Diagnosis of Coronavirus Disease 2019 (COVID-19) With Structured Latent Multi-View Representation Learning

Hengyuan Kang, Liming Xia, Fuhua Yan, Zhibin Wan, Feng Shi⬛, Huan Yuan, Huiting Jiang, Dijia Wu, He Sui, Changqing Zhang⬛, and Dinggang Shen⬛

Abstract— Recently, the outbreak of Coronavirus Disease 2019 (COVID-19) has spread rapidly across the world. Due to the large number of infected patients and heavy labor for doctors, computer-aided diagnosis with machine learning algorithm is urgently needed, and could largely reduce the efforts of clinicians and accelerate the diagnosis process. Chest computed tomography (CT) has been recognized as an informative tool for diagnosis of the disease. In this study, we propose to conduct the diagnosis of COVID-19 with a series of features extracted from CT images. To fully explore multiple features describing CT images from different views, a unified latent representation is learned which can completely encode information from different aspects of features and is endowed with promising class structure for separability. Specifically, the completeness is guaranteed with a group of backward neural networks (each for one type of features), while by using class labels the representation is enforced to be compact within COVID-19/community-acquired pneumonia (CAP) and also a large margin is guaranteed between different types of pneumonia. In this way, our model can well avoid overfitting compared to the case of directly projecting high-dimensional features into classes. Extensive experimental results show that the proposed method outperforms all comparison methods, and rather stable performances are observed when varying the number of training data.

Index Terms— COVID-19, Pneumonia, Chest computed tomography (CT), Multi-view representation learning.

I. Introduction

A NOVEL coronavirus disease 2019 (COVID-19) was recognized in December 2019, in Wuhan, China and has rapidly spread over the world [1]–[5]. Recently, COVID-19 has been threatening all the world and the world health organization (WHO) has declared that COVID-19 becomes a global pandemic. The current clinical experience implies that the RT-PCR detection of viral RNA has a low sensitivity especially in the early stage [6]–[9]. As a form of pneumonia, inflammation of air sacs in lungs has been found, and it has shown that bilateral lung involvement could be observed for early, intermediate, and late stage patients. Accordingly, a high proportion of abnormal chest CT images were obtained from patients with this disease [10]–[13]. Then, it is necessary to complement nucleic acid testing with automatic technique based on lung CT as one of the early diagnostic criteria for this new type of pneumonia as soon as possible.

In this study, we focus on conducting diagnosis for COVID-19 and community-acquired pneumonia [14]–[16], i.e., characterizing the relationships between multiple types of features from CT images and these diseases, which provides a possible pipeline for automatic diagnosis and investigation. Specifically, multiple types of features are extracted and the correlations to diagnosis are extensively evaluated by conducting experiments with multiple baseline methods. The experiment results show that both radiomic features and handcrafted features are helpful for classifying these two different categories of lung diseases. Therefore, we propose a novel pipeline which can effectively integrate information from different views. We also note that, although deep-learning methods [17], especially CNN-based models, have shown the power in image classification, they usually need large-scale training data and have difficulty in exploiting expert prior. Although there are thousands of CT images available which is a quite large dataset for medical image analysis,

Hengyuan Kang, Zhibin Wan, and Changqing Zhang are with the College of Intelligence and Computing, Tianjin University, Tianjin 300350, China (e-mail: kanghengyuan@tju.edu.cn; wanzhibin@tju.edu.cn; zhangchangqing@tju.edu.cn).

Liming Xia is with the Department of Radiology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China (e-mail: xialiming2017@outlook.com).

Fuhua Yan is with the Department of Radiology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China (e-mail: yfh11655@rjh.com.cn).

Feng Shi, Huan Yuan, Huiting Jiang, Dijia Wu, and Dinggang Shen are with the Department of Research and Development, Shanghai United Imaging Intelligence Company Ltd., Shanghai 201210, China (e-mail: feng.shi@united-imaging.com; huan.yuan@united-imaging.com; huiting.jiang@united-imaging.com; dijia.wu@united-imaging.com; dinggan'g.shen@gmail.com).

He Sui is with the Department of Radiology, China–Japan Union Hospital of Jilin University, Changchun 130033, China (e-mail: suihe910402@126.com).
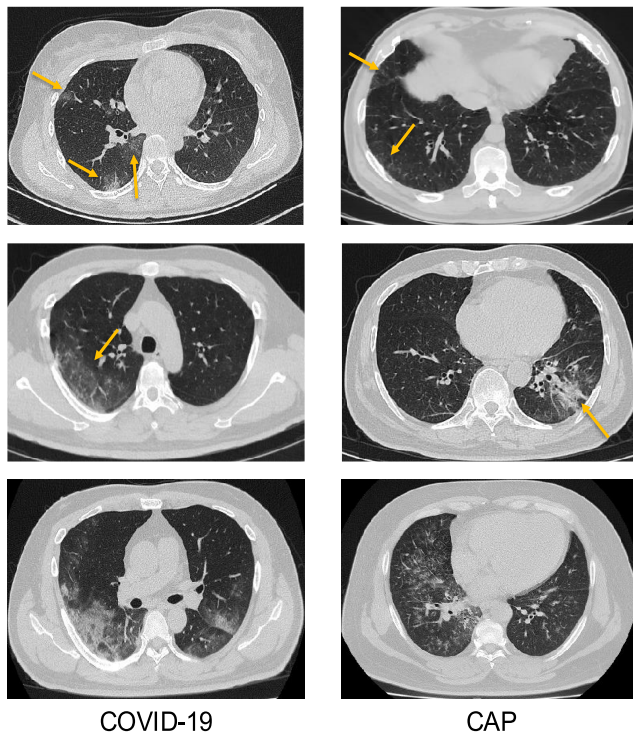
Fig. 1. Examples of chest CT images with infection of COVID-19 (left) and community-acquired pneumonia (right). The pneumonia becomes more serious from top to bottom, and the yellow arrows indicate the representative infection areas. It can be observed that it is quite similar for the CT images of high-severity CAP and mild-severity COVID-19.

it is still not comparable with the large-scale image dataset in the computer vision field, i.e., ImageNet [18]. Therefore, extracting manually designed features to incorporate expert prior is a reasonable and probably preferred approach which could alleviate the overfitting problem in machine learning.

For computer-assisted medical diagnosis, a number of methods automatically learn classification models based on features extracted with expert prior [19], [20]. However, some of them are only applicable to single type of features thus cannot well explore the complementary information from multiple types of features. Fortunately, multi-view representation learning provides tools for exploiting multiple types of heterogeneous features [21], [22]. Although significant progress achieved, existing multi-view representation learning methods cannot guarantee the information completeness and promising structure for separability. This makes them tend to overfit training data thus harm the performance in the testing stage.

Based on the above analysis, we propose a new classification framework to diagnose these diseases, where the main aim is to identify the COVID-19 from CAP. It is worth noting that these different types of features are not directly used as input for a classifier to output the final decision [23]–[25] in our method. Instead, both the training and testing samples are mapped into a promising latent space [26], where these latent representations are expected to encode complementary information from different types of features and have promising structures revealing the underlying class distribution.

First, since there are different types of features which are quite different in distribution and quality as shown in Fig. 5, it is very challenging and important to effectively integrate these different types of features. For this issue, a novel integration strategy is proposed with a group of neural networks, where each one encodes the information from one type of features into a latent representation. Second, we conduct projection learning to build an accurate model to map a subject with these multiple types of features into a latent representation, thus a latent-representation regressor is obtained which can be applied on new subjects. Third, a final classifier is trained based on the latent representation, instead of the original features. We should emphasize the advantages of our model over existing methods that often directly learn projections from original features into class labels. The first advantage is the latent representation, which is usually compact and thus may be more effective since it can avoid to overfit the high-dimensional data and has better generalization in the testing stage. Second, the proposed pipeline can encode information from different types of features and produce a structured representation which imposes simple bias on the model to further enhance the generalization performance. Moreover, the learned representation could be used in different classification models, and the performance with the learned latent representations clearly outperforms that of original features for all baseline classifiers used in experiments. The main contributions of this study are summarized as follows:

- We propose to conduct diagnosis of COVID-19 with multi-view representation learning, where the complementarity among different types of features is well explored, achieving clear performance gain in classification.
- We propose a full pipeline for diagnosis COVID-19 from community-acquired pneumonia, which is quite different from existing models that directly project features into the class space. The key component is the structured latent representation learning which brings robustness, generalization and stability into the pipeline.
- The learned latent representation can be widely used for different classifiers to promote the diagnosis accuracy. Specifically, the latent representations are adopted by several baseline models, and the results clearly demonstrate the effectiveness of the latent representation compared with original features.
- Extensive experiments on the CT images validate that the proposed model can obtain a well-structured latent representation, thus significantly promoting the diagnosis in terms of accuracy, sensitivity and specificity compared with different methods.

II. MATERIAL

There are 2522 CT images involved in this study, where 1495 cases are from COVID-19 patients and the left 1027 cases are from community-acquired pneumonia (CAP) patients. These COVID-19 infected subjects were confirmed with positive nucleic acid testing and confirmed by Chinese Centers for Disease Control and Prevention (CDC). The distributions of subjects in terms of gender and age are shown

TABLE I
SUBJECT DISTRIBUTIONS IN TERMS OF GENDER

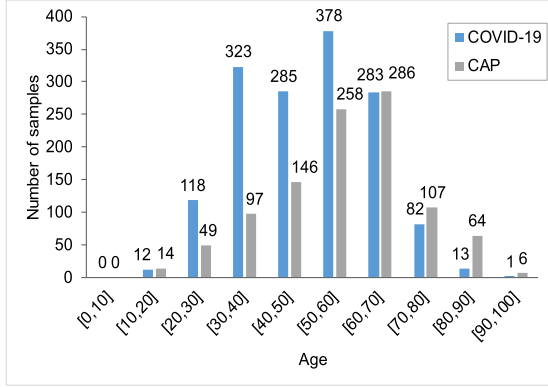| Gender	COVID-19	CAP	Total
Male	770	488	1258
Female	725	539	1264
Total	1495	1027	2522

Fig. 2. Subject distributions in terms of age.

in Table I and Fig. 2. Specifically, according to Table I, we can find that the number of males with COVID-19 is slightly larger than that of females, while, for CAP, the contrary is the case. These subjects cover patients from 12-year old to 98-year old. In addition, basically, the average age of patients infected with COVID-19 is younger than that of CAP according to Fig. 2. These CT images were provided by Tongji Hospital of Huazhong University of Science and Technology, China-Japan Union Hospital of Jilin University, Ruijin Hospital of Shanghai Jiao Tong University, and their collaborators. The COVID-19 images were acquired from Jan. 9, 2020 to Feb. 14, 2020, and CAP images were obtained from Jul. 30, 2018 to Feb. 22, 2020.

Chest CT scans were performed on all patients with thin section. The CT scanners used in the three hospitals mentioned include uCT 780 from UIH, Optima CT520, Discovery CT750, LightSpeed 16 from GE, Aquilion ONE from Toshiba, SOMATOM Force from Siemens, and SCENARIA from Hitachi. Furthermore, CT scanners were carried out with the protocol which include: 120 kV, reconstructed CT thickness ranges from 0.625 to 2mm, with breath hold at full inspiration.

In this study, all images were preprocessed with a V-Net model [27] to extract the lung, lung lobes, and pulmonary segments. The infected lesions were also segmented [28]. After that, based on the lesion region, 189-dimensional features were extracted from each CT image in total. First, according to different approaches of extracting, we divide the features into radiomic features and handcrafted features. Furthermore, we split the radiomic features into two types of features that characterize the CT images from different perspectives: **Gray features** are composed of the first-order statistics which describe the distribution of voxel intensities within the image

TABLE II
ABSTRACT FOR DIFFERENT TYPES OF FEATURES

	Feature groups	# of features
Radiomic feature	Gray features	19
	Texture features	74
Handcrafted feature	Histogram features	30
	Number features	24
	Intensity features	2
	Surface features	7
	Volume features	33

region, such as maximum, minimum, median and so on. **Texture features** are derived from gray level co-occurrence matrix (GLCM), gray level size zone matrix (GLSZM), gray level run length matrix (GLRLM), neighboring gray tone difference matrix (NGTDM) and gray level dependence matrix (GLDM) [29]. The handcrafted features are divided into five groups based on different characteristics of the lesions: **Histogram features** are composed of frequency of intensity level in the infection area at 30 equal bins which are divided from the intensity value (between -1350 and 150 [30]) range of lung area image. **Number features** are composed of the total number of infected areas in the bilateral lungs, lung lobes, and pulmonary segments, respectively. **Intensity features** are composed of the mean and variance value of infection areas. **Volume features** are composed of volume of infected areas of the whole lung, each lobe and pulmonary segment, and the percentage of the infected areas of the whole lung and two lobes. **Surface features** are composed of the infection surface and the lung boundary surface. Furthermore, they contain the distance of each infection surface vertex to the nearest lung boundary surface, and are divided into 5 ranges (3, 6, 9, 12 and 15 voxels (voxel spacing is 1.5mm)). We also calculate the number of infection surface vertices to the lung wall in terms of each range of distance, the percentage of infection vertex number against the whole infection vertices in each range, and the percentage of infection vertex number. These 189 features in total are spit into 7 different non-overlapping groups as shown in Table II.

III. METHOD

There are different types of heterogeneous features from CT images which provide complementary information to diagnosis the COVID-19, hence we employ multi-view machine learning technique [31]–[33] for our task. Inspired by our previous network (CPM-Nets) [26], we further develop a novel diagnosis pipeline to classify COVID-19 and community-acquired pneumonia (CAP). Specifically, these diverse types of features extracted from CT images have extremely different properties, therefore it is unreasonable and ineffective to directly concatenate them without prepossessing or machine learning technique, and this is also validated in experiments.
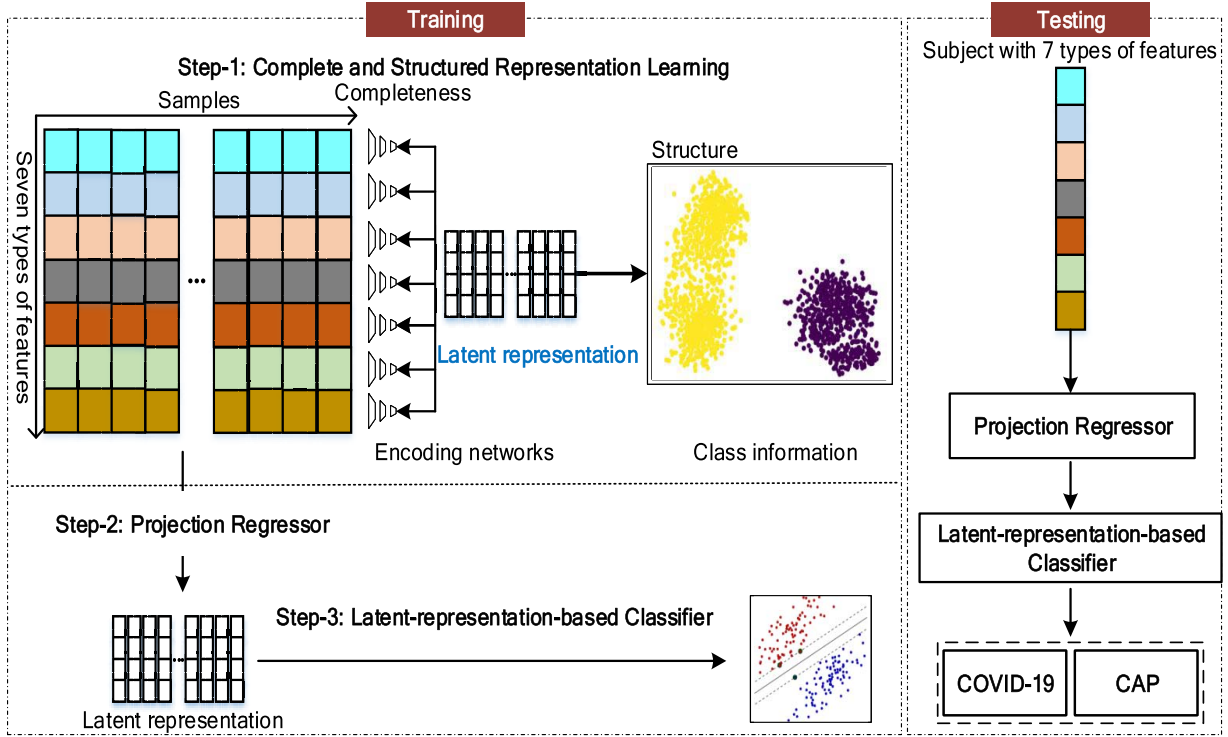
Fig. 3. Overview of the proposed latent-representation-based diagnosis framework. For the input, different colors indicate different types of features, while for the class information yellow and purple indicate COVID-19 and CAP, respectively.

To effectively exploit these multiple types of features from CT images, we propose a latent-representation-based diagnosis pipeline, which is composed of three components in the training stage as shown in Fig. 3. First, based on the CPM-Nets we learn latent representations with information completeness and promising class structure. The latent representations act as bridge of different components. This step is termed as Complete and Structured Representation Learning. Second, for the consistency of latent space [26] between training and testing, we train a projection model termed as Latent-representation Regressor between the 7 types of original features and the latent representations. Third, a latent-representation-based classifier for diagnosis is trained. Accordingly, in the testing stage, the original features are projected into latent space with latent-representation regressor and then the final diagnosis result can be obtained with the latent-representation-based classifier.

A. Step-1: Complete and Structured Representation Learning

Considering our aim is to discriminate two types of CT images associated with COVID-19 and CAP, we learn latent representations which not only encode information of heterogeneous features but also reflect the class distribution. Then, the latent representations will be both informative and separable. For clarification, we first give the formal definition of notations as follows. Given the training set $\{\mathcal{X}_n, y_n\}_{n=1}^N$, where $\mathcal{X}_n = \{\mathbf{x}_n^{(v)}\}_{v=1}^V$ is a multi-view sample and y_n is the corresponding class label (i.e., $y_n = 1$ or 0 indicates the subject is infected with COVID-2019 or CAP, respectively).

N and V are the number of subjects in the training stage and the number of types of features (i.e., $V = 7$ in current experiments), respectively.

1) Completeness for Latent Representation: First, we aim to flexibly and effectively integrate different types of information for each subject into a low-dimensional space, where the desired latent representation should involve information from all types of features. From the perspective of reconstruction [34], if a latent representation $\mathbf{h}$ can well reconstruct each type of features with a stable mapping $f_v(\cdot)$, i.e., $\mathbf{x}^{(v)} = f_v(\mathbf{h})$, then it encodes the intrinsic information of different types of features. Therefore, here we try to reconstruct each type of features from the learned latent representation to guarantee the information completeness.

Based on the above analysis, we can integrate information from different types of features into a latent representation as follows:

$$\ell_r(\mathcal{X}_n, \mathbf{h}_n) = \sum_{v=1}^V \| f_v(\mathbf{h}_n; \Theta_r^{(v)}) - \mathbf{x}_n^{(v)} \|^2 \tag{1}$$

where $f_v(\cdot; \Theta_r^{(v)})$ is the reconstruction network for the vth type of features parameterized by $\Theta_r^{(v)}$. $\mathbf{h}_n$ represents the learned latent representation. Ideally, by minimizing Eq. (1), we can well encode information from different types of features into the complete representation $\mathbf{h}_n$.

2) Structure for Latent Representation: Second, we aim to make the learned latent representation to be well structured with respect to these two different pneumonia diseases. Specifically, the loss for structured representations is
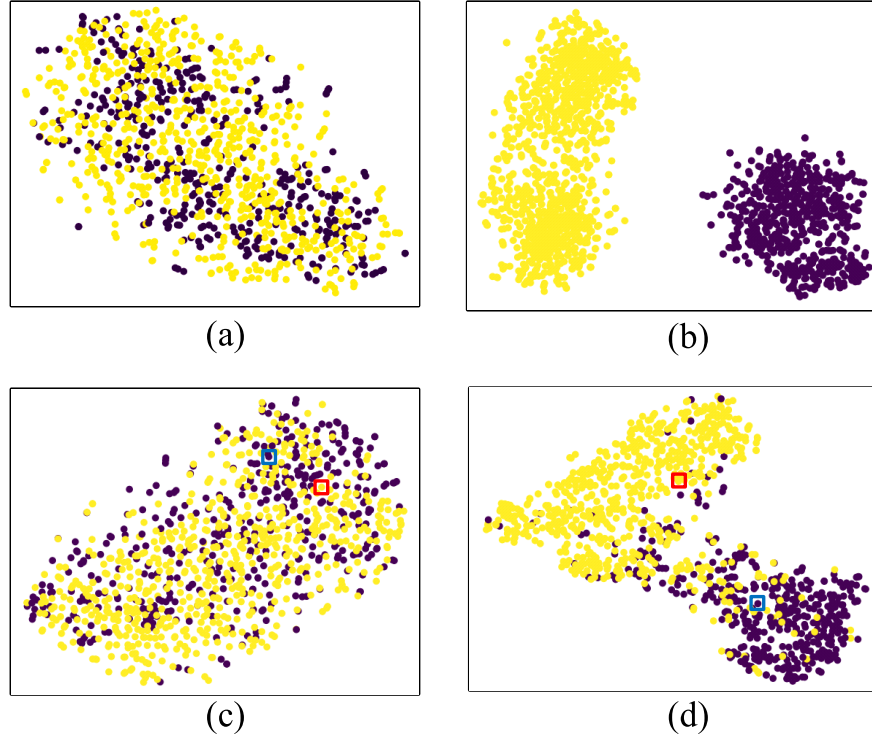
Fig. 4. Visualization of the latent representations in the training and testing stages. Given the original features, the visualization in (a) indicates that the underlying class structure is not well revealed, while the learned latent representations in (b) are much better structured and consistent with classes. Similar case is also observed in the testing stage, as shown in (c) and (d). The red and blue boxes in (c) and (d) indicate the same pair examples for COVID-19 and CAP, respectively.

specified as:

$$\Delta(y_n, y) = \Delta(y_n, g(\mathbf{h}_n; \Theta_c)), \tag{2}$$

$$\text{with } g(\mathbf{h}_n; \Theta_c) = \arg\min_{y \in \mathcal{Y}} \mathbb{E}_{\mathbf{h} \sim \mathcal{T}(y)} F(\mathbf{h}, \mathbf{h}_n) \tag{3}$$

where $F(\mathbf{h}, \mathbf{h}_n) = \phi(\mathbf{h}; \Theta_c)^T \phi(\mathbf{h}_n; \Theta_c)$ with $\phi(\cdot; \Theta_c)$ being the feature mapping function for $\mathbf{h}$ parameterized by Θ_c, and $\mathcal{T}(y)$ being the set of latent representation from pneumonia class y. In practice, we set $\mathbf{h} = \phi(\mathbf{h}; \Theta_c)$ for simplicity which makes the loss non-parametric. This loss will enforce the compactness within the same type of pneumonia, and a margin is between COVID-19 and CAP, guaranteeing the separability. Then, we should minimize the following loss function

$$\ell_c(y_n, y, \mathbf{h}_n) = \max_{y \in \mathcal{Y}} \Big(0, \Delta(y_n, y) + \mathbb{E}_{\mathbf{h} \sim \mathcal{T}(y)} F(\mathbf{h}, \mathbf{h}_n)$$
$$- \mathbb{E}_{\mathbf{h} \sim \mathcal{T}(y_n)} F(\mathbf{h}, \mathbf{h}_n) \Big). \tag{4}$$

Based on above analysis, by jointly considering informativeness and separability, we should optimize the following objective function

$$\arg\min_{\Theta_r} \frac{1}{N} \sum_{n=1}^{N} \ell_r(\mathcal{X}_n, \mathbf{h}_n; \Theta_r) + \lambda \ell_c(y_n, y, \mathbf{h}_n). \tag{5}$$

where $\lambda > 0$ is a balance factor between completeness and class labels.

B. Step-2: Learning Projection From Original Features to Latent Representation

In the Step-1, one low-dimensional latent representation $\mathbf{h}$ is obtained for each subject in the training set. As shown in Fig. 3, these representations are rather promising for the diagnosis of COVID-2019 and CAP since the representations associated with COVID-19 and CAP are compact and there is a clear margin between these two types of pneumonia. However, we should note that we cannot obtain the latent representation in the testing stage now. Therefore, we target to design a latent-representation regressor to accurately transform the original features of a subject into a latent representation.

The latent-representation regressor is implemented with fully connected neural networks to learn a mapping $\Gamma(\cdot)$, i.e., $\mathbf{h}'_n = \Gamma(\mathcal{X}_n; \Theta_e)$ with parameter Θ_e from the original features, i.e., $\mathbf{x}_n$ to the corresponding latent representations. Specifically, the regressor is composed of four fully-connected layers and two sigmoid layers. The optimization objective is to minimize the MSE (Mean Squared Error) loss between the output of the regressor and the corresponding latent representations. It can be formulated as

$$\arg\min_{\Theta_e} \frac{1}{N} \sum_{n=0}^{N} \Big(\mathbf{h}_n - \Gamma(\mathcal{X}_n; \Theta_e) \Big)^2. \tag{6}$$

Then, given multiple types of original features of a CT image, the corresponding latent representation can be calculated.

C. Step-3: Latent-Representation-Based Classifier

After obtaining the latent representation, we then target to train a latent-representation-based classifier which can diagnosis the subjects between COVID-2019 and CAP. For simplicity, we employ a neural network with three fully-connected layers as the latent-representation-based classifier. The widely used cross-entropy loss is employed in our classification tasks. Then, we should minimize the following loss function

$$\ell = -\frac{1}{N} \sum_{n=0}^{N-1} \left[y_n \log(\Phi_{\boldsymbol{\theta}}(\mathbf{h}'_n)) + (1 - y_n) \log(1 - \Phi_{\boldsymbol{\theta}}(\mathbf{h}'_n)) \right] \tag{7}$$

where $\Phi_{\boldsymbol{\theta}}(\mathbf{h}'_n)$ indicates the prediction of pneumonia type from the classifier with $\mathbf{h}'$ as the input.

D. Testing Stage

After training, a full pipeline for diagnosis of COVID-19 and CAP is available. The latent-representation regressor and latent-representation-based classifier play an essential role in the testing phase, as shown on the right side of Fig. 3. Specifically, subjects with different types of features are first transformed into latent representation and then the diagnosis result can be obtained with the latent-representation-based classifier.

IV. EXPERIMENT AND RESULT

A. Experimental Setting

We conduct extensive experiments on the CT images data to evaluate the proposed pipeline. The dataset is randomly divided into 70% and 30% for training and testing, respectively. Furthermore, we adopt 5-fold cross-validation strategy on the training data to tune the parameter λ from the set {0.1, 1, 10, 100}. In practice, promising performance can be expected given a relatively large value for λ, which is fixed as 100 in our experiments.

Data Preprocessing: The original features extracted from CT images are of rather different scales. Accordingly, data preprocessing is necessary before using them as input for learning algorithm. There are several data preprocessing strategies, e.g., normalization and standardization. Specifically, the standardization for K features is computed as:

$$\hat{x^i} = \frac{x^i - \mu^i}{\sigma^i}, \quad i = 1, 2, \ldots, K \tag{8}$$

where μ^i and σ^i are mean value and standard deviation of the feature x^i, respectively. $\hat{x^i}$ denotes the standardization feature of original feature x^i. The features of normalization are calculated as:

$$\hat{x^i} = \frac{x^i - x^i_{min}}{x^i_{max} - x^i_{min}}, \quad i = 1, 2, \ldots, K \tag{9}$$

where x^i_{min} and x^i_{max} are minimum and maximum values of the feature x^i respectively. Accordingly, $\hat{x^i}$ indicates the normalization feature of original feature x^i.

We conduct experiments on several baseline models by concatenating all the original features, normalized features

TABLE III
EVALUATION OF PREPROCESSING OF FEATURES

Method		Original	Normalized	Standardized
LR	ACC	63.50±3.88	89.01±1.21	**89.19±1.21**
	SEN	1.00±0.00	90.64±1.04	90.66±2.08
	SPE	0.00±0.00	86.15±2.05	86.82±2.38
SVM	ACC	61.26±5.21	86.40±2.73	**89.40±1.21**
	SEN	1.00±0.00	90.51±2.31	89.90±1.92
	SPE	0.00±0.00	74.52±4.12	88.59±2.23
GNB	ACC	71.69±3.79	73.01±2.97	**75.60±2.13**
	SEN	96.42±1.98	89.32±2.14	85.22±2.75
	SPE	23.31±6.92	71.22±3.46	71.98±5.26
KNN	ACC	64.62±3.24	**87.10±2.03**	86.40±2.44
	SEN	1.00±0.00	89.44±2.78	89.22±3.35
	SPE	0.00±0.00	78.33±3.36	76.00±4.75
NN	ACC	70.43±3.23	92.31±1.87	**93.90±2.04**
	SEN	98.21±0.56	91.76±1.52	94.60±2.16
	SPE	21.79±4.82	82.01±3.11	91.70±2.53

and standardized features, respectively. The effects of the data preprocessing for diagnosis are shown in Table III. Specifically, the performance of using the original features is relatively low, the main reason of which may be the large scale difference among different features. Fortunately, in terms of accuracy, both of these preprocessing methods obtain a significant improvement (1.32% ∼ 25.69%) on all classification models. For clarification and comparison fairness, we employ the standardized data for all methods in the following experiments. We compare the proposed method with the following methods in the diagnosis task, including SVM, Logistic-Regression (LR), Gaussian-Naive-Bayes (GNB), K-Nearest-Neighbors (KNN), and Fully-Connected-Neural-Networks (NN). For all these methods, we repeat 10 times and report the mean and standard deviation performance. Diagnostic performance is evaluated in terms of accuracy (ACC), sensitivity (SEN) and specificity (SPC).

B. Performance Evaluation

1) Discrimination Power of These Different Types of Features: In order to investigate the discrimination power (related to diagnosis) of different types of features, we visualize them with t-distributed stochastic neighbor embedding (t-SNE) [35]. Fig. 5 demonstrates different distributions for these 7 types of features and concatenated features (7 types). Furthermore, to quantitatively evaluate these features, we conduct experiments on each type of features for diagnosis task with baseline algorithms. Table IV presents the diagnostic performance. First, we can find that large performance gaps exist between different types of features. For example, the baselines with gray features and texture features achieve clearly better performances than number features and intensity features. There are different manifestations reported between COVID-19 and other types of pneumonia, such as Influenza-A viral pneumonia [24]. As expected, radiomic features including gray and texture features have better discrimination ability. However, the number

TABLE IV

DIAGNOSIS PERFORMANCE WITH BASELINE METHODS USING DIFFERENT TYPES OF FEATURES

Method		Radiomic	Handcrafted	GF	TF	HF	NF	IF	SF	VF
SVM	ACC(%)	85.23±2.21	86.47±2.34	81.14±1.92	85.11±1.22	76.05±3.64	70.09±4.67	69.35±4.32	84.24±2.32	80.76±2.06
	SEN(%)	86.40±1.86	88.97±1.33	91.14±1.02	87.52±1.27	90.47±1.52	77.56±3.89	93.92±2.04	88.33±2.86	89.06±2.18
	SPC(%)	83.33±2.03	82.59±2.01	65.04±4.23	87.00±1.96	54.48±4.54	59.27±5.28	30.44±6.43	77.33±3.87	67.62±4.88
LR	ACC(%)	87.33±2.18	89.19±2.22	81.56±2.67	84.91±3.02	69.83±3.98	69.45±4.32	70.94±3.77	81.56±2.57	78.21±2.41
	SEN(%)	93.41±1.19	88.21±2.31	91.54±1.67	89.21±1.92	92.72±1.02	71.64±3.01	95.08±1.17	86.76±2.51	86.70±2.33
	SPC(%)	77.33±3.63	90.77±1.22	65.53±4.02	77.83±4.33	33.33±5.81	65.84±4.55	27.22±6.32	72.58±3.08	66.06±2.32
GNB	ACC(%)	79.30±2.84	75.04±2.44	72.30±3.28	73.50±3.88	71.30±3.52	64.00±3.85	68.70±4.02	81.20±2.98	66.60±4.23
	SEN(%)	83.39±2.14	83.44±2.46	88.23±1.64	76.78±2.65	83.32±2.33	71.82±3.17	92.33±2.01	89.93±1.87	75.43±2.89
	SPC(%)	70.66±3.32	70.02±3.21	77.72±2.66	80.86±2.15	61.12±3.98	46.48±5.61	23.88±7.33	70.25±2.83	74.34±2.21
KNN	ACC(%)	82.30±3.06	86.50±2.55	81.20±3.12	82.60±2.89	75.70±3.67	67.90±4.27	62.50±3.84	80.40±2.75	79.30±3.64
	SEN(%)	85.33±2.24	86.23±2.31	89.32±1.75	84.23±2.61	88.32±2.16	70.87±3.11	88.22±2.07	84.02±2.36	80.39±2.84
	SPC(%)	76.82±2.83	72.33±3.11	78.23±3.12	83.22±2.11	60.32±4.38	50.75±3.61	43.65±3.32	73.89±2.91	67.33±4.22
NN	ACC(%)	87.60±1.77	89.41±1.21	81.62±1.42	87.23±1.96	78.33±2.13	70.96±2.53	70.81±2.12	84.33±1.72	84.91±1.36
	SEN(%)	90.11±1.03	91.16±1.25	86.62±1.86	88.63±1.77	92.24±1.32	72.55±2.85	95.42±1.02	87.62±1.27	88.55±1.83
	SPC(%)	83.06±2.21	87.33±1.89	70.91±2.63	84.12±1.63	52.76±3.26	68.43±3.07	28.03±4.48	79.25±2.53	76.31±3.16

Note: GF, TF, HF, NF, IF, SF, VF represent gray features, texture features, histogram features, number features, intensity features, surface features and volume features, respectively. LR, GNB and NN are shorts for Logistic-Regression, Gaussian-Naive-Bayes algorithm and Fully-Connected-Neural-Networks, respectively.

Gray features Texture features Histogram features Volume features

Number features Intensity features Surface features Concatenated features
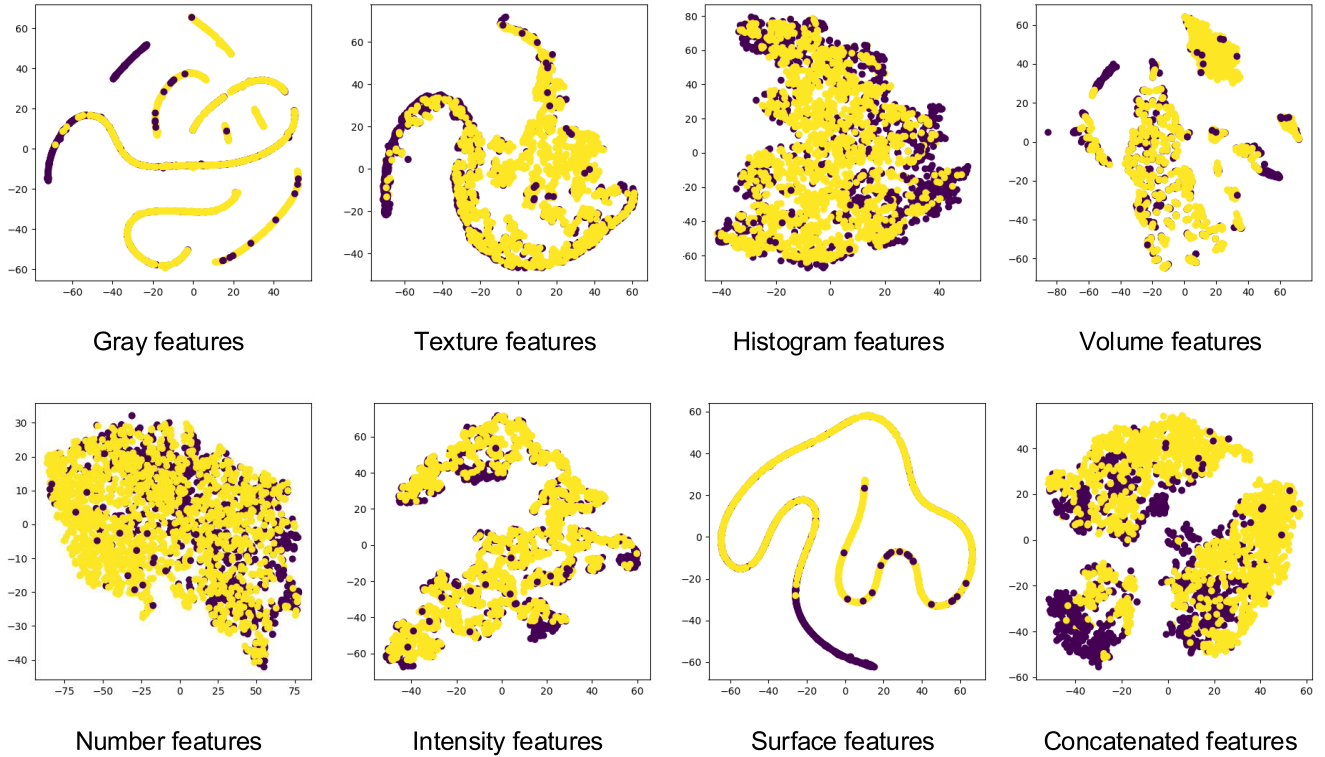
Fig. 5. Visualization of each type of original features and concatenated features using t-distributed stochastic neighbor embedding (t-SNE) [35], which is a nonlinear dimensionality reduction technique well-suited for embedding high-dimensional data for visualization in a low-dimensional space.

features and intensity features are a little less discriminative, and the possible reason is that the number of lesions and the intensity in lung may be not quite different for COVID-19 and CAP. Note that, although different types of features have different power in diagnosis, they are complementary to each other. As shown in Table IV, the concatenated features (i.e., radiomic and handcrafted features) perform much better than the case of using each individual type of features in terms of accuracy, which strongly supports the necessity of jointly using different types of features.

2) Effectiveness of Latent Representation Compared With the Original Features: With latent-representation regressor, Fig. 4 intuitively demonstrates the effectiveness of the learned latent representation with informativeness and structure compared with the original features. Specifically, according to Fig. 4(a), it is observed that the original (concatenated) features are not well structured, while the learned latent representations (Fig. 4(b)) encoding information from original features and class labels can better reveal the underlying class structure. As expected, the counterparts Fig. 4(c) and Fig. 4(d) in the
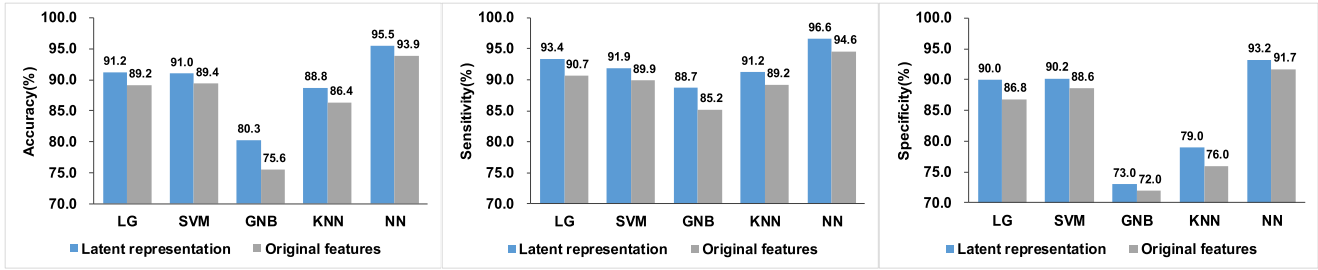
Fig. 6. Evaluation for the latent representation on different classifiers.

Fig. 7. Stability of our method with different ratios of training data.

testing stage further validate the generalization ability of our model. Therefore, promising performance in diagnosis can be expected by using the learned latent representation.

For quantitatively evaluation, as shown in Fig. 6, both conventional learning models and neural networks achieve significant improvement with the learned latent representation in terms of all three metrics. Specifically, Gaussian-Naive-Bayes obtains a clear performance improvement (4.72% and 3.44%) in terms of accuracy and sensitivity, respectively, while Logistic-Regression improves the performance by 3.13% in terms of specificity. This further validates that the learned latent representation is effective in diagnosis for COVID-19 and CAP.

3) Comparison for Different Methods: Fig. 6 also shows diagnosis performance of the proposed method and compared methods. It is observed that the proposed method achieves the best accuracy, up to 95.50%. Compared to all baselines which directly learn projections from the original features into class labels, our latent-representation-based model improves the diagnosis accuracy by 6.1% ~ 19.9%. In terms of sensitivity and specificity, our method also demonstrates the best performance, improving the performance by 4.61% ~ 21.22% compared to the comparison methods. To further investigate the effectiveness of our latent representation, we compare different diagnosis models by using the original features and our latent representation as shown in Fig. 6. We can find that consistent better performances are achieved by using the latent representation for all classifiers. Furthermore, neural network with original features achieves promising performance, while the same structure neural networks using our latent representation achieve higher performance in terms of all metrics. This further validates the advantage and potential of the latent representation learned from our pipeline.

4) Stability of Proposed Method: We conduct experiments to verify the stability of our proposed model under different proportions of training data. For fair comparison, the testing set is fixed in each experiment. Fig. 7 reveals the fluctuation of performance as the ratio of training data changing from 2% to 80%. We can find that the performance becomes clearly better as the number of training samples increases. However, when the ratio of training set exceeds 40%, the stability of performance could be observed, which also reflects the law of diminishing returns. For example, when 60% of data are used, the model achieves the best results on all three metrics. While, the worst performance is only about 1% lower than the best. Therefore, the promising performance and stable training results empirically validate that the proposed method can accurately and stably identify COVID-19 from CAP.

V. CONCLUSION AND FUTURE WORK

In this study, we proposed a novel automatic diagnosis pipeline for COVID-19 which can fully leverage different types of features extracted from CT images. We investigated these different types of features and found they are complementary to each other. Then, with the proposed multi-view representation learning technique, diagnosis performance is promoted to 95.5%, 96.6% and 93.2% in terms of accuracy, sensitivity and specificity, respectively. More importantly, compared to the original features, the learned latent representation has potential for utilization in different classifiers. In the future, we will consider diagnosis with more classes (i.e., normal, different COVID-19 severity, and CAP) instead of only two types of disease (i.e., COVID-19 and CAP). Moreover, clinical characteristics for patients might be beneficial for diagnosis, which can be flexibly integrated into our framework for performance promotion.

REFERENCES

[1] J. T. Wu, K. Leung, and G. M. Leung, "Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in wuhan, China: A modelling study," *Lancet*, vol. 395, no. 10225, pp. 689–697, Feb. 2020.

[2] H. Shi *et al.*, "Radiological findings from 81 patients with COVID-19 pneumonia in wuhan, China: A descriptive study," *Lancet Infectious Diseases*, vol. 20, no. 4, pp. 425–434, Apr. 2020.

[3] F. Song *et al.*, "Emerging 2019 novel coronavirus (2019-nCoV) pneumonia," *Radiology*, vol. 295, no. 1, pp. 210–217, Apr. 2020.

[4] World Health Organization *et al.*, "Coronavirus disease 2019 (COVID-19): Situation report, 67," World Health Org., Geneva, Switzerland, Tech. Rep. 67, 2020.

[5] Z. Xu *et al.*, "Pathological findings of COVID-19 associated with acute respiratory distress syndrome," *Lancet Respiratory Med.*, vol. 8, no. 4, pp. 420–422, Apr. 2020.

[6] Y. Li and L. Xia, "Coronavirus disease 2019 (COVID-19): Role of chest CT in diagnosis and management," *Amer. J. Roentgenol.*, pp. 1–7, Mar. 2020, doi: 10.2214/AJR.20.22954.

[7] Y. Fang *et al.*, "Sensitivity of chest CT for COVID-19: Comparison to RT-PCR," *Radiology*, Feb. 2020, Art. no. 200432.

[8] F. Shi *et al.*, "Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for COVID-19," *IEEE Rev. Biomed. Eng.*, early access, doi: 10.1109/RBME.2020.2987975.

[9] J. P. Kanne, "Chest CT findings in 2019 novel coronavirus (2019-nCoV) infections from wuhan, China: Key points for the radiologist," *Radiology*, vol. 295, no. 1, pp. 16–17, Apr. 2020.

[10] A. Bernheim *et al.*, "Chest CT findings in coronavirus disease-19 (COVID-19): Relationship to duration of infection," *Radiology*, Feb. 2020, Art. no. 200463.

[11] J. P. Kanne, B. P. Little, J. H. Chung, B. M. Elicker, and L. H. Ketai, "Essentials for radiologists on COVID-19: An update—Radiology scientific expert panel," Tech. Rep., 2020.

[12] W. Zhao, Z. Zhong, X. Xie, Q. Yu, and J. Liu, "Relation between chest CT findings and clinical conditions of coronavirus disease (COVID-19) pneumonia: A multicenter study," *Amer. J. Roentgenol.*, vol. 214, no. 5, pp. 1072–1077, May 2020, doi: 10.2214/AJR.20.22976.

[13] M. Chung *et al.*, "CT imaging features of 2019 novel coronavirus (2019-nCoV)," *Radiology*, vol. 295, no. 1, pp. 202–207, Apr. 2020.

[14] J. R. Zech, M. A. Badgeley, M. Liu, A. B. Costa, J. J. Titano, and E. K. Oermann, "Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study," *PLOS Med.*, vol. 15, no. 11, 2018, Art. no. e1002683.

[15] D. S. Kermany *et al.*, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131.e9, Feb. 2018.

[16] P. Rajpurkar *et al.*, "Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists," *PLOS Med.*, vol. 15, no. 11, 2018, Art. no. e1002686.

[17] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[18] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[19] Y. Mori, S.-E. Kudo, T. Berzin, M. Misawa, and K. Takeda, "Computer-aided diagnosis for colonoscopy," *Endoscopy*, vol. 49, no. 08, pp. 813–819, Aug. 2017.

[20] R. Zhang *et al.*, "Automatic detection and classification of colorectal polyps by transferring low-level CNN features from nonmedical domain," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 1, pp. 41–47, Jan. 2017.

[21] M. Liu, D. Zhang, E. Adeli, and D. Shen, "Inherent structure-based multiview learning with multitemplate feature representation for Alzheimer's disease diagnosis," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1473–1482, Jul. 2016.

[22] M. Zhang, Y. Yang, F. Shen, H. Zhang, and Y. Wang, "Multi-view feature selection and classification for Alzheimer's Disease diagnosis," *Multimedia Tools Appl.*, vol. 76, no. 8, pp. 10761–10775, 2017.

[23] S. Wang *et al.*, "A deep learning algorithm using ct images to screen for corona virus disease (covid-19)," *medRxiv*, 2020.

[24] X. Xu *et al.*, "Deep learning system to screen coronavirus disease 2019 pneumonia," 2020, *arXiv:2002.09334*. [Online]. Available: http://arxiv.org/abs/2002.09334

[25] K. R. Gray, P. Aljabar, R. A. Heckemann, A. Hammers, and D. Rueckert, "Random forest-based similarity measures for multi-modal classification of Alzheimer's disease," *NeuroImage*, vol. 65, pp. 167–175, Jan. 2013.

[26] C. Zhang *et al.*, "CPM-Nets: Cross partial multi-view networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 557–567.

[27] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. Fourth Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 565–571.

[28] F. Shan *et al.*, "Lung infection quantification of COVID-19 in CT images with deep learning," 2020, *arXiv:2003.04655*. [Online]. Available: http://arxiv.org/abs/2003.04655

[29] A. Zwanenburg, S. Leger, M. Vallières, and S. Löck, "Image biomarker standardisation initiative," 2016, *arXiv:1612.07003*. [Online]. Available: http://arxiv.org/abs/1612.07003

[30] F. Shi *et al.*, "Large-scale screening of COVID-19 from community acquired pneumonia using infection size-aware classification," 2020, *arXiv:2003.09860*. [Online]. Available: http://arxiv.org/abs/2003.09860

[31] C. Xu, D. Tao, and C. Xu, "A survey on multi-view learning," 2013, *arXiv:1304.5634*. [Online]. Available: http://arxiv.org/abs/1304.5634

[32] C. Zhang, H. Fu, Q. Hu, P. Zhu, and X. Cao, "Flexible multi-view dimensionality co-reduction," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 648–659, Feb. 2017.

[33] C. Zhang, Q. Hu, H. Fu, P. Zhu, and X. Cao, "Latent multi-view subspace clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4279–4287.

[34] T. Sing Lee, "Image representation using 2D Gabor wavelets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 10, pp. 959–971, Oct. 1996.

[35] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.