

Lung and Pancreatic Tumor Characterization in the Deep Learning Era: Novel Supervised and Unsupervised Learning Approaches

Sarfraz Hussein, Pujan Kandel, Candice W. Bolan, Michael B. Wallace, and Ulas Bagci*, *Senior Member, IEEE.*

Abstract—Risk stratification (characterization) of tumors from radiology images can be more accurate and faster with computer-aided diagnosis (CAD) tools. Tumor characterization through such tools can also enable non-invasive cancer staging, prognosis, and foster personalized treatment planning as a part of precision medicine. In this study, we propose both supervised and unsupervised machine learning strategies to improve tumor characterization. Our *first approach* is based on supervised learning for which we demonstrate significant gains with deep learning algorithms, particularly by utilizing a 3D Convolutional Neural Network and Transfer Learning. Motivated by the radiologists’ interpretations of the scans, we then show how to incorporate task dependent feature representations into a CAD system via a graph-regularized sparse Multi-Task Learning (MTL) framework.

In the *second approach*, we explore an unsupervised learning algorithm to address the limited availability of labeled training data, a common problem in medical imaging applications. Inspired by learning from label proportion (LLP) approaches in computer vision, we propose to use proportion-SVM for characterizing tumors. We also seek the answer to the fundamental question about the goodness of “deep features” for unsupervised tumor classification. We evaluate our proposed supervised and unsupervised learning algorithms on two different tumor diagnosis challenges: lung and pancreas with 1018 CT and 171 MRI scans, respectively, and obtain the state-of-the-art sensitivity and specificity results in both problems.

Index Terms—Unsupervised Learning, Lung cancer, 3D CNN, IPMN, Pancreatic cancer.

I. INTRODUCTION

Approximately 40% of people will be diagnosed with cancer at some point during their lifetime with an overall mortality of 171.2 per 100,000 people per year (based on deaths between 2008-2012) [1]. Lung and pancreatic cancers are two of the most common cancers. While lung cancer is the largest cause of cancer-related deaths in the world, pancreatic cancer has the poorest prognosis with a 5-year survival rate of only 7% in the United States [1]. With regards to pancreatic cancer, specifically in this work, we focus on the challenging problem of automatic diagnosis of Intraductal Papillary Mucinous Neoplasms (IPMN). IPMN is a pre-malignant condition and if left untreated, it can progress to invasive cancer. IPMN is mucin-producing neoplasm that can be found in the main pancreatic duct and its branches. They are radiographically

* indicates corresponding author (ulasbagci@gmail.com).

S. Hussein is with Center for Advanced Machine Learning (CAML) at Symantec Corporation. U. Bagci is with Center for Research in Computer Vision (CRCV) at University of Central Florida (UCF), Orlando, FL; P. Kandel, C. Bolan, and M. Wallace are with Mayo Clinic, Jacksonville, FL

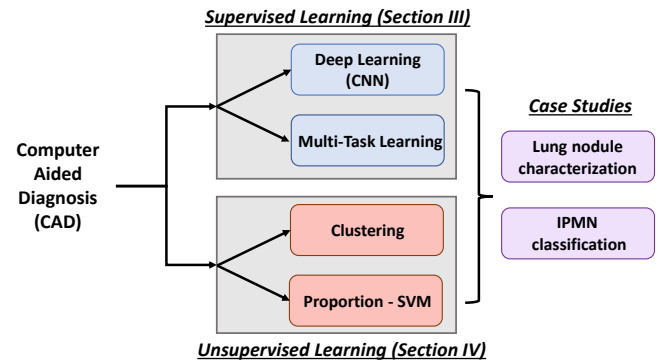


Figure 1: A block diagram to represent different schemes, methods and experimental case studies presented in this paper. We develop both supervised and unsupervised learning algorithms to characterize tumors. For the supervised learning scheme, we propose a new 3D CNN architecture based on a Graph Regularized Sparse Multi-task learning and perform evaluations for lung nodule characterization from CT scans. For unsupervised learning scheme, we propose a new clustering algorithm, \propto SVM, and test it for the categorization of lung nodules from CT scans and pancreatic cysts (IPMN) from MRI scans.

identifiable precursors to pancreatic cancer [2]. Detection and characterization of these lung and pancreatic tumors can aid in early diagnosis; hence, increased survival chance through appropriate treatment/surgery plans.

Conventionally, the CAD systems are designed to assist radiologists in making accurate and fast decisions by reducing the number of false positives and false negatives. For diagnostic decision making, a higher emphasis is laid on increased sensitivity: a false-flag is more tolerable than a tumor being missed or incorrectly classified as benign. In this regard, a computerized analysis of imaging features becomes a key instrument for radiologists to improve their diagnostic decisions. In the literature, automated detection and diagnosis methods had been developed for tumors in different organs such as breast, colon, brain, lung, liver, prostate, and others. As typical in such studies, a CAD includes preprocessing and feature engineering steps (including feature extraction and selection) followed by a classification step [3], [4], [5], [6]. However, with the success of deep learning, a transition from feature engineering to feature learning has been observed in

medical image analysis literature. Those systems comprise Convolutional Neural Networks (CNN) as feature extractor followed by a conventional classifier such as Random Forest (RF) [7], [8]. In scenarios where a large number of labeled training examples are available, however, end-to-end trainable deep learning approaches can be employed [9].

This paper includes two main approaches for tumor characterization from radiology scans: supervised and unsupervised learning algorithms. In the first part, we focus on novel supervised algorithms, which is a significant extension to our IPMI 2017 study [10]. Specifically, we first present a novel supervised learning strategy to perform risk-stratification of lung nodules from low-dose CT scans. For this strategy, we perform a 3D CNN based discriminative feature extraction from radiology scans. We contend that 3D networks are important for the characterization of lung nodules in CT images which are inherently 3-dimensional. The use of conventional 2D CNN methods, whereas, leads to the loss of vital volumetric information which can be crucial for precise risk assessment of lung nodules. In the absence of a large number of labeled training examples, we utilize a pre-trained 3D CNN architecture and fine-tune the network with lung nodules dataset. Also, inspired by the significance of lung nodule attributes for clinical determination of malignancy [11], we utilize the information about six high-level nodule attributes such as calcification, spiculation, sphericity, lobulation, margin, and texture (Figure 2-A) to improve automatic benign-malignant classification. Then, we integrate these high-level features into a novel graph regularized multi-task learning (MTL) framework to yield the final malignancy output. We analyze the impact of the aforementioned lung nodule attributes in-depth for malignancy determination and find these attributes to be complementary when obtaining the malignancy scores. From a technical perspective, we also exploit different regularizers and multi-task learning approaches such as trace-norm and graph regularized MTL for regression.

In the second part of the paper, inspired by the successful application of unsupervised learning methods in other domains, we explore the potential of unsupervised learning strategies in lung nodule and IPMN classification. First, we extract discriminative information from a large amount of unlabeled imaging data. We analyze both hand-crafted and deep learning features and assess how good those features are when applied to tumor characterization. In order to obtain an initial set of labels in an unsupervised fashion, we cluster the samples into different groups in the feature domain. We next propose to train Proportion-Support Vector Machine (\propto SVM) algorithm using label proportions rather than instance labels. The trained model is then employed to learn malignant-benign categorization of the tumors.

This paper is organized as follows. Section 2 describes related work pertaining to supervised and unsupervised learning for the diagnosis of lung nodules and IPMN. We present our MTL based supervised learning algorithm in Section 3. In Section 4, we introduce an unsupervised learning method adapted for the diagnosis of lung nodules and IPMN from CT and MRI scans, respectively. The experiments and results are discussed in Section 5. In the same section, we also

study the influence of different deep learning features for unsupervised learning and establish an upper bound on the classification performance using supervised learning methods. Finally, Section 6 states discussions and concluding remarks.

II. RELATED WORK

This section summarizes the advances in machine learning applied to medical imaging and CAD systems developed for lung cancer diagnosis. Since the automatic characterization of IPMN from MRI scans has not been extensively studied in the literature, relevant works are mostly selected from the clinical studies. Our work is the first in this regard.

Imaging Features and Classifiers: Conventionally, the risk stratification (classification) of lung nodules may require nodule segmentation, computation and selection of low-level features from the image, and the use of a classifier/regressor. In the approach by [12], different physical statistics including intensity measures were extracted and class labels were obtained using Artificial Neural Networks. In [3], lung nodules were segmented using appearance-based models followed by shape analysis using spherical harmonics. The last step involved k -nearest neighbor based classification. Another approach extended 2D texture features including Local Binary Patterns, Gabor and Haralick to 3D [4]. Classification using Support Vector Machine (SVM) was performed as the final step. In a different study, Way et al. [5], implemented nodule segmentation via 3D active contours, and then applied rubber band straightening transform. A Linear Discriminant Analysis (LDA) classifier was applied to get class labels. Lee et al. [6] introduced a feature selection based approach utilizing both clinical and imaging data. Information content and feature relevance were measured using an ensemble of genetic algorithm and random subspace method. Lastly, LDA was applied to obtain final classification on the condensed feature set. In a recent work, spherical harmonics features were fused with deep learning features [8] and then RF classification was employed for lung nodule characterization. Hitherto, the application of CNN for nodule characterization has been limited to 2D space [13], thus falling short of incorporating vital contextual and volumetric information. In another approach, Shin et al. [14] employed CNN for the classification of lung nodules. Other than not completely 3D CNN, the approach didn't take into account high-level nodule attributes and required training an off-the-shelf classifier such as RF and SVM.

The information about different high-level image attributes had been found useful in the malignancy characterization of lung nodules. In a study exploring the correlation between malignancy and nodule attributes, [11] found that 82% of the lobulated, 93% of the ragged, 97% of the densely spiculated, and 100% of the halo nodules were malignant in a particular dataset. Automatic determination of lung nodule attributes and types had been explored in [15]. The objective was to perform the classification of six different nodule types such as solid, non-solid, part-solid, calcified, periferfissural and spiculated nodules. However, the approach is based on 2D CNN and fell short

of estimating the malignancy of lung nodules. Furthermore, 66% of the round nodules were determined as benign.

Although, not an objective in this paper, the detection of lung nodules has also been an active subject of interest among researchers [16], [17], [18]. A short but informative review of the most recent detection studies can be found in [18].

Pancreatic Cysts (IPMN): Although there has been considerable progress in developing automatic approaches to segment pancreas and its cysts [19], [20], the use of advanced machine learning algorithms to perform fully automatic risk-stratification of IPMNs is limited. The approach by Hanania et al. [21] investigated the influence of 360 imaging features ranging from intensity, texture, and shape to stratify subjects as low or high-grade IPMN. In another example, Gazit et al. [22] extracted texture and features from the solid component of segmented cysts followed by a feature selection and classification scheme. Both of these approaches [21], [22] required segmentation of cysts or pancreas and are evaluated on CT scans only.

Unsupervised Learning: Typically, the visual recognition and classification tasks are addressed using labeled data (supervision). However, for tasks where manually generating labels corresponding to large datasets is laborious and expensive, the use of unsupervised learning methods is of significant value. Unsupervised techniques had been used to solve problems in various domains ranging from object categorization [23], speech processing [24] and audio classification [25]. These methods conventionally relied on some complementary information provided with the data to improve learning, which may not be available for several classification tasks in medical imaging.

In medical imaging, there have been different approaches that used unsupervised learning for detection and diagnosis problems. The approach by Shin et al. [26] used stacked autoencoders for multiple organ detection in MRI scans. Vaidhya et al. [27] presented a brain tumor segmentation method with stacked denoising autoencoder evaluated on multi-sequence MRI images. In a work by Sivakumar et al. [28], the segmentation of lung nodules is performed with unsupervised clustering methods. In another study, Kumar et al. [7] used features from autoencoder for lung nodule classification. These auto-encoder approaches, however, did not yield satisfactory classification results. Other than these, unsupervised deep learning has also been explored for mammographic risk prediction and breast density segmentation [29].

Unsupervised feature learning remains an active research area for the medical imaging community, more recently with Generative Adversarial Networks (GAN) [30]. In order to explore the information from unlabeled images, Zhang et al. [31] described a semi-supervised method for the classification of four types of nodules. In sharp contrast to these approaches, the unsupervised learning strategies presented in this paper don't involve feature learning using auto-encoders. Using sets of hand-crafted as well as pre-trained deep learning features, we propose a new unsupervised learning algorithm where an initially estimated label set is

progressively improved via proportion-SVM.

Our Contributions

A block diagram representing different supervised and unsupervised schemes is presented in Figure 1. Overall, our main contributions in this work can be summarized as follows:

- For lung nodule characterization, we present a 3D CNN based supervised learning approach to fully appreciate the anatomical information in 3D, which would be otherwise lost in the conventional 2D approaches. We use fine-tuning strategy to avoid the requirement for a large number of volumetric training examples for 3D CNN. For this purpose, we use a pre-trained network (which is trained on 1 million videos) and fine-tune it on the CT data.
- We introduce a graph regularized sparse MTL platform to integrate the complementary features from lung nodule attributes so as to improve malignancy prediction. Figure 2-A shows high-level lung nodule attributes having varying levels of prominence.
- We evaluate the proposed supervised and unsupervised learning algorithms to determine the characterization of lung nodules and IPMN cysts (Table I). In the era where the wave of deep learning has swept into almost all domains of visual analysis, we investigate the contribution of features extracted from different deep learning architectures. To the best of our knowledge, this is the first work to investigate the automatic diagnosis of IPMNs from MRI.
- In the proposed unsupervised learning algorithm, instead of hard assigning labels, we estimate the label proportions in a data-driven manner. Additionally, to alleviate the effect of noisy labels (i.e. mislabels) obtained during clustering, we propose to employ \propto SVM, which is trained on label proportions only.

III. SUPERVISED LEARNING METHODS

A. Problem Formulation

Let $X = [x_1, x_2 \dots x_n]^T \in \mathbb{R}^{n \times d}$ represent the input features obtained from n images of lung nodules each having a dimension d . Each data sample has an attribute/malignancy score given by $Y = [y_1, y_2 \dots y_n]$, where $Y^T \in \mathbb{R}^{n \times 1}$. While X consists of features extracted from radiology images, and Y represents the malignancy score over 1-5 scale where 1 represents benign and 5 represents malignant. In supervised learning, the labeled training data is used to learn the coefficient vector or the regression estimator $W \in \mathbb{R}^d$. In testing, W is used to estimate Y for an unseen feature/example. For regression, a regularizer is often added to prevent overfitting. Hence, a classical least square regression turns into a *constrained optimization* problem with ℓ_1 regularization as:

$$\min_W \|XW - Y\|_2^2, \text{ s.t. } \|W\|_1 \leq t. \quad (1)$$

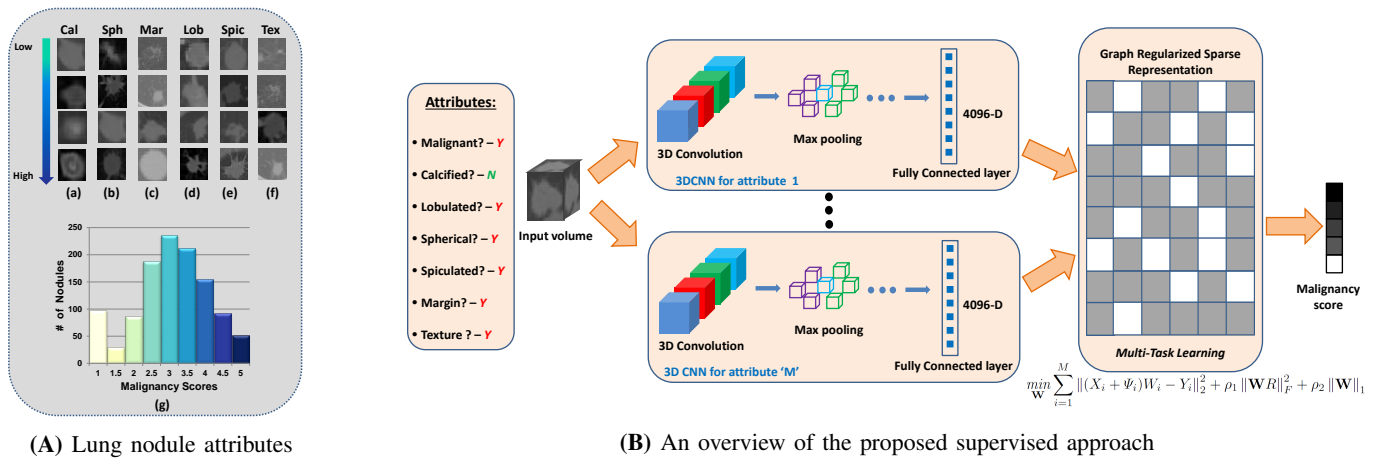


Figure 2: (A) A visualization of lung nodules having different levels of attributes. On moving from the top (attribute absent) to the bottom (attribute prominently visible), the prominence level of the attribute increase. Different attributes including calcification, sphericity, margin, lobulation, spiculation and texture can be seen in (a-f). The graph in (g) depicts the number of nodules with different malignancy levels in our experiments using the publicly available dataset [32]. An overview of the proposed 3D CNN based graph regularized sparse MTL approach is presented in (B).

In the above equation, the sparsity level of the coefficient vector $W = [w_1, w_2 \dots w_d]$ is controlled by a parameter t . Since the function in Eq. (1) is convex and the constraints define a convex set, a local minimizer of the objective function is subjected to constraints corresponding to a global minimizer. In the following subsections, we extend this supervised learning setting with deep learning and MTL concepts to characterize lung nodules as benign or malignant.

B. 3D Convolution Neural Network (CNN) and Fine-Tuning

We use 3D CNN [33] trained on Sports-1M dataset [34] and fine-tune it on the lung nodule CT dataset. The Sports-1M dataset consists of 487 classes with 1 million videos. As the lung nodule dataset doesn't have a large number of training examples, fine-tuning is done to acquire dense feature representation from the Sports-1M. The 3D CNN architecture consists of 5 sets of convolution, 2 fully-connected and 1 softmax classification layers. Each convolution set is followed by a max-pooling layer. The input to the 3D CNN comprises dimensions of $128 \times 171 \times 16$, where 16 denotes the number of slices. Note that the images in the dataset are resized to have consistent dimensions such that the number of channels is 3 and the number of slices is fixed to 16. Hence, the overall input dimension can be considered as $3 \times 16 \times 128 \times 171$. The number of filters in the first 3 convolution layers are 64, 128 and 256 respectively, whereas there are 512 filters in the last 2 layers. The fully-connected layers have a dimension 4096 which is also the length of feature vectors used as an input to the MTL framework. Implementation details are mentioned in section V-C.

C. Multi-task learning (MTL)

Multi-task learning is an approach of learning multiple tasks simultaneously while considering disparities and similarities across those tasks. Given M tasks, the goal is to improve

the learning of a model for task i , ($i \in M$) by using the information contained in the M tasks. We formulate the malignancy prediction of lung nodules as an MTL problem, where visual attributes of lung nodules are considered as distinct tasks (Figure 2A). In a typical MTL problem, initially, the correlation between M tasks and the shared feature representations are not known. The aim in the MTL approach is to learn a joint model while exploiting the dependencies among visual attributes (tasks) in feature space. In other words, we utilize visual attributes and exploit their feature level dependencies so as to improve regressing malignancy using other attributes.

As shown in Figure 2B, we design lung tumor characterization as an MTL problem, where each task has model parameters W_i , which are utilized to characterize the corresponding task i . When $\mathbf{W} = [W_1, W_2 \dots W_M] \in \mathbb{R}^{d \times M}$ constitutes a rectangular matrix, rank can be considered as a natural extension to cardinality, and nuclear/trace norm leads to low rank solutions. In some cases nuclear norm regularization can be considered as the ℓ_1 -norm of the singular values [35]. Trace norm, the sum of singular values, is the convex envelope of the rank of a matrix (which is non-convex), where the matrices are considered on a unit ball. After substituting, ℓ_1 -norm by trace norm, the least square loss function with trace norm regularization can be formulated as:

$$\min_{\mathbf{W}} \sum_{i=1}^M \|\mathbf{X}_i W_i - \mathbf{Y}_i\|_2^2 + \rho \|\mathbf{W}\|_*, \quad (2)$$

where ρ adjusts the rank of the matrix \mathbf{W} , and $\|\mathbf{W}\|_* = \sum_{i=1} \sigma_i(\mathbf{W})$ is the trace-norm where σ denotes singular values. However, as in trace-norm, the assumption about models sharing a common subspace is restrictive for some applications.

As the task relationships are often unknown and are learned from data, we represent tasks and their relations in the form

of a graph. Let $\Upsilon = (V, E)$ represent a complete graph in which nodes V correspond to the tasks and the edges E model any affinity between the tasks. In such case, a regularization can be applied on the graph modeling task dependencies [36]. The complete graph can be modeled as a structure matrix $S = [e^1, e^2 \dots e^{\|E\|}] \in \mathbb{R}^{M \times \|E\|}$ where the deviation between the pairs of tasks can be regularized as:

$$\|\mathbf{W}S\|_F^2 = \sum_{i=1}^{\|E\|} \|\mathbf{W}e^i\|_2^2 = \sum_{i=1}^{\|E\|} \|\mathbf{W}_{e_a^i} - \mathbf{W}_{e_b^i}\|_2^2, \quad (3)$$

here, e_a^i, e_b^i are the edges between the nodes a and b , and $e^i \in \mathbb{R}^M$. The matrix S defines an incidence matrix where e_a^i and e_b^i are assigned to 1 and -1, respectively, if nodes a and b are connected in the graph. Eq. (3) can be further explained as:

$$\|\mathbf{W}S\|_F^2 = \text{tr}((\mathbf{W}S)^T(\mathbf{W}S)) = \text{tr}(\mathbf{W}SS^T\mathbf{W}^T) = \text{tr}(\mathbf{W}\mathcal{L}\mathbf{W}^T) \quad (4)$$

where $\mathcal{L} = SS^T$ is the Laplacian matrix and ‘tr’ represents the trace of a matrix. The method to compute structure matrix S is discussed in Section V-C.

The malignancy prediction equation can be further regularized because there are still other uncertainties to consider, i.e., disagreement between radiologists’ visual interpretations for a given nodule. For instance, while one radiologist may give a malignancy score of x_1^j for a nodule j , the other may give a score of x_2^j for the same nodule. In order to reflect these uncertainties in our algorithm, we formulate a scoring function which models such inconsistencies:

$$\Psi(j) = \left(\exp\left(-\frac{\sum_r (x_r^j - \mu^j)^2}{2\sigma^j}\right) \right)^{-1}. \quad (5)$$

For a particular example j , this inconsistency measure can be represented as $\Psi(j)$. x_r^j is the score given by the r^{th} radiologist (expert) whereas μ^j and σ^j represent mean and standard deviation of the scores, respectively. We calculate this inconsistency score for all the tasks under consideration and for simplicity we have omitted the index for the tasks. The final objective function of graph regularized sparse least square optimization with the inconsistency measure can be expressed as:

$$\min_{\mathbf{W}} \sum_{i=1}^M \underbrace{\|(\mathbf{X}_i + \Psi_i)\mathbf{W}_i - \mathbf{Y}_i\|_2^2}_{\textcircled{1}} + \underbrace{\rho_1 \|\mathbf{W}S\|_F^2}_{\textcircled{2}} + \underbrace{\rho_2 \|\mathbf{W}\|_1}_{\textcircled{3}}, \quad (6)$$

where ρ_1 tunes the penalty degree for graph structure and ρ_2 handles the sparsity level. In Eq. (6), the least square loss function ① observes decoupling of tasks whereas ② and ③ model their interdependencies, so as to learn joint representation.

D. Optimization

In order to solve Eq. (6), the conventional approach is to use standard gradient descent as an optimization algorithm. However, standard gradient descent cannot be applied here

because the ℓ_1 -norm is not differentiable at $\mathbf{W} = 0$ and gradient descent approach fails to provide sparse solutions [37]. Since the optimization function in the above equation has both smooth and non-smooth convex parts, it can be solved after replacing the non-smooth part with its estimates. In other words, the ℓ_1 -norm in the above equation is the non-smooth part and the proximal operator can be used for its estimation. For this purpose, we utilize *accelerated proximal gradient method* [38], the first order gradient method having a convergence rate of $O(1/m^2)$, where m controls the number of iterations.

IV. UNSUPERVISED LEARNING METHODS

Since annotating medical images is laborious, expensive and time-consuming, in the second part of this paper, we explore the potential of unsupervised learning approaches for tumor characterization problems. As illustrated in Figure 3, our proposed unsupervised framework includes three steps. First, we perform clustering on the appearance features obtained from the images to estimate an initial set of labels. Then, using the obtained initial labels, we compute label proportions corresponding to each cluster. Finally, we use the initial cluster assignments and label proportions to learn the categorization of tumors.

A. Initial Label Estimation

Let $X = [x_1, x_2 \dots x_n]^T \in \mathbb{R}^{n \times d}$ represent the input matrix which contains features from n images such that $x \in \mathbb{R}^d$. We then cluster the data into $2 \leq k < n$ clusters using k -means algorithm. Let A represent $|X| \times k$ assignment matrix which denotes the membership assignment of each sample to a cluster. The optimal clustering would minimize the following objective function:

$$\begin{aligned} & \underset{\mu_v, A}{\text{argmin}} \sum_{v=1}^k A(u, v) \|x_u - \mu_v\|^2, \\ & \text{s.t. } A(u, v) = 0 \vee 1, \sum_v A(u, v) = 1 \end{aligned} \quad (7)$$

where μ_v is the mean of the samples in cluster v . The assignment matrix A can then be used to estimate labels c . These labels are only used for estimating label proportions of the clustered data for the purpose of training a new algorithm which we adapt for our problem, i.e., proportion-SVM (\propto SVM). The rationale behind this proportion comes from the clustering notion where data is divided into groups/clusters and each cluster corresponds to a particular class. In our work, specifically, clustering is only an initial step to estimate cluster assignments that are progressively refined in the subsequent steps.

B. Learning with the Estimated Labels

Since our initial label estimation approach is unsupervised, there are uncertainties associated with them. It is, therefore, reasonable to assume that learning a discriminative model based on these noisy instance level labels can deteriorate classification performance. In order to address this issue, we

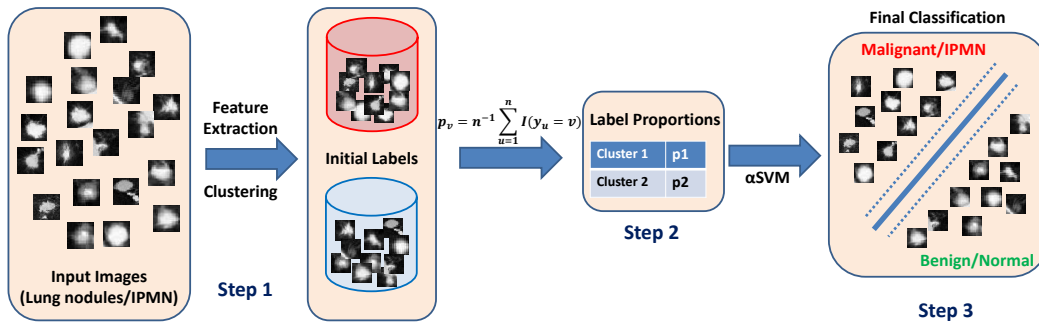


Figure 3: An outline of the proposed unsupervised approach. Given the input images, we compute GIST features and perform k -means clustering to get the initial set of labels which can be noisy. Using the set of labels, we compute label proportions corresponding to each cluster/group (Eq. (9)). We finally employ α SVM to learn a discriminative model using the features and label proportions.

TABLE I: List and details of different experiments performed for supervised and unsupervised learning along with their evaluation sets.

Experiments	Details	Evaluation Set
E1	Supervised learning, 3D CNN based Multi-task learning with attributes, fine-tuning (C3D) network	3D dataset: Malignancy score regression of Lung nodules (CT)
E2	Unsupervised learning, GIST features, Proportion-SVM	2D dataset: Lung nodules (CT) and IPMN classification (MRI)
E3	Unsupervised learning, features from different layers of 2D VGG network	
E4	Supervised learning to establish classification upper-bound, GIST and VGG features with SVM and RF	

model the instance level labels as *latent* variables and thereby consider group/bag level labels.

Inspired by α SVM approach [39], which models the latent instance level variables using the known group level label proportions, we formulate our learning problem such that clusters are analogous to the groups. In our formulation, each cluster v can be represented as a group such that the majority of samples belong to the class v . Considering the groups to be disjoint such that $\bigcup_{v=1}^k \Omega_v = 1, 2, \dots, n$, and Ω represents groups; the objective function of the large-margin α SVM after convex relaxation can be formulated as:

$$\min_{\mathbf{c} \in \mathcal{C}} \min_w \left(\frac{1}{2} w^T w + K \sum_{u=1}^n L(c_u, w^T \phi(x)) \right) \quad (8)$$

$$\mathcal{C} = \left\{ \mathbf{c} \mid |\tilde{p}_v(\mathbf{c}) - p_v| \leq \epsilon, c_u \in \{-1, 1\}^{\forall v=1} \right\},$$

where \tilde{p} and p represent the estimated and true label proportions, respectively. In Eq. (8), \mathbf{c} is the set of instance level labels, $\phi(\cdot)$ is the input feature, K denotes cost parameter and $L(\cdot)$ represents the hinge-loss function for maximum-margin classifiers such as SVM. An alternative approach based on training a standard SVM classifier with clustering assignments is discussed in Section V-D.

The optimization in Eq. (8) is, in fact, an instance of Multiple Kernel Learning, which can be solved using the cutting plane method where the set of active constraints is

incrementally computed. The goal is to find the most violated constraint, however, the objective function still decreases even by further relaxation and aiming for any violated constraint. Further details about optimization can be studied in [39].

C. Calculating Label Proportions

In the conventional α SVM approach, the label proportions are known a priori. Since our approach is unsupervised, both instance level labels and group label proportions are unknown. Moreover, establishing strong assumptions about the label proportions may affect learning. It is, however, reasonable to assume that a large number of instances in any group carry the same label and there may be a small number of instances which are outliers. The label proportions serve as a soft-label for a bag where a bag can be considered as a super-instance. In order to determine the label proportions in a data-driven manner, we use the estimated labels obtained from clustering. The label proportion p_v corresponding to the group v can be represented as:

$$p_v = n^{-1} \sum_{u=1}^n I(y_u = v), \quad (9)$$

where $I(\cdot)$ is the indicator function which yields 1 when $y_u = v$. The α SVM is trained using the image features and label proportions to classify the testing data. It is important to mention that the ground truth labels (benign/malignant labels) are used only to evaluate the proposed framework and are

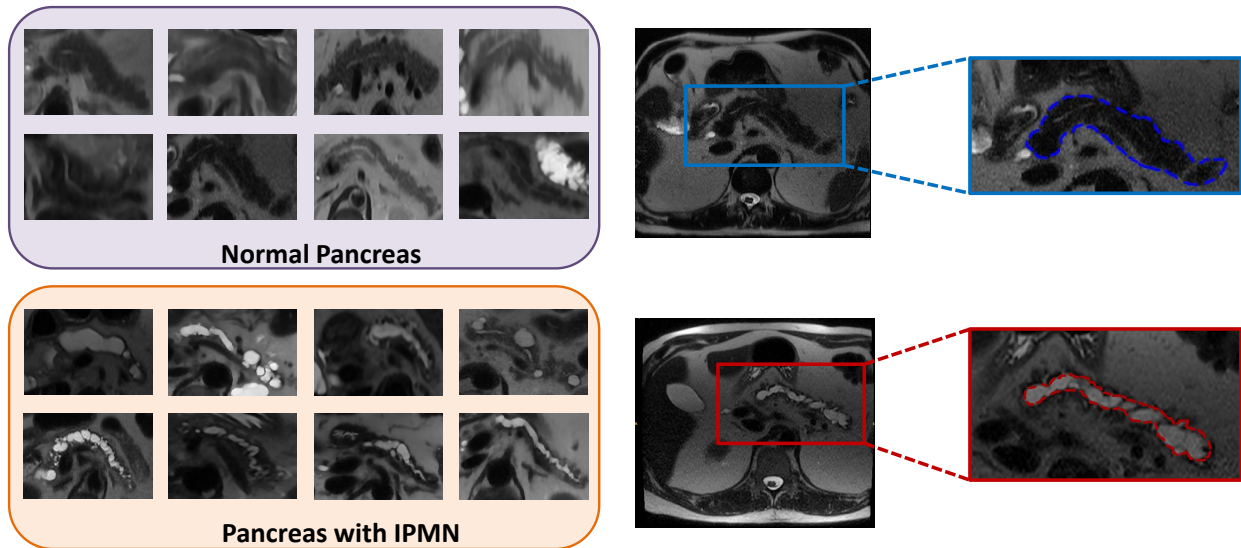


Figure 4: Axial T2 MRI scans illustrating pancreas. The top row shows different ROIs of pancreas, along with a magnified view of a normal pancreas (outlined in blue). The bottom row shows ROIs from subjects with IPMN in the pancreas, which is outlined in red.

not used in estimating label proportions or training of the proportion-SVM. In addition, clustering and label proportion calculation are only performed on the training data and the testing data remains completely unseen for \propto SVM. The number of clusters is fixed as 2, i.e. benign and malignant classes and the result was checked to assign benign and malignant labels to the clusters.

V. EXPERIMENTS

A. Data for Lung Nodules

For test and evaluation, we used LIDC-IDRI dataset from Lung Image Database Consortium [32], which is one of the largest publicly available lung nodule dataset. The dataset comprises 1018 CT scans with a slice thickness varying from 0.45 mm to 5.0 mm. At most four radiologists annotated those lung nodules which have diameters equal to or greater than 3.0 mm.

We considered nodules which were interpreted by at least three radiologists for evaluations. The number of nodules fulfilling this criterion was 1340. As a nodule may have different malignancy and attribute scores provided by different radiologists, their mean scores were used. The nodules have scores corresponding to these six attributes: (i) calcification, (ii) lobulation, (iii) spiculation, (iv) sphericity, (v) margin and (vi) texture as well as malignancy (Figure 2). The malignancy scores ranged from 1 to 5 where 1 denoted benign and 5 meant highly malignant nodules. To account for malignancy indecision among radiologists, we excluded nodules with a mean score of 3. The final evaluation set included 509 malignant and 635 benign nodules. As a pre-processing step, the images were resampled to be isotropic so as to have 0.5 mm spacing in each dimension.

B. Data for IPMN

The data for the classification of IPMN contains T2 MRI axial scans from 171 subjects. The scans were labeled by a radiologist as normal or IPMN. Out of 171 scans, 38 subjects were normal, whereas the rest of 133 were from subjects diagnosed with IPMN. The in-plane spacing (xy-plane) of the scan was ranging from 0.468 mm to 1.406 mm. As pre-processing, we first employ N4 bias field correction [40] to each image in order to normalize variations in image intensity. We then apply curvature anisotropic image filter to smooth image while preserving edges. For experiments, 2D axial slices with pancreas (and IPMN) are cropped to generate Region of Interest (ROI) as shown in Figure 4. The large intra-class variation, especially due to varying shapes of the pancreas can also be observed in Figure 4. A list of different supervised and unsupervised learning experiments along with their evaluation sets is tabulated in Table I.

C. Evaluation and Results- Supervised Learning

We fine-tuned the 3D CNN network trained on Sports-1M dataset [34] which had 487 classes. In order to train the network with binary labels for malignancy and the six attributes we used the mid-point as a pivot and labeled samples as positive (or negative) based on their scores being greater (or lesser) than the pivot. In our context, malignancy and attributes are characterized as tasks. The C3D was fine-tuned with these 7 tasks and 10 fold cross-validation was conducted. The requirement to have a large amount of labeled training data was evaded by fine-tuning the network. Since the input to the network required 3 channel image sequences with at least 16 slices, we concatenated the gray level axial channel as the other two channels.

Additionally, in order to ascertain that all input volumes have 16 slices, we performed interpolation where warranted. The final feature representation was obtained from the

TABLE II: The comparison of the proposed approach with other methods using regression accuracy and mean absolute score difference for lung nodule characterization.

Methods	Accuracy %	Mean Score Difference
GIST features + LASSO	76.83	0.675
GIST features + RR	76.48	0.674
3D CNN features + LASSO (<i>Pre-trained</i>)	86.02	0.530
3D CNN features + RR (<i>Pre-trained</i>)	82.00	0.597
3D CNN features + LASSO (<i>Fine-tuned</i>)	88.04	0.497
3D CNN features + RR (<i>Fine-tuned</i>)	84.53	0.550
3D CNN MTL with Trace norm	80.08	0.626
Proposed (3D CNN with Multi-task Learning- Eq. 7)	91.26	0.459

first fully connected layer of 3D CNN consisting of 4096-dimensions.

For computing structure matrix S , we calculate the correlation between different tasks by estimating the normalized coefficient matrix \mathbf{W} via least square loss function with lasso followed by the calculation of correlation coefficient matrix [36]. In order to get a binary graph structure matrix, we thresholded the correlation coefficient matrix. As priors in Eq. (6) we used ρ_1 and ρ_2 as 1 and 10 respectively. Finally, to obtain the malignancy score for test images, the features from the network trained on malignancy were multiplied with the corresponding task coefficient vector W .

We evaluated our proposed approach using both classification and regression metrics. For classification, we considered a nodule to be successfully classified if its predicted score lies in ± 1 of the ground truth score. For regression, we calculated average absolute score difference between the predicted score and the true score. The comparison of our proposed MTL approach with approaches including GIST features [41], 3D CNN features from pre-trained network + LASSO, Ridge Regression (RR) and 3D CNN MTL+trace norm is tabulated in Table II. It can be observed that our proposed graph regularized MTL performs significantly better than other approaches both in terms of classification accuracy as well as the mean score difference. The gain in classification accuracy was found to be 15% and 11% for GIST and trace-norm respectively. In comparison with the pre-trained network, we obtain an improvement of 5% with proposed MTL. In addition, our proposed approach reduces the average absolute score difference for GIST by 32% and for trace-norm by 27%.

D. Evaluations and Results- Unsupervised Learning

For unsupervised learning, evaluations were performed on both lung nodules and IPMN datasets. In order to compute image level features, we used GIST descriptors [41]. The number of clusters is fixed as 2, which accounts for benign and malignant classes. The clustering result was checked to assign benign and malignant labels to the clusters. We used 10 fold cross-validation to evaluate our proposed approach. The training samples along with the label proportions generated

using clustering served as the input to α SVM with a linear kernel.

To evaluate our unsupervised approach we used accuracy, sensitivity and specificity as metrics. It can be observed in Table III that the proposed combination of clustering and α SVM significantly outperforms other approaches in accuracy and sensitivity. In comparison with clustering+SVM, the proposed framework yields almost 21% improvement in sensitivity for lung nodules and around 7% improvement for IPMN classification. The low sensitivity and high specificity of clustering, clustering+SVM, and clustering+RF approaches can be explained by disproportionate assignment of instances as benign (normal) by these approaches, which is not found in the proposed approach. At the same time, the proposed approach records around 24% and 9% improvement in accuracy as compared to clustering for lung nodules and IPMN, respectively.

Are Deep Features good for Unsupervised Classification?

Given the success of deep learning features for image classification and their popularity with the medical imaging community, we explored their performance to classify lung nodules and IPMN in an unsupervised manner. For this purpose, we used a pre-trained deep CNN architecture to extract features and then perform clustering to obtain baseline classification performance. We extracted features from fully-connected layers 7 and 8 of Fast-VGG [42] with and without applying ReLU non-linearity. Classification accuracy, using clustering over these features is shown in Figure 5.

It can be seen in Figure 5 that the features with non-linearity (ReLU) are more discriminative for classification using clustering as compared to without ReLU. The same trend can be observed for both lung nodules and IPMN classification using VGG-fc7 and VGG-fc8 layers. Owing to the larger evaluation set, the influence of ReLU is more prominent for lung nodules as compared to IPMN. Although the results between VGG-fc7 and VGG-fc8 are not substantially different, the highest accuracy for IPMN can be obtained by using VGG-fc7-ReLU features and for lung nodules by using VGG-fc8-ReLU features. The non-linearity induced by ReLU clips the negative values to

TABLE III: Average classification accuracy, sensitivity, and specificity of the proposed *unsupervised* approach for IPMN and lung nodule classification with other methods

Evaluation Set	Methods	Accuracy	Sensitivity	Specificity
<i>IPMN Classification</i>	Clustering	49.18%	45.34%	62.83%
	Clustering + RF	53.20%	51.28%	69.33%
	Clustering + SVM	52.03%	51.96%	50.5%
	Proposed approach	58.04%	58.61%	41.67%
<i>Lung Nodule Classification</i>	Clustering	54.83%	48.69%	60.04%
	Clustering + RF	76.74%	58.59%	91.40%
	Clustering + SVM	76.04%	57.08%	91.28%
	Proposed approach	78.06%	77.85%	78.28%

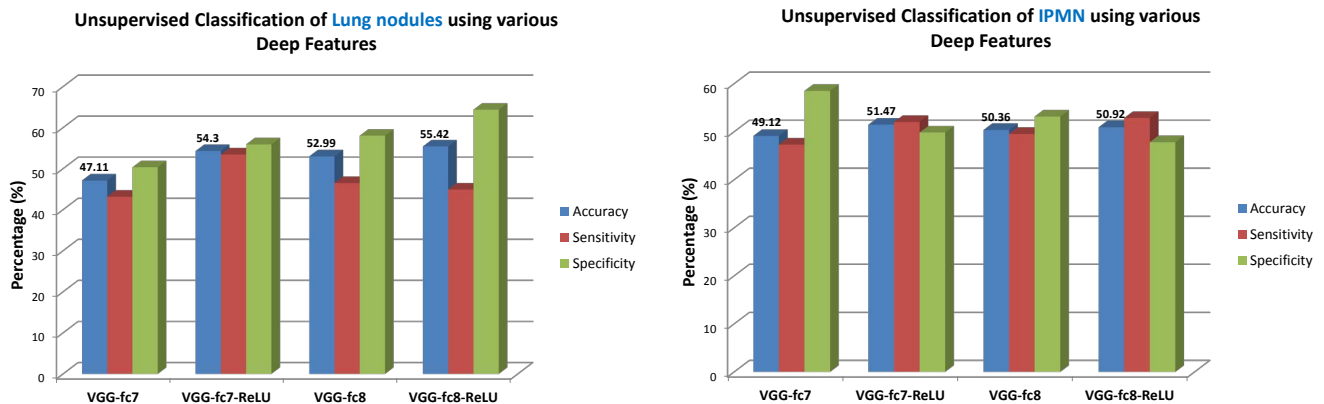


Figure 5: Influence of deep learning features obtained from different layers of a VGG network with and without ReLU non-linearities. The graph on the left shows accuracy, sensitivity and specificity for unsupervised lung nodule classification (clustering), whereas the right one shows the corresponding results for IPMN.

zero, which can sparsify the feature vector and can reduce overfitting. Additionally, it can be seen that GIST features yield comparable performance than deep features (Table III). This can be explained by the fact that the deep networks were trained on ImageNet dataset so the filters in the networks were more tuned to the variations in natural images than medical images. Classification improvement can be expected with unsupervised feature learning techniques such as GANs [30].

Classification using Supervised Learning

In order to establish the upper-bound on the classification performance, we trained linear SVM and Random Forest using GIST and different deep learning features with ground truth labels on the same 10 fold cross-validations sets. Table IV lists the classification accuracy, sensitivity, and specificity using GIST, VGG-fc7 and VGG-fc8 features for both IPMN and lung nodules. For both VGG-fc7 and VGG-fc8, we used features after ReLU since they are found to be more discriminative (Figure 5). Interestingly, for lung nodules, VGG-fc7 features along with RF classifier are reported to have comparable results to the combination of GIST and RF classifier. This can be explained by the fact that deep networks are pre-trained on ImageNet dataset as compared to handcrafted features such as GIST, which don't

require any training. On the other hand, for smaller datasets such as IPMN, deep features are found to perform better as compared to GIST. In order to balance the number of positive (IPMN) and negative (normal) examples, which can be a critical drawback otherwise, we performed Adaptive Synthetic Sampling [43]. This was done to generate synthetic examples in terms of features from the minority class (normal).

VI. DISCUSSION AND CONCLUDING REMARKS

In this study, we present a framework for the malignancy determination of lung nodules with 3D CNN based graph regularized sparse MTL. To the best of our knowledge, this is the first work where MTL and transfer learning are studied for 3D deep networks to improve risk stratification of lung nodules. Usually, the data sharing for medical imaging is highly regulated and the accessibility of experts (radiologists) to label these images is limited. As a consequence, the access to the crowdsourced and publicly gathered and annotated data such as videos may help in obtaining discriminative features for medical image analysis.

We also analyzed the significance of different imaging attributes corresponding to lung nodules including spiculation, texture, calcification and others for risk assessment. Instead

TABLE IV: Classification of IPMN and Lung Nodules using different features and supervised learning classifiers.

Evaluation Set	Features	Classifiers	Accuracy (%)	Sensitivity (%)	Specificity (%)
<i>IPMN Classification</i>	GIST	SVM	76.05	83.65	52.67
		RF	81.9	93.69	43.0
	VGG-fc7	SVM	84.18	96.91	44.83
		RF	81.96	94.61	42.83
	VGG-fc8	SVM	84.22	97.2	46.5
		RF	80.82	93.4	45.67
<i>Lung Nodule Classification</i>	GIST	SVM	81.56	71.31	90.02
		RF	81.64	76.47	85.97
	VGG-fc7	SVM	77.97	75.2	80.6
		RF	81.73	78.24	84.59
	VGG-fc8	SVM	78.76	74.67	82.29
		RF	80.51	76.03	84.24

of manually modeling these attributes we utilized 3D CNN to learn rich feature representations associated with these attributes. The graph regularized sparse MTL framework was employed to integrate 3D CNN features from these attributes. We have found the features associated with these attributes complementary to those corresponding to malignancy.

In the second part of this study, we explored the potential of unsupervised learning for malignancy determination. Since in most medical imaging tasks radiologists are required to get annotations, acquiring labels to learn machine learning models is more cumbersome and expensive as compared to other computer vision tasks. In order to address this challenge, we employed clustering to obtain an initial set of labels and progressively refined them with α SVM. We obtained promising results and our proposed approach outperformed the other methods in evaluation metrics.

Following up on the application of deep learning for almost all tasks in the visual domain, we studied the influence of different pre-trained deep networks for lung nodule classification. For some instances, we found that commonly used imaging features such as GIST have comparable results as those obtained from pre-trained network features. This observation can be explained by the fact that the deep networks were trained on ImageNet classification tasks so the filters in CNN were more tuned to the nuances in natural images as compared to medical images.

To the best of our knowledge, this is one of the first and the largest evaluation of a CAD system for IPMN classification. CAD systems for IPMN classification are relatively newer research problems and there is a need to explore the use of different imaging modalities to improve classification. Although MRI remains the most common modality to study pancreatic cysts, CT images can also be used as a complementary imaging modality due to its higher resolution and its ability to capture smaller cysts. Additionally, a combination of T2-weighted, contrast-enhanced and unenhanced T1-weighted sequences can help improve detection and diagnosis of IPMN [44]. In this regard, multi-modal deep learning architectures can be deemed useful [45]. The detection and segmentation of pancreas can also be useful to make a better prediction about the presence of IPMN and cysts. Due to its anatomy, the pancreas is a

challenging organ to segment, particularly in MRI images. To address this challenge, other imaging modalities can be utilized for joint segmentation and diagnosis of pancreatic cysts and IPMN. Furthermore, visualization of activation maps can be quite useful for the clinicians to identify new imaging biomarkers that can be employed for diagnosis in the future.

The future prospects of using different architectures to perform unsupervised representation learning using GAN are promising. Instead of using hand-engineered priors of sampling in the generator, the work in [46] learned priors using denoising auto-encoders. For measuring the sample similarity for complex distributions such as those in the images, [47] jointly trained variational autoencoders and GANs. Moreover, the applications of CatGAN [48] and InfoGAN [49] for semi-supervised and unsupervised classification tasks in medical imaging are worth exploring as well.

Medical imaging has unique challenges associated with the scarcity of labeled examples. Moreover, unless corroborated by biopsy, there may exist a large variability in labeling from different radiologists. Although fine-tuning has helped to address the lack of annotated examples, the performance is limited due to large differences in domains. It is comparatively easier to obtain scan level labels than slice level labels. In this regard, weakly supervised approaches such as multiple instance learning (MIL) can be of great value. Active learning can be another solution to alleviate the difficulty in labeling. In addition to these directions, unsupervised learning approaches will surely be pursued to address unique medical imaging challenges.

REFERENCES

- [1] Society, A.C.: Cancer Facts & Figures. American Cancer Society (2016)
- [2] Sadot, E., Basturk, O., Klimstra, D.S., Gönen, M., Anna, L., Do, R.K.G., D'Angelica, M.I., DeMatteo, R.P., Kingham, T.P., Jarnagin, W.R., et al.: Tumor-associated neutrophils and malignant progression in intraductal papillary mucinous neoplasms: an opportunity for identification of high-risk disease. *Annals of surgery* 262(6), 1102 (2015)
- [3] El-Baz, A., Nitzken, M., Khalifa, F., Elnakib, A., Gimelfarb, G., Falk, R., El-Ghar, M.A.: 3D shape analysis for early diagnosis of malignant lung nodules. In: *IPMI*. pp. 772–783. Springer (2011)
- [4] Han, F., Wang, H., Zhang, G., Han, H., Song, B., Li, L., Moore, W., Lu, H., Zhao, H., Liang, Z.: Texture feature analysis for computer-aided diagnosis on pulmonary nodules. *Journal of Digital Imaging* 28(1), 99–115 (2015)

- [5] Way, T.W., Hadjiiski, L.M., Sahiner, B., Chan, H.P., Cascade, P.N., Kazerooni, E.A., Bogot, N., Zhou, C.: Computer-aided diagnosis of pulmonary nodules on CT scans: segmentation and classification using 3D active contours. *Medical Physics* 33(7), 2323–2337 (2006)
- [6] Lee, M., Boroczky, L., Sungur-Stasik, K., Cann, A., Borczuk, A., Kawut, S., Powell, C.: Computer-aided diagnosis of pulmonary nodules using a two-step approach for feature selection and classifier ensemble construction. *Artificial Intelligence in Medicine* 50(1), 43–53 (2010)
- [7] Kumar, D., Wong, A., Clausi, D.A.: Lung nodule classification using deep features in CT images. In: *Computer and Robot Vision (CRV)*, 2015 12th Conference on. pp. 133–138. IEEE (2015)
- [8] Buty, M., Xu, Z., Gao, M., Bagci, U., Wu, A., Mollura, D.J.: Characterization of Lung Nodule Malignancy Using Hybrid Shape and Appearance Features. In: *MICCAI*. pp. 662–670. Springer (2016)
- [9] Saouli, R., Akil, M., Kachouri, R., et al.: Fully automatic brain tumor segmentation using end-to-end incremental deep neural networks in mri images. *Computer methods and programs in biomedicine* 166, 39–49 (2018)
- [10] Hussein, S., Cao, K., Song, Q., Bagci, U.: Risk Stratification of Lung Nodules Using 3D CNN-Based Multi-task Learning. In: *International Conference on Information Processing in Medical Imaging*. pp. 249–260. Springer (2017)
- [11] Furuya, K., Murayama, S., Soeda, H., Murakami, J., Ichinose, Y., Yauuchi, H., Katsuda, Y., Koga, M., Masuda, K.: New classification of small pulmonary nodules by argin characteristics on high-resolution CT. *Acta Radiologica* 40(5), 496–504 (1999)
- [12] Uchiyama, Y., Katsuragawa, S., Abe, H., Shiraiishi, J., Li, F., Li, Q., Zhang, C.T., Suzuki, K., Doi, K.: Quantitative computerized analysis of diffuse lung disease in high-resolution computed tomography. *Medical Physics* 30(9), 2440–2454 (2003)
- [13] Chen, S., Ni, D., Qin, J., Lei, B., Wang, T., Cheng, J.Z.: Bridging computational features toward multiple semantic features with multi-task regression: A study of CT pulmonary nodules. In: *MICCAI*. pp. 53–60. Springer (2016)
- [14] Shen, W., Zhou, M., Yang, F., Yang, C., Tian, J.: Multi-scale convolutional neural networks for lung nodule classification. In: *IPMI*. pp. 588–599. Springer (2015)
- [15] Ciompi, F., Chung, K., Van Riel, S.J., Setio, A.A.A., Gerke, P.K., Jacobs, C., Scholten, E.T., Schaefer-Prokop, C., Wille, M.M., Marchianò, A., et al.: Towards automatic pulmonary nodule management in lung cancer screening with deep learning. *Scientific reports* 7, 46479 (2017)
- [16] Setio, A.A.A., Ciompi, F., Litjens, G., Gerke, P., Jacobs, C., van Riel, S.J., Wille, M.M.W., Naqibullah, M., Sánchez, C.I., van Ginneken, B.: Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. *IEEE TMI* 35(5), 1160–1169 (2016)
- [17] Setio, A.A.A., Traverso, A., De Bel, T., Berens, M.S., van den Bogaard, C., Cerello, P., Chen, H., Dou, Q., Fantacci, M.E., Geurts, B., et al.: Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the luna16 challenge. *Medical image analysis* 42, 1–13 (2017)
- [18] Khosravan, N., Bagci, U.: S4ND: Single-Shot Single-Scale Lung Nodule Detection. *arXiv preprint arXiv:1805.02279* (2018)
- [19] Zhou, Y., Xie, L., Fishman, E.K., Yuille, A.L.: Deep Supervision for Pancreatic Cyst Segmentation in Abdominal CT Scans. *arXiv preprint arXiv:1706.07346* (2017)
- [20] Cai, J., Lu, L., Zhang, Z., Xing, F., Yang, L., Yin, Q.: Pancreas segmentation in MRI using graph-based decision fusion on convolutional neural networks. In: *MICCAI*. pp. 442–450. Springer (2016)
- [21] Hanania, A.N., Bantis, L.E., Feng, Z., Wang, H., Tamm, E.P., Katz, M.H., Maitra, A., Koay, E.J.: Quantitative imaging to evaluate malignant potential of IPMNs. *Oncotarget* 7(52), 85776 (2016)
- [22] Gazit, L., Chakraborty, J., Attiyeh, M., Langdon-Embry, L., Allen, P.J., Do, R.K., Simpson, A.L.: Quantification of CT Images for the Classification of High- and Low-Risk Pancreatic Cysts. In: *SPIE Medical Imaging*. pp. 101340X–101340X. International Society for Optics and Photonics (2017)
- [23] Sivic, J., Russell, B.C., Efros, A.A., Zisserman, A., Freeman, W.T.: Discovering objects and their location in images. In: *ICCV*. vol. 1, pp. 370–377. IEEE (2005)
- [24] Kamper, H., Jansen, A., Goldwater, S.: Fully unsupervised small-vocabulary speech recognition using a segmental Bayesian model. In: *Interspeech* (2015)
- [25] Lee, H., Pham, P., Largman, Y., Ng, A.Y.: Unsupervised feature learning for audio classification using convolutional deep belief networks. In: *Advances in neural information processing systems*. pp. 1096–1104 (2009)
- [26] Shin, H.C., Orton, M.R., Collins, D.J., Doran, S.J., Leach, M.O.: Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4d patient data. *IEEE transactions on pattern analysis and machine intelligence* 35(8), 1930–1943 (2013)
- [27] Vaidhya, K., Thirunavukkarasu, S., Alex, V., Krishnamurthi, G.: Multi-modal brain tumor segmentation using stacked denoising autoencoders. In: *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. pp. 181–194. Springer (2015)
- [28] Sivakumar, S., Chandrasekar, C.: Lung nodule segmentation through unsupervised clustering models. *Procedia engineering* 38, 3064–3073 (2012)
- [29] Kallenberg, M., Petersen, K., Nielsen, M., Ng, A.Y., Diao, P., Igel, C., Vachon, C.M., Holland, K., Winkel, R.R., Karssemeijer, N., et al.: Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring. *IEEE transactions on medical imaging* 35(5), 1322–1331 (2016)
- [30] Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434* (2015)
- [31] Zhang, F., Song, Y., Cai, W., Zhou, Y., Fulham, M., Eberl, S., Shan, S., Feng, D.: A ranking-based lung nodule image classification method using unlabeled image knowledge. In: *IEEE ISBI*. pp. 1356–1359. IEEE (2014)
- [32] Armato III, S., McLennan, G., Bidaut, L., McNitt-Gray, M.F., Meyer, C.R., Reeves, A.P., Zhao, B., Aberler, D.R., Henschke, C.I., Hoffman, E.A., et al.: The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans. *Medical Physics* 38(2), 915–931 (2011)
- [33] Tran, D., Bourdev, L., Fergus, R., Torresani, L., Paluri, M.: Learning spatiotemporal features with 3D convolutional networks. In: *ICCV*. pp. 4489–4497. IEEE (2015)
- [34] Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., Fei-Fei, L.: Large-scale video classification with convolutional neural networks. In: *IEEE CVPR*. pp. 1725–1732 (2014)
- [35] Recht, B., Fazel, M., Parrilo, P.A.: Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review* 52(3), 471–501 (2010)
- [36] Zhou, J., Chen, J., Ye, J.: MALSAR: Multi-task learning via structural regularization (2012)
- [37] Shalev-Shwartz, S., Tewari, A.: Stochastic methods for l_1 -regularized loss minimization. *Journal of Machine Learning Research* 12(Jun), 1865–1892 (2011)
- [38] Nesterov, Y.: *Introductory lectures on convex optimization: A basic course*, vol. 87. Springer Science & Business Media (2013)
- [39] Yu, F., Liu, D., Kumar, S., Tony, J., Chang, S.F.: α SVM for Learning with Label Proportions. In: *Proceedings of The 30th International Conference on Machine Learning*. pp. 504–512 (2013)
- [40] Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C.: N4ITK: Improved N3 bias correction. *IEEE Transactions on Medical Imaging* 29(6), 1310–1320 (2010)
- [41] Oliva, A., Torralba, A.: Modeling the shape of the scene: A holistic representation of the spatial envelope. *IJCV* 42(3), 145–175 (2001)
- [42] Chatfield, K., Simonyan, K., Vedaldi, A., Zisserman, A.: Return of the devil in the details: Delving deep into convolutional nets. In: *British Machine Vision Conference* (2014)
- [43] He, H., Bai, Y., Garcia, E.A., Li, S.: ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In: *Neural Networks, 2008. IJCNN 2008. (IEEE World Congress on Computational Intelligence)*. IEEE International Joint Conference on. pp. 1322–1328. IEEE (2008)
- [44] Kalb, B., Sarmiento, J.M., Kooby, D.A., Adsay, N.V., Martin, D.R.: MR imaging of cystic lesions of the pancreas. *Radiographics* 29(6), 1749–1765 (2009)
- [45] Ma, L., Lu, Z., Shang, L., Li, H.: Multimodal convolutional neural networks for matching image and sentence. In: *IEEE ICCV*. pp. 2623–2631 (2015)
- [46] Nguyen, A., Yosinski, J., Bengio, Y., Dosovitskiy, A., Clune, J.: Plug & play generative networks: Conditional iterative generation of images in latent space. *arXiv preprint arXiv:1612.00005* (2016)
- [47] Larsen, A.B.L., Sønderby, S.K., Larochelle, H., Winther, O.: Autoencoding beyond pixels using a learned similarity metric. In: *ICML* (2016)
- [48] Springenberg, J.T.: Unsupervised and semi-supervised learning with categorical generative adversarial networks. *arXiv preprint arXiv:1511.06390* (2015)
- [49] Chen, X., Duan, Y., Houthoofd, R., Schulman, J., Sutskever, I., Abbeel, P.: Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In: *Advances in Neural Information Processing Systems*. pp. 2172–2180 (2016)