 2D convolution and $\theta_e = \{W_e, b_e\}$ are the parameters (weights and bias) of the encoder. Here, the bias is processed per feature map. The reconstruction then occurs as follows:

$$\mathbf{y} = \sigma\left(\sum_{n \in M} z^n * \widetilde{W}_d^n + b_d^n\right) \tag{2}$$

where $\widetilde{W}$ is 180° flipped weight matrix, $\theta_d = \{W_d, b_d\}$ are the parameters of the decoder, and M represents the group of latent feature maps. The two parameter sets are generally constrained to have a form of $W_d = W_e^{\mathrm{T}}$, preventing degenerated features from being learned. The cost function to minimise the loss was then implemented as:

$$\mathrm{E}(\theta) = \min_{W\, b} \frac{1}{2} \sum_{i=1}^{p} \|x_i - y_i\|_2^2. \tag{3}$$

As in a standard backpropagation algorithm, the reconstruction error gradient is first back propagated and then the weights are updated using stochastic gradient descent (SGD). Unlike Classic AEs where local connections between image content are ignored, the CAE captures and shares localised information among all locations in the input.

The hidden unit biases that take large negative values when training a CAE can make it difficult to learn non-trivial image characteristics [46, 47]. Negative biases, in conjunction with activation functions such as sigmoid or ReLU, can lead to learning a point attractor rather than multidimensional regions in highly semantic and sparse medical images. This is problematic so we used a zero-bias ReLU activation that fixes the biases (b) of our convolutional and deconvolutional layers to zero at encoding-time.

D. Integration of a Pre-trained CNN

We used AlexNet [48], pre-trained using ImageNet, as the source domain CNN for extracting generic image features from medical images. AlexNet has eight trainable layers: 5 convolutional and 3 fully connected layers. We removed the last 3 fully connected layers and used the output of the last convolutional layer of the pre-trained CNN as an input to subsequent zero-bias CAE learning. AlexNet can take RGB images with a size of 227 × 227 pixels as inputs and so we re-scaled our medical images to this resolution.

E. Multi-layer Zero-bias CAEs

Our zero-bias CAE can be learned in a hierarchical fashion for a potentially improved meaningful image feature representation. The input feature map of $l + 1$ can be computed by applying the convolution operation and learned weights using our zero-bias CAE from layer l. A multi-layer zero-bias CAE is learned in a feedforward manner, using given network parameters such as kernel and filter size for each layer.

F. Context-based Feature Augmentation

Our data augmentation method operates in the feature space. A new augmented feature is derived by considering other nearby (i.e., similar) features. We used the *K*-means algorithm to cluster all the image features into *K* clusters. Data pre-processing, dimensionality reduction and normalisation are crucial to clustering [49]. Since raw data tends to generate many highly correlated cluster centroids rather than the centroids that capture diverse characteristics across whole data we used principal component analysis (PCA) to remove these correlations, followed by l_2 normalisation [50]. The effective initialisation of the centroids in *K*-means is important to avoid empty clusters so we used the setup of the initialisation scheme introduced by Arthur et al. [51] to find centroids that are maximally different.

The new augmented image feature $\tilde{x}_i$ was created by proportionally combining the feature representation of other similar features in the same cluster and the original image feature of x_i at feature i (see Fig. 3). This is defined as follows:

$$\tilde{x}_i = \tau \sum_{j=1}^{N_c} w_{ik_j} \tilde{x}_{k_j} + (1-\tau) x_i \,, \tag{4}$$

where $[k_1, k_2, k_3, \cdots, k_{N_c}]$ indicate the N_c feature labels in cluster k and τ is a weight parameter. The similarity weight w_{ik_j} of each feature i was then created by its normalized similarity according to:

$$w_{ik_j} = \frac{\exp\left(-\frac{\left\|x_i - x_{k_j}\right\|^2}{2\sigma_X{}^2}\right)\left(1-\delta(k_j - i)\right)}{\sum_{j=1}^{N_c} \exp\left(-\frac{\left\|x_i - x_{k_j}\right\|^2}{2\sigma_X{}^2}\right)} \tag{5}$$

where the $\sigma_X{}^2$ denotes the sum of the variance in each of x and $\delta(.)$ is the indicator function. To capture more diverse image characteristics, we created augmented feature sets by setting 10 different number of clusters (k). The number of k controls the size of similar features. The parameter τ which we named as augmentation rate, controls the feature closeness between x_i and $\tilde{x}_i$. Multiple different clusters allow the extraction of more diverse image characteristics and reduce the risk of augmenting irrelevant image features. Our augmented training dataset was 11x larger than the original training dataset (10 augmented clusters plus the original training set).

IV. Experimental Setup

A. Evaluation

We evaluated our approach by comparing it to other unsupervised and supervised learning methods, as follows:

1) Conventional unsupervised feature learning methods: SC [27], Independent Component Analysis (ICA) and 2-layer stacked sparse auto-encoder (SSAE). We set the number of filters (i.e., weights) for the first layer of the SC, ICA and SSAE to 1600 and the filters for the second layer of SSAE to 1024.

TABLE I
The network parameters of our zero-bias CAE.

Type	Size
Kernel	3x3
Convolutional Stride	1
Pad	1
Pool	2
Filter	4096

TABLE II
The parameters of our Context-based Feature Augmentation.

Dataset	Minimum to maximum number of clusters (k)	Augmentation rate (τ)
ImageCLEF 2016	30-75*	0.5
ISIC 2017	5-50*	0.5
ImageCLEF 2017	30-75*	0.5

* The increment of 5 was used between the minimum and maximum for each cluster size.

2) State-of-the-art unsupervised feature learning methods: convolutional sparse kernel network (CSKN) [30] and *K*-means-based CNN [33].
3) State-of-the-art supervised pre-trained CNNs (with natural images): AlexNet [48], VGG [1], GoogLeNet [3], and ResNet [2]. These CNNs have achieved high rankings in object recognition and localisation from the ImageNet Challenge. For all pre-trained CNNs models, the last convolutional layer and final fully-connected layers were used as the feature extractors.
4) State-of-the-art supervised fine-tuned CNNs: we used the same models as in the pre-trained baselines above: AlexNet [48], VGG [1], GoogLeNet [3] and ResNet [2, 39]. These models were dominantly used in all ImageCLEF 2016, ISIC 2017 and ImageCLEF 2017 competitions. The fine-tunned AlexNet for ImageCLEF 2017 dataset were trained for 60 epochs. We used the same batch size and learning rate as the training for our model (see Section IV.B).

B. Implementation Details

We trained our zero-bias CAE for 150 epochs with a batch size of 512 and an initial learning rate of 10^{-5}. We used learning rate annealing, decaying the rate by a factor of 10 when the error plateaued. Table I shows the network parameters. The main parameter is the size of filters in the zero-bias CAE network. We used an empirical process to discover the appropriate filter size (via trials using size of 512, 1024, 2048, 4096 and 8192) and set 4096 in our all experiments. We also empirically set the number of layers as 2 (after pilot tests on 2-7 layers). We used the same network parameters defined in the first layer of our zero-bias CAE for subsequent convolutional layers (see Table I). We used the same settings for all three datasets.

Our context-based feature augmentation has two parameters; the number of clusters (k) and augmentation rate (τ) need to be determined. We set the minimum and maximum number of clusters by observing the distributions of the image samples in clusters for all 3 datasets. The augmentation rate (τ) was set to 0.5 to ensure that 50% of the original features are still preserved. The parameters we used are shown in Table II.

For all learned features (SC, ICA, SSAE and our approach), we used the setup of the multi-class linear SVM introduced by Yang et al. [52], who used a differentiable quadratic hinge loss so that the training could easily be done with simple gradient-based optimisation methods. We used the standard Limited memory Broyden Fletcher Goldfarb Shanno (LBFGS) with a learning rate of 0.1 and a regularization parameter of 1, consistent with the parameters specified by [30, 52].

For MDR detection from 3D CT images, we extracted image slices from each patient volume in the axial plane and used these 2D slices for model training. We removed the slices that did not contain lungs, using the lung masks provided from the competition organiser [53] and this is consistent to what is done for other studies [54]. We averaged the prediction scores of image slices in the volume for each patient study as the final prediction scores.

C. Evaluation Metrics

We used the top 1 accuracy (the correctness of the predicted label), which is the standard performance measure adopted in

TABLE III
TOP 1 IMAGE CLASSIFICATION RESULTS ON IMAGECLEF 2016 DATASET (MEDICAL IMAGE MODALITY CLASSIFICATION) COMPARED TO OTHER METHODS.

Methods	Accuracy (%)
SC + SVM	57.08
ICA + SVM	58.79
SSAE + SVM	65.17
CSKN [30]	70.99
K-means-based CNN [33]	74.51
Pre-trained AlexNet (LC) + SVM	68.41
Pre-trained GoogLeNet (FC) +SVM	78.61
Pre-trained AlexNet (FC) + SVM	79.21
Fine-tuned AlexNet + SVM [20]	79.60
Fine-tuned GoogLeNet + SVM [20]	80.75
An ensemble of fine-tuned AlexNet and GoogLeNet [20]	82.48
Our method	82.98
An ensemble of fine-tuned GoogLeNet and ResNet [62]	83.14
Synergic ResNet [39]	87.97

LC is last convolutional layer and FC is final fully connected layer.

recent CNN studies for the medical image modality classification [20], and the area under curve (AUC) from the receiver operating characteristics (ROC) curve as our evaluation metrics for skin disease classification [43] and MDR detection [45].

D. Training Computation

It took 12 hours for our 2-layer zero-bias CAE to be trained with a GeForce GTX 1080 Ti GPU (11GB memory). The required computational cost to train an individual convolutional layer in our zero-bias CAE can be approximated as f^2mmco floating point operations per second (FLOPS) [55, 56], where f denotes the size of 2D convolutional kernel, $m \times m \times c$ is the size of input feature map and o is the filter size of output feature map. The required computational cost of our zero-bias CAE is equivalent to that of standard supervised convolutional neural networks (CNNs) such as AlexNet.

V. RESULTS

The results of the image modality classification experiments are shown in Table III. Our approach had greater accuracy than other unsupervised feature learning methods and other pre-trained CNN models; there was an over 10% improvement from the baseline pre-trained AlexNet. When compared to other baseline fine-tuned CNNs, our approach had a top 1 accuracy of 82.98%. The best performing approach was the synergic deep learning model that used multiple fine-tuned ResNets (87.97%) [39]. Table IV shows that the larger size of filters generally improved the final feature representation at the cost of increased computational complexity.

The results of skin disease classification experiments are shown in Table V. We show visual results of skin disease classification in Fig. 4. Our approach had greater accuracy than other unsupervised methods, achieving a mean AUC accuracy of 83.15%. Furthermore, it also performed better than other baseline fine-tuned CNNs, consistent to the results of the image modality classification experiments. The top performing methods were all based on well-established supervised CNNs including AlexNet [57], GoogLeNet [57], Inception v3 [58] and ResNet [59, 60]. The best method reported in the competition was the fine-tuned ResNet with a mean AUC accuracy of 91.10%. An example of how our zero-bias CAE constrained the transformation of generic features from the pre-trained AlexNet to relevant visual features of melanoma is shown in Fig. 4.

The results of MDR detection from 3D CT images are shown in Table VI. Our approach had a greater AUC than other pre-trained CNNs, with an AUC of 65.52%. This result was an improvement from the baseline pre-trained AlexNet at 59.3% and the fine-tuned AlexNet at 59.85%. The top performing methods in the Competition also used fine-tuned CNNs e.g. ResNet [54]. The best method was a non-CNN-based approach using 3D texture features, with an AUC of 58.3% [61].

Fig. 5 shows the relative improvement in classification accuracy due to extraction of more relevant feature representations from our zero-bias CAE with context-based feature augmentation scheme. Our experiments using 2-layer zero-bias CAE also had classification improvements across all 3 datasets (see Fig. 5). This is attributed to its ability to extract discriminative medical image features that are more representative of the local imaging characteristics, which may

TABLE IV
RESULTS OF IMAGE MODALITY CLASSIFICATION PERFORMANCE WITH DIFFERENT FILTER SIZES WITH/WITHOUT OUR CONTEXT-BASED FEATURE AUGMENTATION.

Filter size / Top 1 Accuracy (%)	Without Feature Augmentation	With Feature Augmentation
512	77.00	77.15
1024	79.48	79.99
2048	80.15	80.33
4096	81.33	82.36
8192	80.49	81.04

1-layer zero-bias CAE

TABLE V
AUC RESULTS ON ISIC 2017 DATASET (SKIN DISEASE CLASSIFICATION) COMPARED TO OTHER METHODS.

Methods	Seborrheic keratosis AUC (%)	Melanoma AUC (%)	Mean AUC (%)
ICA + SVM	49.00	58.62	53.81
SC + SVM	66.87	56.88	61.87
SSAE + SVM	68.28	60.98	64.63
Pre-trained (FC) ResNet-152 + SVM	79.43	65.24	72.35
Pre-trained (FC) VGG-19 + SVM	80.20	73.14	76.67
Fine-tuned Inception v3 [58]	81.70	68.40	75.00
Fine-tuned AlexNet + SVM	82.19	72.02	77.10
An ensemble of fine-tuned AlexNet and GoogLeNet [57]	84.00	80.50	82.30
Our method	88.08	78.21	83.15
Fine-tuned ResNet-101 [60]	87.00	92.10	89.60
Best method – fine-tuned ResNet-50 [59]	95.30	86.80	91.10

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2020.2971258, IEEE Transactions on Medical Imaging

TABLE VI
AUC RESULTS ON IMAGECLEF 2017 DATASET (MDR DETECTION) COMPARED TO OTHER METHODS.

Methods	AUC (%)
Pre-trained ResNet-152 (FC) + SVM	53.96
Pre-trained GoogLeNet (FC) + SVM	55.94
Pre-trained AlexNet (FC) + SVM	59.30
Fine-tune AlexNet	59.85
Our method	65.52

TABLE VII
TRAINING TIME OF ZERO-BIAS CAE PER EPOCH ON IMAGECLEF 2016 DATASET (MODALITY CLASSIFICATION) COMPARED TO OTHER METHODS.

Methods	Average training time per epoch (s)
Fine-tuned ResNet-152	6961
Fine-tuned VGG-16	6299
Fine-tuned GoogLeNet	748
Our method	280

be subtly different. Our experiments using higher number of convolutional layers such as 3- to 7-layer architectures did not improve performance (see Fig. 6).

The results of using different augmentation rates (τ) are shown in Fig. 7. Our experiments using augmentation rate of 0.5 had the highest accuracies from all three datasets.

The training time of our zero-bias CAE compared to other fine-tunned CNNs are shown in Table VII. Our zero-bias CAE required less training time than the other fine-tunned CNNs, with an average training time of 280s per epoch.

VI. DISCUSSION

Our main findings are that our zero-bias CAE: a) outperformed other conventional unsupervised feature learning methods; b) had competitive accuracy with the state-of-the-art supervised CNNs for medical image modality, skin disease classification and MDR detection; c) constrained the transformation of generic features from a source (natural images) to non-redundant and locally relevant features for the target domains (medical images); d) improved feature representation of medical images using our context-based feature augmentation.

Our zero-bias CAE outperformed other unsupervised feature learning methods, and we attribute this to our robust combination of a zero-bias CAE and a CNN that was pre-

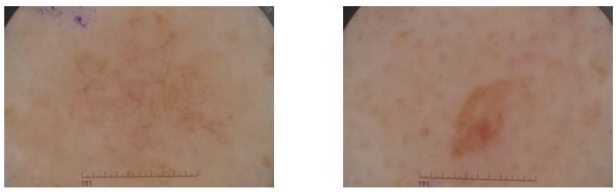

Fig. 4. Sample images of melanoma, where the both left and right skin lesions are correctly classified by our zero-bias CAE but misclassified by the baseline pre-trained AlexNet as benign nevi.

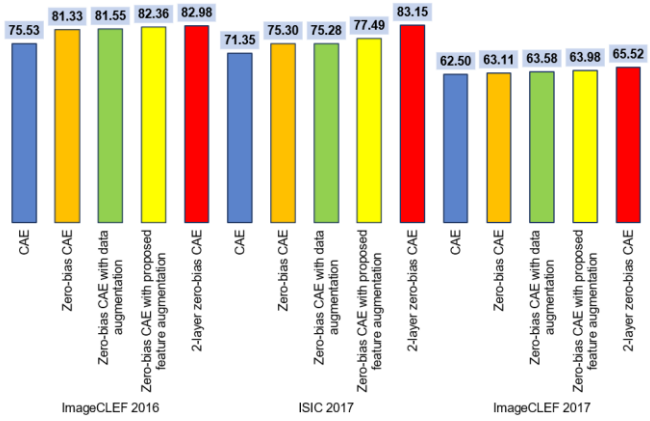

Fig. 5. Top 1 accuracy (%) and AUC (%) of CAE, zero-bias CAE, zero-bias CAE with data augmentation (horizontal flips and translations), our zero-bias CAE with context-based feature augmentation and our 2-layer zero-bias CAE with context-based feature augmentation.

trained from natural image domain. It preserves the generic image features from the pre-trained CNN and encodes specific and relevant local characteristics that lie within the medical images.

Our zero-bias CAE also outperformed all other pre-trained CNNs and achieved competitive accuracies to fine-tunned CNNs. This is attributed to our zero-bias ReLU activation that forces neurons with large negative values of biases to activate (in particular those ones whose inner product with the weight vector is small), thereby removing the possibility of ignoring subtle and sparse characteristics of medical images (see Fig. 5). It acts as a form of regularisation that allows to learn parameters that are more discriminative for medical image data. In addition, our context-based feature augmentation further improves feature representation of medical images based upon multiple sources from multiple similar images (neighbours in the feature space). This is different to standard data augmentation schemes that create a synthetic image from a single source. We suggest that creating image features using the distribution of features in clusters of similar images via our context-based feature augmentation improves the learning outcome. The clusters of similar images allowed us to create a new fuzzy feature space that contains diverse and relevant visual characteristics of a particular (diagnostic) class.

The results from Fig. 6 show that the use of 2-layer zero-bias CAE had the highest classification accuracies across all 3 datasets. However, when the number of layers was > 3, accuracy decreased. While more layers could extract a higher level of feature representation, it increases the number of weights that need to be learned, which could make the model more prone to overfitting. We suggest that a supervised fine-tuning step could be a potential mechanism to achieve better performance as we increase the depth of the network.

The results from Fig. 7 show that the use of higher augmentation rates (0.1 to 0.5) generally improves the feature representation. However, the use of a higher augmentation rate, such as 0.6, can derive image features that are less relevant to the original data, and thus resulting in lower accuracies.

Medical image modality classification – Our approach outperformed all other unsupervised approaches and achieved a competitive accuracy (82.98%) compared to all supervised CNNs reported in the ImageCLEF 2016 competition. Our

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2020.2971258, IEEE Transactions on Medical Imaging

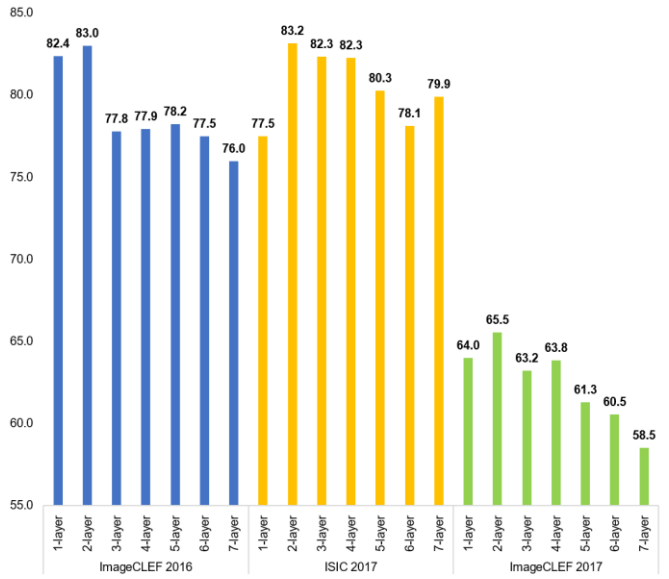

Fig. 6. Top 1 accuracy (%) and AUC (%) using higher number of convolutional layers.

results show that the quality of the features learned with conventional unsupervised feature learning methods such as SC, ICA and SSAE were not as robust as that of our approach. It also had a higher accuracy than the pre-trained CNNs; the image features extracted from these methods were not optimised to contain characteristics of medical images, and as such have limited ability to extract most discriminative features. As expected, the fine-tuning of AlexNet (79.60%) and GoogLeNet (80.75%) produced better results (as in [20] [62]) relative to the baseline pre-trained methods. The best performing method was a synergic deep learning model that used multiple fine-tuned ResNets with an accuracy of 87.97% [39]; we note that this model was trained with an manually expanded dataset from another source. The performances of these supervised methods were dependent on the availability of the labelled dataset.

Skin disease classification – The quality of image features extracted using conventional unsupervised methods such as SC, ICA and SSAE were not as robust as that of our approach, consistent with the results of medical image modality classification. Ensemble of CNNs (e.g., AlexNet and GoogLeNet) and deeper CNNs (e.g., ResNet with 50 or 101) layers) generally had a higher accuracy than shallower CNNs (see Table V). The best method reported in ISIC 2017 challenge used the ResNet with 50 layers and added extra labelled data from additional sources, which contributed to its overall accuracy.

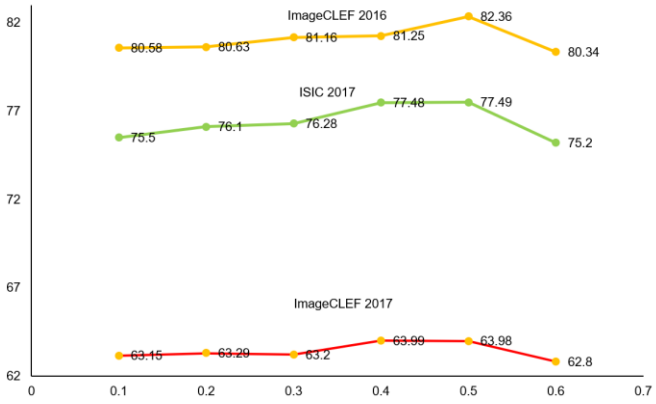

Fig. 7. Top 1 accuracy (%) and AUC (%) using a single-layer zero-bias CAE with different augmentation rates (τ).

MDR detection – Our approach outperformed all pre-trained CNNs and had a higher AUC (65.52%) than the baseline fine-tuned AlexNet (59.85%). Our results show that the image features extracted from shallower pre-trained CNNs such as AlexNet had a higher AUC (59.30%) compared to the deeper CNNs such as GoogLeNet (55.94%) and ResNet (53.96%). This is attributed to the fact that the image features extracted from deeper CNNs represented image characteristics that were more relevant to natural images, i.e., they had less image characteristics that were relevant to describing subtle features in CT images. As expected, the fine-tunning of AlexNet, produced a better result relative to the baseline pre-trained AlexNet, consistent with the results from other datasets. The top performing methods reported in the competition, using the test dataset, also used fine-tuned CNNs such as ResNet [54]. The best method was a non-CNN-based approach using 3D texture features, achieving an AUC of 58.25% [61]. This indicates that the accuracy of fine-tuned CNNs are dependent on the availability of labelled data from medical image domain. In the ImageCLEF 2017 (MDR TB detection) dataset there were 230 patients in the training dataset. As a consequence, the performance of our zero-bias CAE, but also all other baseline CNNs was poorer when compared to the ImageCLEF 2016 (6776 images) and ISIC 2017 (2000 images) datasets. Nevertheless, our result shows that our approach was able to drive domain-specific image features without reliance on labels. As such we suggest that it could be easily scalable to many different transfer learning or domain adaptation approaches where they require some form of labelled data from either source or target domain.

A. Limitations and future work

Although our approach demonstrated improvements in the feature representation of medical images in an unsupervised manner, it is fundamentally dependent on the quality of image features provided from the pre-trained CNNs (in our experiments, AlexNet). Moreover, medical image features extracted from the pre-trained CNNs should contain sufficient specific local representations. It is possible that deeper pre-trained CNNs such as GoogLeNet and ResNet as base models (source domains), may potentially provide different feature representations, some of which could potentially be more discriminative for different target domains. Investigation of identifying optimal layers to extract image features from the CNNs is also important.

Although our context-based feature augmentation scheme improves the learning outcomes, the parameter k, i.e., number of clusters needs to be derived by observing the distributions of the image samples in clusters. A hierarchical clustering that do not require any predefined parameters could potentially provide more meaningful feature representation and we will explore such approaches in the future.

One of the key advantages of using unsupervised feature learning methods is the ability to use all available unlabelled medical image data. When our approach is used with the all

available unlabelled data, it could potentially enable the derivation of semantically more meaningful representations of image data. We note that our zero-bias CAE needs to learn fewer parameters across fewer layers when compared to other supervised CNNs, and so it can be efficiently coupled with supervised fine-tuning approaches without a large computational overhead. The investigation of how this would impact different clinical scenarios is a substantial research study and we will pursue this as part of our future work.

VII. Conclusion

We developed a deep unsupervised feature extractor to better characterise medical image data. Our 2-layer zero-bias CAE and context-based feature augmentation improved learning outcomes and feature representation in medical image classification. We compared our approach to other unsupervised and supervised methods on 3 large public datasets and showed that our approach was competitive with the state-of-the-art supervised CNNs. Our findings indicate that our approach can be generally applied across different medical image data and we suggest that it can benefit many medical image analysis tasks when there are none or limited annotated training data.

References

[1] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[2] 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.

[3] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1-9.

[4] O. Heri, A. Hiba, and A. Zarandy, "Automatic skin lesion analysis using relatively small learning set," in *CNNA 2018; The 16th International Workshop on Cellular Nanoscale Networks and their Applications*, 2018, pp. 1-4: VDE.

[5] H.-C. Shin, M. R. Orton, D. J. Collins, S. J. Doran, and M. O. Leach, "Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4D patient data," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 8, pp. 1930-1943, 2013.

[6] E. Ahn, A. Kumar, J. Kim, C. Li, D. Feng, and M. Fulham, "X-ray image classification using domain transferred convolutional neural networks and local sparse spatial pyramid," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, 2016, pp. 855-858.

[7] Y. Bengio, A. Bergeron, N. Boulanger–Lewandowski, T. Breuel, Y. Chherawala, M. Cisse *et al.*, "Deep learners benefit more from out-of-distribution examples," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, 2011, pp. 164-172.

[8] T. DeVries and G. W. Taylor, "Dataset augmentation in feature space," *arXiv preprint arXiv:1702.05538*, 2017.

[9] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette *et al.*, "Domain-adversarial training of neural networks," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 2096-2030, 2016.

[10] K. Bousmalis, G. Trigeorgis, N. Silberman, D. Krishnan, and D. Erhan, "Domain separation networks," in *Advances in Neural Information Processing Systems*, 2016, pp. 343-351.

[11] P. O. Pinheiro and A. Element, "Unsupervised domain adaptation with similarity learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8004-8013.

[12] M. Ghifary, W. B. Kleijn, M. Zhang, D. Balduzzi, and W. Li, "Deep reconstruction-classification networks for unsupervised domain adaptation," in *European Conference on Computer Vision*, 2016, pp. 597-613: Springer.

[13] H. T. H. Phan, A. Kumar, J. Kim, and D. Feng, "Transfer learning of a convolutional neural network for HEp-2 cell image classification," in *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*, 2016, pp. 1208-1211: IEEE.

[14] A. Kumar, P. Sridar, A. Quinton, R. K. Kumar, D. Feng, R. Nanan *et al.*, "Plane identification in fetal ultrasound images using saliency maps and convolutional neural networks," in *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*, 2016, pp. 791-794: IEEE.

[15] F. Ciompi, B. de Hoop, S. J. van Riel, K. Chung, E. T. Scholten, M. Oudkerk *et al.*, "Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box," *Medical image analysis*, vol. 26, no. 1, pp. 195-202, 2015.

[16] S. Choi, "X-ray image body part clustering using deep convolutional neural network: SNUMedinfo at ImageCLEF 2015 medical clustering task," in *ImageCLEF 2015*, 2015.

[17] Y. Anavi, I. Kogan, E. Gelbart, O. Geva, and H. Greenspan, "A Comparative Study for Chest Radiograph Image Retrieval using Binary, Texture, and Deep Learning Classification," in *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015.

[18] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues *et al.*, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285-1298, 2016.

[19] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway *et al.*, "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.

[20] A. Kumar, J. Kim, D. Lyndon, M. Fulham, and D. Feng, "An Ensemble of Fine-Tuned Convolutional Neural Networks for Medical Image Classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 31-40, 2017.

[21] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60-88, 2017/12/01/ 2017.

[22] R. Bermúdez-Chacón, C. Becker, M. Salzmann, and P. Fua, "Scalable unsupervised domain adaptation for electron microscopy," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2016, pp. 326-334: Springer.

[23] T. Heimann, P. Mountney, M. John, and R. Ionasec, "Learning without labeling: Domain adaptation for ultrasound transducer localization," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2013, pp. 49-56: Springer.

[24] K. Kamnitsas, C. Baumgartner, C. Ledig, V. Newcombe, J. Simpson, A. Kane *et al.*, "Unsupervised domain adaptation in brain lesion segmentation with adversarial networks," in *International Conference on Information Processing in Medical Imaging*, 2017, pp. 597-609: Springer.

[25] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," *arXiv preprint arXiv:1405.3531*, 2014.

[26] L. Bi, J. Kim, E. Ahn, A. Kumar, D. Feng, and M. Fulham, "Step-wise integration of deep class-specific learning for dermoscopic image segmentation," *Pattern Recognition*, vol. 85, pp. 78-89, 2019.

[27] H. Lee, A. Battle, R. Raina, and A. Y. Ng, "Efficient sparse coding algorithms," in *Advances in neural information processing systems*, 2006, pp. 801-808.

[28] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural computation*, vol. 18, no. 7, pp. 1527-1554, 2006.

[29] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807-814.

[30] E. Ahn, A. Kumar, M. Fulham, D. Feng, and J. Kim, "Convolutional Sparse Kernel Network for Unsupervised Medical Image Analysis," *Medical Image Analysis*, vol. 56, pp. 140-151, 2019/06/12/ 2019.

[31] J. Mairal, P. Koniusz, Z. Harchaoui, and C. Schmid, "Convolutional kernel networks," in *Advances in Neural Information Processing Systems*, 2014, pp. 2627-2635.

[32] J. Mairal, "End-to-end kernel learning with supervised convolutional kernel networks," in *Advances in neural information processing systems*, 2016, pp. 1399-1407.

[33] E. Ahn, A. Kumar, D. Feng, M. Fulham, and J. Kim, "Unsupervised Feature Learning with K-means and An Ensemble of Deep

Convolutional Neural Networks for Medical Image Classification," *arXiv preprint arXiv:1906.03359*, 2019.

[34] M. Gong, J. Liu, H. Li, Q. Cai, and L. Su, "A multiobjective sparse feature learning model for deep neural networks," *IEEE transactions on neural networks and learning systems*, vol. 26, no. 12, pp. 3263-3277, 2015.

[35] E. Hosseini-Asl, J. M. Zurada, and O. Nasraoui, "Deep learning of part-based representation of data using sparse autoencoders with nonnegativity constraints," *IEEE transactions on neural networks and learning systems*, vol. 27, no. 12, pp. 2486-2498, 2016.

[36] E. Ahn, A. Kumar, D. Feng, M. Fulham, and J. Kim, "Unsupervised Deep Transfer Feature Learning for Medical Image Classification," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 2019, pp. 1915-1918.

[37] M. Villegas, H. Müller, A. G. S. de Herrera, R. Schaer, S. Bromuri, A. Gilbert *et al.*, "General Overview of ImageCLEF at the CLEF 2016 Labs," in *International Conference of the Cross-Language Evaluation Forum for European Languages*, 2016, pp. 267-285: Springer.

[38] A. García Seco de Herrera, R. Schaer, S. Bromuri, and H. Müller, "Overview of the medical tasks in ImageCLEF 2016," in *Proceedings of the 7th International Conference of CLEF Association (CLEF) 2016*, 2016, no. CONFERENCE: 5-8 September 2016.

[39] J. Zhang, Y. Xie, Q. Wu, and Y. Xia, "Medical image classification using synergic deep learning," *Medical image analysis*, vol. 54, pp. 10-19, 2019.

[40] A. Kumar, J. Kim, W. Cai, M. Fulham, and D. Feng, "Content-Based Medical Image Retrieval: A Survey of Applications to Multidimensional and Multimodality Data," *Journal of Digital Imaging*, journal article vol. 26, no. 6, pp. 1025-1039, December 01 2013.

[41] J. Kalpathy-Cramer and W. Hersh, "Multimodal medical image retrieval: image categorization to improve search precision," in *Proceedings of the international conference on Multimedia information retrieval*, 2010, pp. 165-174: ACM.

[42] H. Müller, J. Kalpathy–Cramer, I. Eggel, S. Bedrick, S. Radhouani, B. Bakke *et al.*, "Overview of the CLEF 2009 medical image retrieval track," in *Workshop of the Cross-Language Evaluation Forum for European Languages*, 2009, pp. 72-84: Springer.

[43] N. C. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza *et al.*, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)," in *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on*, 2018, pp. 168-172: IEEE.

[44] D. T. Hoagland, J. Liu, R. B. Lee, and R. E. Lee, "New agents for the treatment of drug-resistant Mycobacterium tuberculosis," *Advanced drug delivery reviews*, vol. 102, pp. 55-72, 2016.

[45] Y. D. Cid, A. Kalinovsky, V. Liauchuk, V. Kovalev, and H. Müller, "Overview of the ImageCLEF 2017 Tuberculosis Task-Predicting Tuberculosis Type and Drug Resistances," in *CLEF (Working Notes)*, 2017.

[46] K. Konda, R. Memisevic, and D. Krueger, "Zero-bias autoencoders and the benefits of co-adapting features," *arXiv preprint arXiv:1402.3337*, 2014.

[47] A. Coates, A. Ng, and H. Lee, "An analysis of single-layer networks in unsupervised feature learning," in *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, 2011, pp. 215-223.

[48] 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.

[49] A. Coates and A. Y. Ng, "Learning feature representations with k-means," in *Neural networks: Tricks of the trade*: Springer, 2012, pp. 561-580.

[50] A. Hyv, J. Hurri, and P. Hoyer, "Natural Image Sta-tistics: A Probabilistie Approach to Early Computa-tional Vision," ed: Berlin: Springer, 2009.

[51] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," in *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, 2007, pp. 1027-1035: Society for Industrial and Applied Mathematics.

[52] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 1794-1801: IEEE.

[53] Y. D. Cid, O. A. J. Del Toro, A. Depeursinge, and H. Müller, "Efficient and fully automatic segmentation of the lungs in CT volumes," in *VISCERAL Challenge@ ISBI*, 2015, pp. 31-35.

[54] J. Sun, P. Chong, Y. X. M. Tan, and A. Binder, "ImageCLEF 2017: ImageCLEF Tuberculosis Task-the SGEast Submission," in *CLEF (Working Notes)*, 2017.

[55] M. Engelcke, D. Rao, D. Z. Wang, C. H. Tong, and I. Posner, "Vote3deep: Fast object detection in 3d point clouds using efficient convolutional neural networks," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 1355-1361: IEEE.

[56] B. Graham and L. van der Maaten, "Submanifold sparse convolutional networks," *arXiv preprint arXiv:1706.01307*, 2017.

[57] R. T. Sousa and L. V. de Moraes, "Araguaia Medical Vision Lab at ISIC 2017 Skin Lesion Classification Challenge," *arXiv preprint arXiv:1703.00856*, 2017.

[58] D. H. Murphree and C. Ngufor, "Transfer Learning for Melanoma Detection: Participation in ISIC 2017 Skin Lesion Classification Challenge," *arXiv preprint arXiv:1703.05235*, 2017.

[59] K. Matsunaga, A. Hamada, A. Minagawa, and H. Koga, "Image classification of melanoma, nevus and seborrheic keratosis by deep neural network ensemble," *arXiv preprint arXiv:1703.03108*, 2017.

[60] L. Bi, J. Kim, E. Ahn, and D. Feng, "Automatic skin lesion analysis using large-scale dermoscopy images and deep residual networks," *arXiv preprint arXiv:1703.04197*, 2017.

[61] Y. D. Cid, K. Batmanghelich, and H. Müller, "Textured Graph-model of the Lungs for Tuberculosis Type Classification and Drug Resistance Prediction: Participation in ImageCLEF 2017," in *CLEF (Working Notes)*, 2017.

[62] S. Koitka and C. M. Friedrich, "Optimized Convolutional Neural Network Ensembles for Medical Subfigure Classification," in *International Conference of the Cross-Language Evaluation Forum for European Languages*, 2017, pp. 57-68: Springer.