The Institution of Engineering and Technology | WILEY

ORIGINAL RESEARCH

DCACorrCapsNet: A deep channel-attention correlative capsule network for COVID-19 detection based on multi-source medical images

Aite Zhao (ID) | **Huimin Wu** | **Ming Chen** | **Nana Wang**

College of Computer Science and Technology, Qingdao University, Qingdao, Shandong, China

Correspondence
Aite Zhao, College of Computer Science and Technology, Qingdao University, Qingdao, Shandong, China.
Email: zhaoaite@qdu.edu.cn

Abstract

The raging trend of COVID-19 in the world has become more and more serious since 2019, causing large-scale human deaths and affecting production and life. Generally speaking, the methods of detecting COVID-19 mainly include the evaluation of human disease characterization, clinical examination and medical imaging. Among them, CT and X-ray screening is conducive to doctors and patients' families to observe and diagnose the severity and development of the COVID-19 more intuitively. Manual diagnosis of medical images leads to low the efficiency, and long-term tired gaze will decline the diagnosis accuracy. Therefore, a fully automated method is needed to assist processing and analysing medical images. Deep learning methods can rapidly help differentiate COVID-19 from other pneumonia-related diseases or healthy subjects. However, due to the limited labelled images and the monotony of models and data, the learning results are biased, resulting in inaccurate auxiliary diagnosis. To address these issues, a hybrid model: deep channel-attention correlative capsule network, for channel-attention based spatial feature extraction, correlative feature extraction, and fused feature classification is proposed. Experiments are validated on X-ray and CT image datasets, and the results outperform a large number of existing state-of-the-art studies.

1 | INTRODUCTION

COVID-19 refers to pneumonia caused by the 2019 novel coronavirus infection. Since December 2019, acute respiratory infectious diseases caused by novel coronavirus infection have become a global pandemic [1, 2]. The main clinical manifestations of novel coronavirus pneumonia are fever, dry cough, and fatigue. A small number of patients are accompanied by upper respiratory and gastrointestinal symptoms such as nasal congestion, runny nose, and diarrhoea [3–5].

COVID-19 can be detected by routine sputum examination, nasopharyngeal swab, chest CT and chest fluoroscopy. Compared with other primary detection methods, chest CT shows strong advantages. Direct tomography of the lung produces more CT images, and the structure of each image shows in different angles, which significantly displays the infection status and development trend of the lung. However, X-ray examination is superior to CT examination in cost, efficiency and speed, and has become a more popular way of detection [6].

The chest X-rays of early COVID-19 are mostly patchy shadows, which are obvious in the extrapulmonary band. If the COVID-19 nucleic acid test is positive, it is generally recommended that the patient detect the chest CT, because the chest X-ray has an anterior-posterior projection. However, it is easy for some lesions to be blocked by the shadow of the ribs, including the intercostal muscle, diaphragm, and heart shadow. This is one of the reasons why the diagnostic accuracy of chest X-ray is lower than that of CT. In the literature using X-ray for COVID-19 detection, deep learning and machine learning algorithms [5, 7–9] are used to extract features from X-ray images, and the probability of success in identifying COVID-19 patients is high, which verifies the feasibility of these schemes.

The CT findings of novel coronavirus pneumonia showed multiple small patches and interstitial changes at the early stage,

with obvious external lung involvement. Then it develops into multiple ground glass shadow and infiltration shadow of both lungs. In severe cases, lung consolidation can occur, and pleural effusion is rare. Lung CT is one of the conditions of clinical manifestations for the diagnosis of suspected cases of novel coronavirus pneumonia, and lung CT is also a very reliable indicator for dynamic observation of patients' condition changes. Even if the clinical symptoms are not too severe, lung CT will have characteristic manifestations of viral infection, so several experts advocate making lung CT to diagnose COVID-19. Several state-of-the-art approaches [10, 11] also analyse CT images of COVID-19, including CNN-RNN model, vision transformer, ResNet, U-Net, etc.

There are about 300 CT images of a COVID-19 patient, and the time required to analyse the CT images of the case is about 5–15 min, which brings a lot of pressure to the doctor's clinical diagnosis. Therefore, many deep learning methods have been applied to the field of chest image analysis for COVID-19. The most used are automated diagnostic methods based on convolutional neural networks (CNNs) [8, 12–14]. It takes an average of less than 20 s to identify a case, which greatly improves the efficiency of diagnosis and reduces the pressure on doctors. In addition, these neural networks can directly calculate the proportion of diseased parts, and then evaluate the severity of the disease, which greatly improves the efficiency of clinical diagnosis.

Although the use of deep learning models can improve detection efficiency, the lack of open source COVID-19 image data, low image resolution, similarities between images of different pneumonias, and limitations of algorithms can hinder the accuracy of pre-diagnosis results. For instance, the spatial features extracted using convolutional neural networks ignore the structural representation of the image itself, and do not consider the correlation information between convolutional features, and it is difficult to deal with dense numerical features. Temporal models such as long and short-term memory networks [15–19] only analyse the temporal information of the image and are not enough to fully express the spatial features of the image, but do not focus on the key points of the image, such as ground glass opacity (GGO) and other regions of interest (RoI). Additionally, some commonly used machine learning algorithms, support vector machine (SVM) [20–22], naive Bayes (NB) [23], AdaBoost[24], gradient boosting decision tree (GBDT) [25], are difficult to deal with sparse category features, cannot be updated online, and cannot handle large-scale data. Therefore, an autonomous learning method is required to solve the challenges brought by the above problems.

In this paper, as an assistive tool in diagnosis, a hybrid model is proposed for COVID-19 pneumonia detection by analysing X-Ray and CT images of the lung. There are three modules in this model, that is, a multi-feature extractor, a multi-level capsules, and a classification layer. The multi-feature extractor performs data augmentation, Fisher vector generation in time- and frequency- domain, and channel-attention based convolutional feature extraction of the input data. The multi-level capsules include three capsule layers: primary capsules, correlative capsules, and digit capsules. In the classification layer,

fused Fisher vector, channel-attention based convolutional feature, correlative feature, and digit feature are connected together as a fused feature and input into the DeepGBM model. The DeepGBM Model includes two modules: CatNN mainly uses embedding technology to convert high-dimensional sparse vectors into dense vectors; GBDT2NN uses GBDT to divide the data into regions and cluster them into leaves, and then uses neural networks to approximate the tree structure output and distil and extend to multi-tree approximation neural network model.

The main contributions of our work are summarized as follows:

- For spatial feature refining of the COVID-19 medical images, that is, X-ray and CT images, we first develop a multi-feature extractor that can generate Fisher vector in time- and frequency- domain and channel-attention based convolutional feature. Since the number of X-ray and CT images is only a few hundred, we utilize rotation, translation, staggered transformation and colour transformation for data augmentation.
- For correlative feature extraction, a multi-level capsules is embedded with primary capsules, correlative capsules and digit capsules, which can learn and capture the correlation between primary capsules after the convolution operation. Through the dynamic routing algorithm and weight updating, the best matching digital capsules can be selected from the primary capsules.
- For sparse categorical features and dense numerical features extraction, we use the four outputs of the previous two modules as input, and use CatNN and GBDT2NN in DeepGBM to train separately to obtain the results of multi-feature fusion and achieve classification.

Afterward, recent work will introduce the application of machine learning and deep learning in the detection and diagnosis of COVID-19. Then, the overview and internal structure details of our proposed model are described in Section 3, and the experimental procedure, setup and comparison results are provided in Section 4.

2 | RELATED WORK

The outbreak and rapid spread of COVID-19 have made many experts and scholars focus on improving the efficiency of diagnosis and treatment of COVID-19. At the same time, the extensive application of automated learning algorithms such as machine learning and deep learning in various fields has also attracted the attention of these outstanding researchers. The sparks of the combination of medicine and computer science bloom in this cross-cutting field, and have made outstanding contributions to the detection, diagnosis, treatment and rehabilitation of COVID-19.

As the pioneering force of deep learning, machine learning undoubtedly shines in the research direction of image recognition and signal processing [26–28]. The X-ray and CT images of

COVID-19 collected using different sensors fit right into the application category of machine learning algorithms, and the combination of the two has proven to advance the research and treatment of COVID-19.

As early as 2020, support vector machines (SVM) were introduced into data analysis applications for COVID-19 [29], which was introduced to generate real-time prediction of confirmed, dead and recovered cases of COVID-19 [30], and applied in the process of analysing of COVID-19 chest X-ray images [31]. Subsequently, random forest (RF) was adopted to classify COVID-19 CT images due to the advantages of ensemble strategy and high classification accuracy [32]. In 2021, several integrated machine learning algorithms entered researchers' attention. The fusion algorithm of AdaBoost and Random Forest was used to predict the age of death of COVID-19. Then, traditional algorithms such as random selection, principal component analysis, singular value decomposition, etc., were also used to speed up the training process [33]. XGBoost was paired with DenseNet169 deep neural network to extract features from COVID-19 chest X-ray images and detect the COVID-19 patient [34].

In addition to the excellent results brought by the above-mentioned machine learning algorithms, deep learning models are emerging one after another, and the collaborative operation with traditional algorithms simplifies the process of pneumonia detection and diagnosis. Integrating CNN with sparse autoencoder and feed forward neural network, a diagnosis system was established to extract features from X-ray images, and distinguish COVID-19 and non-COVID-19 cases [35]. Moreover, two lightweight least explored CNN models, namely SqueezeNet and ShuffleNet were evaluated with CT scan images for the diagnosis of COVID-19 [36]. In the work of [37], four fully automatic approaches was presented for the classification of chest X-ray images under the analysis of COVID-19, pneumonia and healthy controls, which compared the results of DenseNet-121, DenseNet-161, ResNet-18, ResNet-34, VGG-12, VGG-19, and other state-of-the-art studies. Furthermore, a new fusion method combined the outputs of multiple state-of-the-art CNN to build an optimal integration to determine which basic learners should be aggregated, which can recognize COVID-19 infected patients via chest X-ray images [38].

Although the above approaches have mature research on the analysis and processing of CT and X-ray images, there is still no method that can simultaneously consider the structural information, spatial features and attention-directed key localization of these lung images, and model inter-layer correlations. Therefore, we develop a deep channel-attention correlative capsule network (DCACorrCapsNet) to solve these problems.

3 | DEEP CHANNEL-ATTENTION CORRELATIVE CAPSULE NETWORK

In this section, we first describe how we construct the framework and then discuss the details of the internal modules, and then we explain the calculation process and parameter meaning of each module, and summarize the performance of the overall model at the end of this section.

3.1 | Model introduction

To successfully identify medical images of the lungs of COVID-19 patients, we propose a method that fuses image augmentation, attention mechanism and hybrid multi-level feature extraction and classification.

The structure of the proposed model DCACorrCapsNet is illustrated in Figure 1. In this work, the CT or X-ray data are first input into the multi-feature extractor, and the data is first used to expand the data. Then, the batch-generated images are input into Fisher vector generator for feature analysis in time domain and frequency domain respectively. At the same time, these images are extracted with convolutional features. After being processed by a channel attention module, the convolutional features have the attentional features and are fused with the Fisher vector to generate the final low-level spatial features. The convolution graph after the convolution operation is input into the multi-level capsules for capsule layer processing, which is divided into primary capsule layer, correlation capsule layer and digital capsule layer. The primary capsule is only processed for the convolutional graph, while the correlation capsule layer and the digital capsule layer derive high-level features by learning the correlation between capsules and selecting the best matching capsule output. Finally, these output features are fed into the classification layer for final fusion feature analysis and classification to obtain accurate sample labels.

First, we formulate the problem for COVID-19 detection. The augmented sample sequences are defined as $X = \{x_i \in \mathbb{R}^{W*H}, i = 1, 2, 3, \dots, N\}$ with corresponding 3-class label sequences L, N is the sample numbers of the input data, W and H are the width and height of each frame of the image.

3.2 | Multi-feature extractor

3.2.1 | Data augmentation

To prevent over-fitting and generate a more robust model, we use the ImageDataGenerator function in the Keras framework to expand the source image data to 3k images in each class. The converted data is shown in Figure 2. In the X-ray and CT datasets used, there are only a few hundred images per classification, while the most popular advanced neural networks require thousands of image data to help with parameter training, and large datasets are a guarantee of good results. In different task contexts, we can use geometric transformations of images, such as rotation, reflection transformation, flip transformation, scaling transformation, translation transformation, and colour transformation, using various combined transforms to increase the amount of input data.
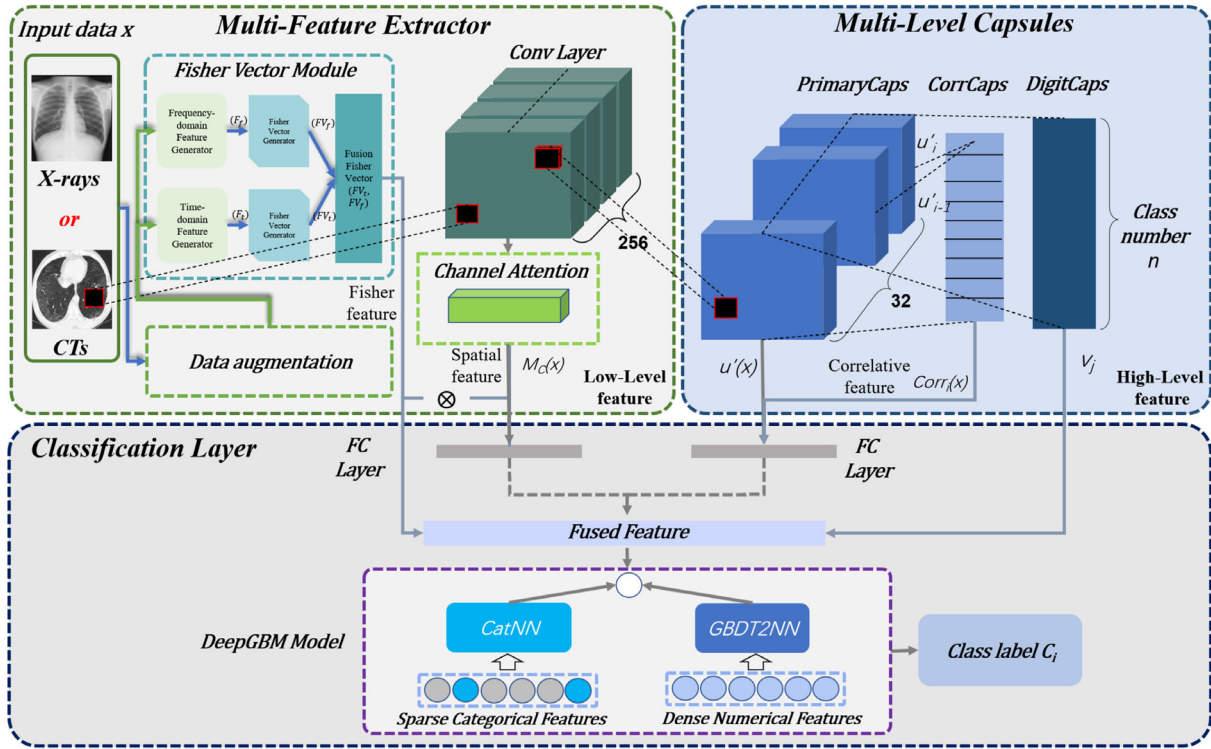
FIGURE 1   The framework of DCACorrCapsNet. It includes three components: multi-feature extractor, multi-level capsules, and classification layer
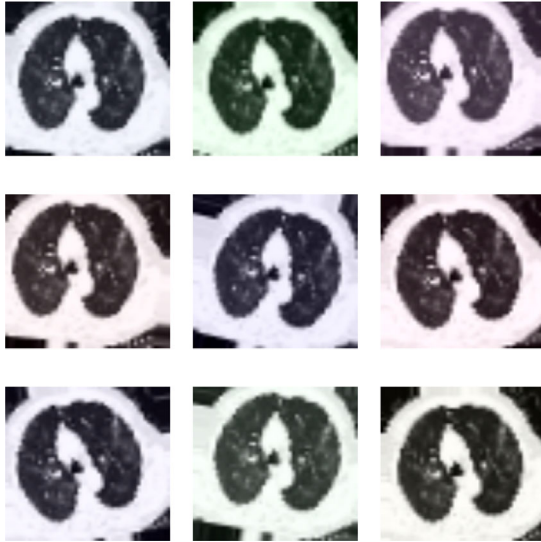

FIGURE 2   The augmented CT image

3.2.2 | Fisher vector module

In this module, we aim to develop a fisher vector generator to analyse the input image data. After having received the augmented data, we first normalize all the data by $L2$ regularization to generalize the statistical distribution of uniform samples. The normalized data is processed separately to extract descriptors in time- and frequency- domain as the input of the Fisher vector generator (Figure 3).
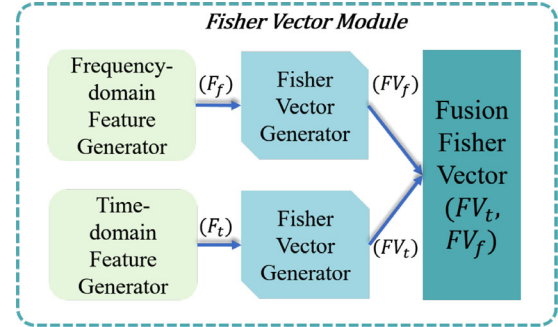

FIGURE 3   The Fisher vector generator. It includes frequency- and time-domain Fisher feature generators

To extract the time- and frequency-domain information of the image, two observations are made. Formally, given a image sequence $X = \{x_i \in \mathbb{R}^{W*H*C}, i = 1, 2, 3, \dots, N\}$, we employ the two-domain feature generators, that is, time- and frequency-domain generators. The time-domain generator calculates statistics including waveform factor, mean, skewness, root mean square, kurtosis, peak factor, margin factor and impulse factor. The frequency-domain generator adopts discrete Fourier transform to learn the frequency features of the images, which displays the data dynamics.

After having obtained the descriptors from the two-domain generators, firstly, each training sample of the input x_i is divided into W H-dimensional vectors $x_t \in \mathbb{R}^H, t = 1, \dots, W$, which are learned automatically. To generate the descriptors of the medical images, $V_j \in \mathbb{R}^{H \times H} (j = 1, 2)$ is a description

parameter matrix that can be updated for different feature types. A GMM model is used to fit all input descriptors, which employs multiple weighted Gaussian density functions to describe the distribution of the data and preserves the most frequent patterns of the original dataset. The probability of each class can be obtained from the projected samples, which can better visualize CT or X-ray images. Assuming that the main parameters of the Gaussian distribution includes weight, mean and variance ($\lambda = \{w_k, \mu_k^b, \sigma_k^b, k = 1, 2, \dots, K\}$, $h = 1, 2, 3, \dots, H$) refers to the hth dimension of $V_j x_t$. We assume that the covariance matrix is diagonal, and its diagonal elements reflect the spread of the data across dimensions. According to the standard Bayesian model, if the descriptor $V_j x_t$ fits the ith Gaussian model, the obtained probability is $\gamma_t(i) = u(i|V_j x_t, \lambda) = w_i u_i(V_j x_t|\lambda)/(\sum_{k=1}^{K} w_k u_k(V_j x_t|\lambda))$. The probability function of GMM can be calculated by:

$$
\begin{aligned}
u(V_j x_t|\lambda) &= \sum_{k=1}^{K} w_k * u_k(V_j x_t|\lambda) \\
&= \sum_{k=1}^{K} w_k * \frac{e^{-\frac{1}{2}(V_j x_t - \mu_k)' \sigma_k^{(-1)}(V_j x_t - \mu_k)}}{(2\pi)^{(\frac{H}{2})} * |\sigma_k|^{\frac{1}{2}}}
\end{aligned}
\tag{1}
$$

In frequency-domain, we denote a parameter $V_1 = \{V_{ab} = (e^{-i\frac{2\pi}{H}})^{a \times b}, a = 1, 2 \dots, H, b = 1, 2 \dots, H\}$ to represent the data, and process it by using Euler's formula as: $V_1 = \{V_{ab} = (\cos\frac{2\pi}{H} - i\sin\frac{2\pi}{H})^{a \times b}, a = 1, 2 \dots, H, b = 1, 2 \dots, H\}$ in which the real and imaginary parts are separated. Similarly, we can also define $V_2 \in \mathbb{R}^{H \times H}$ as the corresponding description parameter to indicate the time-domain feature.

Assuming that $V_j x_t \in \mathbb{R}^H, t = 1, \dots, W$ has a distribution of p, and when the sample number increases, the sample mean converges to the expectation $E_{x \sim p}$. We decompose the Gaussian distribution p into two parts to reduce or eliminate the influence of weak interference and noise on the image: The samples in the stronger part of the signal change have the distribution v. The noise (or weak signal of the image) follows a distribution u_λ. $0 <= w <= 1$ is defined as the ratio of strong signal of the image against the input. The probability of this modelling process is provided as: $p(x) = wv(x) + (1 - w)u_\lambda(x)$.

The Fisher vector G_λ^X is obtained by computing the partial derivatives of the likelihood function $L(X|\lambda) = E_{x \sim p}[\log v(V_j x_t|\lambda)]$ to w_k, μ_k^b, σ_k^b, which can represent the gradient statistics of the local descriptors.

$$
\begin{aligned}
G_\lambda^X &= \nabla_\lambda E_{x \sim p}[\log u_\lambda(x)] \\
&= \nabla_\lambda \int_x p(x)\log u_\lambda(x)dx \\
&= w\nabla_\lambda \int_x u(x)\log u_\lambda(x)dx + (1-w)\nabla_\lambda \int_x u_\lambda(x)\log u_\lambda(x)dx.
\end{aligned}
\tag{2}
$$

The parameter λ is obtained by maximizing the likelihood function in GMM modelling:

$$
\nabla_\lambda \int_x u_\lambda(x)\log u_\lambda(x)dx = \nabla_\lambda E_{x \sim u_\lambda}[\log u_\lambda(x)] \approx 0.
\tag{3}
$$

So Equation (2) is expressed as:

$$
\begin{aligned}
G_\lambda^X &= w\nabla_\lambda \int_x v(x)\log u_\lambda(x)dx \\
&= w\nabla_\lambda E_{x \sim v}[\log u_\lambda(x)]
\end{aligned}
\tag{4}
$$

Finally, calculate the expectations of the diagonal elements of the Fisher matrix for each of the three variables, Fishier vector is generated as: $\mathcal{G}_\lambda^X = \{G_1, \dots G_{(K*(2H+1)-1)}\} = F_\lambda^{-1/2} G_\lambda^X$, where $F_\lambda^{-1/2} = E_X(G_\lambda^X G_\lambda^{X'})$ denotes the expectation, G_λ^X denotes the partial derivatives of the corresponding variables. Then, we obtain the absolute value of each element in the three normalized partial derivative components ($\mathcal{G}_{w_k}^X = F_{w_k}^{-1/2} \frac{w E_{x \sim v}[\log u_\lambda(x)]}{\partial w_k}$, $\mathcal{G}_{\mu_k}^X = F_{\mu_k}^{-1/2} \frac{w E_{x \sim v}[\log u_\lambda(x)]}{\partial \mu_k^b}$, $\mathcal{G}_{\sigma_k}^X = F_{\sigma_k}^{-1/2} \frac{w E_{x \sim v}[\log u_\lambda(x)]}{\partial \sigma_k^b}$).

To emphasize the difference of the three vectors in the same space, we normalize the Fisher vector with L_2 group normalization, as shown in Equation (5). We perform feature augmentation by grouping the three vectors (derived gradients of different parameters), and then apply L_2 normalization to the augmented data. The fusion Fisher vector finally obtained by merging two Fisher vectors FV_t and FV_f in dual domain.

$$
FV_t(FV_f) = L_2\left(\frac{\mathcal{G}_{w_k}^X |\mathcal{G}_{w_k}^X|}{\sum_{k=1}^{K} |\mathcal{G}_{w_k}^X|}, \frac{\mathcal{G}_{\mu_k}^X |\mathcal{G}_{\mu_k}^X|}{\sum_{k=1}^{K} |\mathcal{G}_{\mu_k}^X|}, \frac{\mathcal{G}_{\sigma_k}^X |\mathcal{G}_{\sigma_k}^X|}{\sum_{k=1}^{K} |\mathcal{G}_{\sigma_k}^X|} \right)
\tag{5}
$$

3.2.3 | Channel attention module

After the data augmentation, in addition to generating the Fisher vector, we use 256 5×5 convolution kernels to convolve the input image, input it to the channel attention layer (Figure 4), extract the channel attention feature of the convolution feature, and combine it with the Fisher feature to generate channel-attention based spatial feature [39].

Firstly, the convolutional feature map is compressed in the spatial dimension to obtain a 1D vector. When compressing the spatial dimension of the input feature map, the model considers both average pooling and max pooling. After operating through the two pooling functions, a total of two 1D vectors can be obtained. Global average pooling has feedback for every pixel on the feature map, while global max pooling only has gradient feedback when the response is largest in the feature map during gradient back-propagation computation.
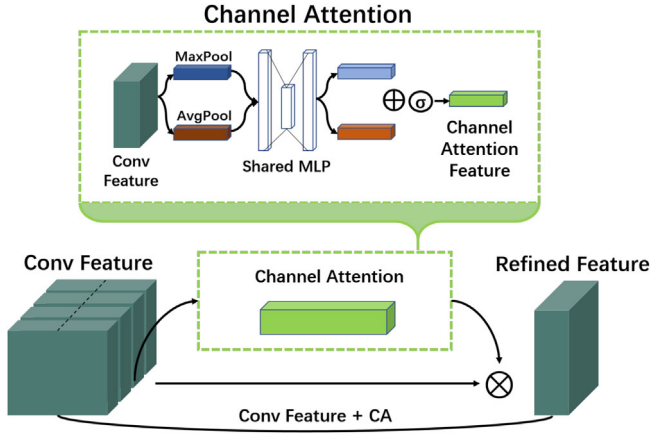
FIGURE 4 Channel attention. It exploits the correlation between feature channels to generate channel attention maps. To efficiently compute channel attention, we compress the spatial dimension of the input feature maps. Average pooling is often used to aggregate spatial information

Meanwhile, average pooling and max pooling are used to aggregate spatial dimension features to generate two spatial dimension descriptors: X_{max}^C and X_{avg}^C, and then pass through MLP (2-layer FC, activation, sigmoid, and squeeze operation) to generate weights for each channel. Finally, the weights are multiplied by the original channel attention. The formula is:

$$M_C(X) = \sigma(MLP(AvgPool(X)) + MLP(MaxPool(X)))$$
$$= \sigma(W_1(W_0(X_{avg}^C)) + W_1(W_0(X_{max}^C))) \quad (6)$$

where X represents the input feature map, X_{avg}^C and X_{max}^C represent the features calculated by global average pooling and global max pooling respectively, W_0 and W_1 represent the two-layer parameters in the multi-layer perceptron model, and the features between W_0 and W_1 in the multi-layer perceptron model need to be processed with ReLU as the activation function.

3.3 | Multi-level capsules

After having extracted features from the multi-feature extractor, we use three-layer capsules to obtain structural and correlation information of the images.

3.3.1 | Capsule layer

The primary capsule layer is computed as a stack of eight parallel convolutional layers, each with a 3D convolution kernel/capsule with stride 1, outputting a vector rather than a scalar. The input vector of the primary capsule layer is equivalent to the scalar input of the traditional neural network neuron, and the calculation of the vector is equivalent to the propagation and connection between the primary capsule and the digital capsule. The input vector computation can be divided into linear combination and dynamic routing. The linear combination and dynamic routing processes can be expressed by the following formula:

$$\begin{cases} \hat{u}'_{j|i} = W_{ij} u'_i, s_j = \sum_i c_{ij} \hat{u}'_{j|i}, \\ c_{ij} = \dfrac{\exp(b_{ij})}{\sum_k \exp(b_{ik})}. \end{cases} \quad (7)$$

where $\hat{u}'_{j|i}$ is a linear combination of neurons u'_i in the primary capsule layer. W_{ij} represents the corresponding weight vector of the ith capsule in this layer connecting to jth capsule in the next layer. s_j is the jth output of $\hat{u}'_{j|i}$ through the dynamic routing process by updating and finding the coupling coefficients c_{ij} iteratively. To calculate and update the c_{ij}, a parameter b_{ij} depends on the location and type of two former capsules is constantly updated with the consistency of the measurement. The prediction vector $\hat{u}'_{j|i}$ is multiplied by c_{ij} and passed to the digit capsule layer.

The input s_j of the different capsule units is the weighted sum of all possible incoming prediction tensors, and by putting this into the squashing nonlinear function, we get the output tensor v_j of the digit capsule layer. The product of v_j and the corresponding prediction vector $\hat{u}'_{j|i}$, summed with the parameter b_{ij}, jointly participates in the update of the coefficient c_{ij} without back-propagation, that is, $b_{ij} = b_{ij} + \hat{u}'_{j|i} * v_j$.

3.3.2 | Correlative capsule layer

For the primary capsule layer, all capsules are independent of each other, and the extracted features can not only represent the local features of the lung (mass, displacement, nodule, cavity, etc.), but also explain the correlation between local features and global features (the whole lung). However, due to the correlation between capsules is ignored, we design a correlative capsule layer to calculate the correlation between capsules:

$$Corr_i = u'_i * (u'_{i+1})^{-1}, \quad i < 32. \quad (8)$$

where $Corr_i$ is the correlation matrix of the capsules in the primary capsule layer, and u'_i and u'_{i+1} are adjacent capsules.

3.4 | Classification layer

We feed the four outputs of multi-feature extractor and multi-level capsules, as the fused feature, into the DeepGBM Model, which contains CatNN and GBDT2NN modules. CatNN mainly relies on embedding technology, which can effectively convert high-dimensional sparse vectors into dense vectors; GBDT2NN distils the trees in the GBDT model into multiple neural networks, which can well extract the features of dense numerical variables [40].

3.4.1 | CatNN

CatNN utilizes factor machine (FM) and feed-forward neural networks to learn interactions on features.

$$y_{Cat}(x) = y_{FM}(x) + y_{Deep}(x) \qquad (9)$$

where the FM component is used to learn linear features and pair-wise interactions, and the deep component is used to learn higher-order feature interactions.

3.4.2 | GBDT2NN

The tree model can extract the characteristics of dense numerical variables very well. If we define the function implemented by the tree model as $C^t(x)$, the role of GBDT2NN is to use a neural network to realize the function of $C^t(x)$. GBDT has multiple trees, and adopts leaf embedding distillation method and tree grouping method to reduce the number of leaves and the number of neural network models, respectively.

Leaf embedding distillation uses the information of the tree model itself for retraining. More specifically, since there is a mapping between leaf indices and leaf values, leaf values can be used to learn embeddings. Learning objectives can be expressed as:

$$\min_{\omega,\omega^0,\omega^t} \frac{1}{n} \sum_{i=1}^{n} L''(\omega^T H(L^{t,i}; \omega^t) + \omega^0, p^{t,i}) \qquad (10)$$

where $H(L^{t,i}; \omega^t)$ represents the representation of the one-hot leaf node index $L^{t,i}$ mapped to the embedding h through the parameter ω^t, $p^{t,i}$ represents the leaf node value of x^i, and L'' represents the loss function. The parameters ω and ω^0 represent the parameter matrix that transforms the embedding result into leaf node values.

In order to reduce the number of neural networks, the trees can be grouped, and then the neural network model is used to distil knowledge to the grouped trees. By utilizing random grouping, the objective function of embedding is extended to:

$$\min_{\omega,\omega^0,\omega^G} \frac{1}{n} \sum_{i=1}^{n} L''(\omega^T H(||_{t \in G}(L^{t,i}); \omega^G) + \omega^0, \sum_{t \in G} p^{t,i}) \quad (11)$$

where $||(...)$ represents concatenate operation, $H^{G,i} = H(||_{t \in G}(L^{t,i}); \omega^G)$ indicates that the result $(||_{t \in G}(L^{t,i}))$ after concatenate the leaf node index of one-hot in group G is mapped to the representation of embedding $H^{G,i}$ through the parameter ω^0. The objective function of the neural network is transformed into:

$$L^G = \min_{\theta^G} \frac{1}{n} \sum_{i=1}^{n} L(N(x^i[I^G]; \theta^G), H^{G,i}) \qquad (12)$$

FIGURE 5 The X-ray cases. From left to right are a patient with COVID-19, a healthy subject and a patient with ordinary pneumonia. The white blurred area indicated by the red arrow is ground-glass opacity (GGO), not found in the lung fields of the healthy subject
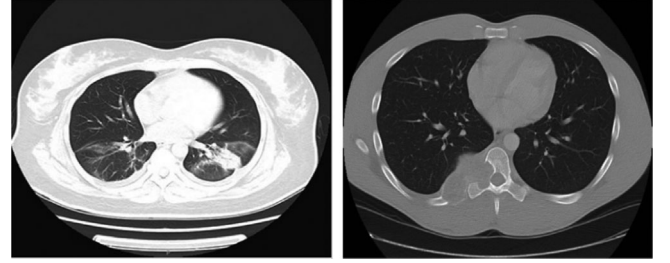

FIGURE 6 The chest CT cases. From left to right are a patient with COVID-19, and a healthy subject. We can see that the chest CT of the COVID-19 on the left has GGO shadows and the chest CT on the right is smooth, which can significantly detect the COVID-19

where I^G represents the feature used in the tree model grouping G. For a tree model with k groups, we get the output of group G and the output of GBDT2NN as:

$$\begin{cases} y_G(x) = \omega^T \times N(x[I^G, \theta^G]) + \omega_0, \\ y_{GBDT2NN}(x) = \sum_{j=1}^{k} y_{G_j}(x) \end{cases} \qquad (13)$$

4 | EXPERIMENT

We perform experiments on two datasets to evaluate the developed DCACorrCapsNet model for COVID-19 recognition. The compared methods are described as follows: GRU (long-short term memory), BiLSTM (bi-directional long-short term memory), CNN (convolutional neural network), DenseNet [41], EfficientNet [42], MobileNetV3 [43], ViT Transformer [44], Swin Transformer [45], Regnet [46], FixEfficientNet-L2 [47] and EfficientNet-V2 [48].

4.1 | Datasets

To avoid privacy concerns, we utilize two open source datasets, that is, X-ray and CT datasets. The acquisition of other X-ray and CT datasets used in other literature mentioned in this paper required more complex procedures and longer time

TABLE 1   Results of the ablation experiment on the X-ray dataset

X-ray dataset

Components						Evaluation			
Data Aug	Fisher vector	Conv+CA	Multi-feature extractor	Multi-level capsules	Deep GBM	Precision	Recall	F1 score	Acc
✓	✓					87.47%	86.93%	87.46%	86.93%
✓		✓				88.48%	88.01%	87.38%	88.01%
✓			✓			88.83%	88.69%	88.64%	88.69%
✓				✓		87.26%	87.12%	87.54%	87.12%
✓					✓	86.37%	86.49%	86.49%	86.49%
✓			✓	✓		89.37%	89.48%	89.94%	89.48%
✓			✓	✓	✓	*90.81%*	*90.43%*	*90.40%*	*90.43%*

TABLE 2   Performance of classification of COVID-19 patients, pneumonia patients and healthy subjects compared with state-of-the-art models on X-ray dataset

Classifier	DT	GBDT	LR	RF	KNN	NB
COVID-19	25.00%	60.00%	77.78%	100.00%	78.57%	29.73%
pneumonia	61.68%	73.68%	69.07%	73.08%	61.00%	63.92%
Healthy	70.93%	72.57%	70.09%	73.00%	71.72%	78.48%
Supervised models	CNN	BiLSTM	LSTM(Attn)	DenseNet121 [41]	EfficientNet-B0 [42]	MobileNetV3 [43]
COVID-19	76.92%	64.29%	78.57%	83.75%	86.49%	85.90%
pneumonia	66.34%	68.57%	51.00%	0.00%	50.00%	25.00%
Healthy	80.81%	85.11%	86.87%	89.19%	80.95%	88.89%
	ViT Transformer [44]	**Swin Transformer [45]**	**Regnet [46]**	**EfficientNet-V2 [48]**	**FixEfficientNet-L2 [47]**	**DCACorrCapsNet**
COVID-19	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
pneumonia	79.13%	80.56%	88.66%	77.19%	75.61%	94.38%
Healthy	82.98%	79.80%	83.96%	82.22%	87.34%	86.36%

for approval. The contents of the two datasets used are as follows:

4.1.1 | X-Ray dataset

The X-ray dataset was obtained from two different sources [49, 50] for COVID-19 detection, divided into three categories, that is, COVID-19 (125), Pneumonia (500) and No-Findings (500).

The dataset was constantly updated with images shared by researchers from different regions. Currently, there were 125 X-ray images diagnosed as positive for COVID-19 in the dataset, with 43 female and 82 male cases. The dataset exposed age information for 26 COVID-19 subjects, showing an average age of about 55. The specific performance characteristics are shown in Figure 5.

4.1.2 | CT dataset

The COVID-CT dataset has 349 CT images containing clinical manifestations of COVID-19 from 216 patients, and 397

CT images negative for COVID-19. Images were collected from COVID-19 related papers by medRxiv[1], bioRxiv[2], NEJM, JAMA, The Lancet, etc. The accuracy of diagnostic decisions may be reduced due to the lower quality of images acquired in the paper. Figure 6 shows one COVID-19 case and one healthy case from the dataset.

4.2 | Experimental settings

The experiments are conducted with Tensorflow and Python libraries using the system of GTX1050Ti GPU, i5-7300HQ CPU and 8G RAM. We will introduce the parameter settings for the three modules: multi-feature extractor, multi-level capsules and classification layer.

In the multi-feature extractor module, the rotation angle of the input image is 15°, the amplitude of the random horizontal offset and the random channel offset of the image is set to 0.05 and 10. The shear strength is 0.3 and the epsilon is 1×10^{-6}. We
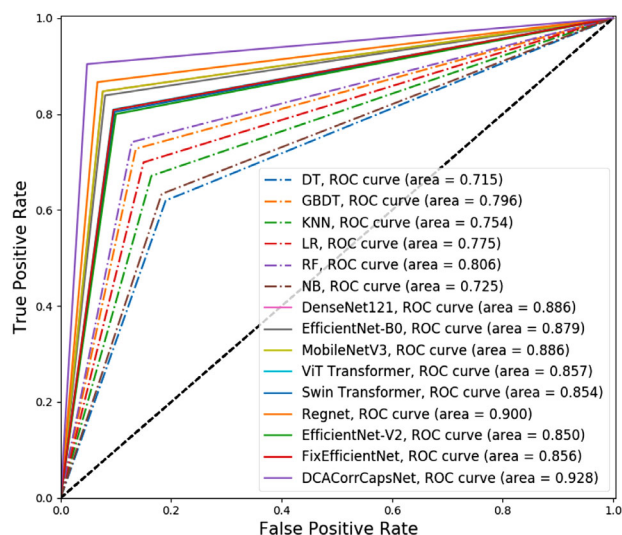
---

[1] https://www.medrxiv.org/
[2] https://www.biorxiv.org/

FIGURE 7 The ROC curve of COVID-19 detection on X-ray dataset

FIGURE 8 The confusion matrix of COVID-19 detection on X-ray dataset

totally generate 3000 images in each class by transformation to avoid over-fitting. In the Fisher vector module, the number of the Gaussian distributions of the GMM model is set to 20 for the data. The convolution layer uses 256 convolution kernels with the size of [5,5] to generate 256 feature maps. Afterward, 32 primary capsules with kernel size of [5,5] can be utilized for correlation computation and digit feature extraction, in the digit capsule layer, there are n (class number) groups of features generated for fused feature combination. In the classification layer, two FC layers are provided for processing the outputs of correlative feature and spatial feature. In DeepGBM model, most of the parameters are automatically learned during the training process and will not be explained here.

We use the fivefold cross validation method to validate the experiment. We divide the dataset into five parts. According to 20% of the data in each part, we randomly take four parts as the training set and one part as the testing set to train the model, find the appropriate model parameters, and prevent over-fitting.

4.3 | COVID-19 detection

In this section, to confirm the improvement over the proposed DCACorrCapsNet, two experiments are conducted for COVID-19 detection. Experiment I is designed to differentiate between 125 lung X-ray images of COVID-19 patients, 500 lung X-ray images of common pneumonia patients, and 500 lung X-ray images of healthy subjects. Experiment II aims to distinguish between 349 lung CT images of COVID-19 patients and 379 lung CT images of healthy subjects.

4.3.1 | COVID-19 detection on X-Ray dataset

In this experiment, COVID-19 (125), No-Findings (500), and Pneumonia (500) are recognized for COVID-19 Detection.
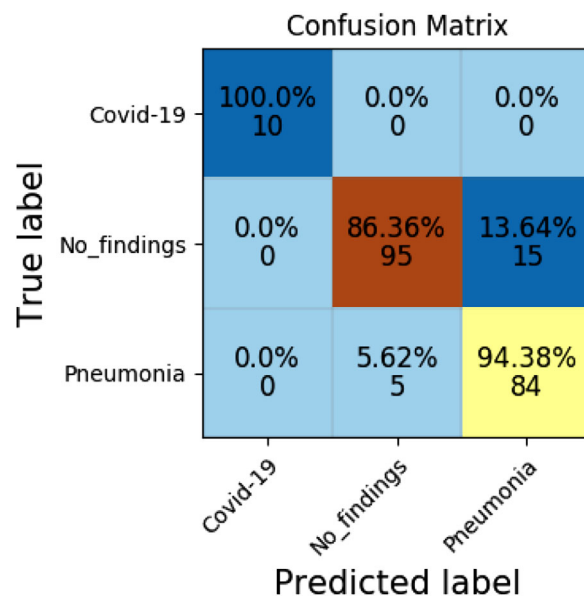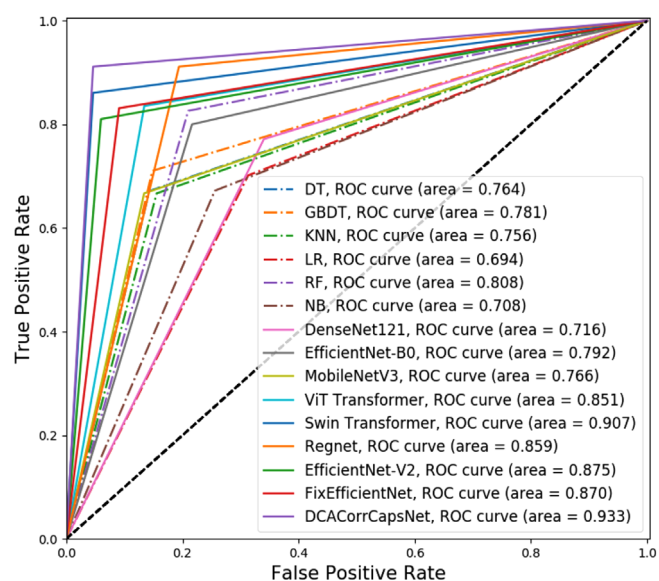

FIGURE 9 The ROC curve of COVID-19 detection on CT dataset

Since all datasets are augmented, the number of images in each class is 3000 to avoid over-fitting.

Ablation study on the composition module
Firstly, we split the hybrid model into six modules for evaluation on two datasets, and verified the feasibility of the fusion model by comparing the recognition results of multiple classifications and the performance of the matching modules. As illustrated in Table 1, seven typical combinations of these six modules are implemented to calculate four evaluation metrics, that is, precision, recall, F1 score, and accuracy, and we use "✓" to indicate the modules that have been selected. Compared

TABLE 3 Results of the ablation experiment on the CT dataset

CT dataset

Components						Evaluation			
Data Aug	Fisher Vector	Conv+CA	Multi-Feature Extractor	Multi-Level Capsules	DeepGBM	Precision	Recall	F1 score	Acc
✓	✓					88.26%	87.46%	88.64%	87.46%
✓		✓				89.06%	89.38%	88.29%	89.75%
✓			✓			90.83%	90.33%	90.90%	90.33%
✓				✓		89.96%	89.49%	88.94%	89.49%
✓					✓	90.29%	89.72%	90.74%	89.72%
✓		✓		✓		91.45%	91.01%	91.07%	91.01%
✓		✓		✓	✓	*93.23%*	*93.06%*	*93.07%*	*93.06%*

TABLE 4 Performance of classification of COVID-19 patients and healthy subjects compared with state-of-the-art models on CT dataset

Classifier	DT	GBDT	LR	RF	KNN	NB
COVID-19	73.33%	77.53%	74.03%	84.21%	79.17%	75.29%
Non COVID-19	81.67%	80.33%	64.39%	77.03%	74.07%	66.15%
Supervised models	**CNN**	**BiLSTM**	**LSTM(Attn)**	**DenseNet121 [41]**	**EfficientNet-B0 [42]**	**MobileNetV3 [43]**
COVID-19	77.78%	75.64%	77.75%	72.73%	78.35%	70.59%
Non COVID-19	73.91%	66.67%	69.50%	71.05%	80.00%	84.34%
	ViT Transformer [44]	**Swin Transformer [45]**	**Regnet [46]**	**EfficientNet-V2 [48]**	**FixEfficientNet-L2 [47]**	**DCACorrCapsNet**
COVID-19	81.69%	84.93%	88.52%	80.77%	82.19%	*89.86%*
Non COVID-19	88.00%	95.77%	84.71%	94.12%	91.43%	96.00%

with the two modules in the Multi-Feature Extractor, the performance of Fisher Vector is slightly inferior to Conv + CA, while the performance of the combined module of the two modules becomes better. The classification effect of Multi-Level Capsules is about 1% higher than DeepGBM. The classification accuracy of the fusion module of Multi-Level Capsules and Multi-Feature Extractor is higher than that of other separation modules, while the fusion performance of Multi-Level Capsules, Multi-Feature Extractor and DeepGBM reaches 90%.

Comparison with state-of-the-arts
In this section, 18 methods are validated on three categories, that is, COVID-19, No-Findings, and Pneumonia, including six mainstream classifiers, and 12 deep learning models. By comparing the accuracy of these models, it can be concluded that the proposed DCACorrCapsNet is the most stable and has the best average performance. The results are shown in Table 2. There are seven models with 100% detection rate for COVID-19 detection, that is, RF and six deep learning models. There are three models with a detection rate of more than 80% for common pneumonia, that is, Swin Transformer, Regnet, and our DCACorrCapsNet, and DCACorrCapsNet has the highest detection rate. It should be noted that DenseNet121 has over-fitting problem, and the detection rate for common pneumonia is 0%. When identifying healthy subjects, all the compared mod-

els exceeded 70%, and 11 models exceeded 80%, all of which are deep learning models. MobileNetV3 performs best excluding over-fitting factors, reaching 88.89% of the recognition rate.

The ROC of all methods used on the X-ray dataset is shown in Figure 7, where the coloured dotted line represents the classifier and the coloured solid line represents the deep learning methods in recent years. AUC represents the area under the ROC curve. The larger the AUC, the better the classification performance and the more stable it is. Overview of the ROC, it can be seen that there are fewer overlapping ROC lines, and the better the performance, the closer to the upper left corner, and completely wrap the line in the lower right corner. The dotted line is completely wrapped by the solid line, which proves that the deep learning network is better than the traditional classifier. The AUC of DenseNet121 is the same as that of MobileNetV3, and the AUC of FixEfficientNet-L2 and ViT Transformer is similar. As shown in Table 2, the recognition rate of Dense121 for ordinary pneumonia is 0%, indicating that there is over-fitting, and its ROC cannot be used as a reference. Here, Regnet and RF can be utilized as the best models for deep learning and classifiers, respectively, for COVID-19 detection, but they are still surpassed by our proposed scheme.

The classification confusion matrix of DCACorrCapsNet is shown in Figure 8. There is a data imbalance in the test set. There are a total of 209 test samples, including ten new crown
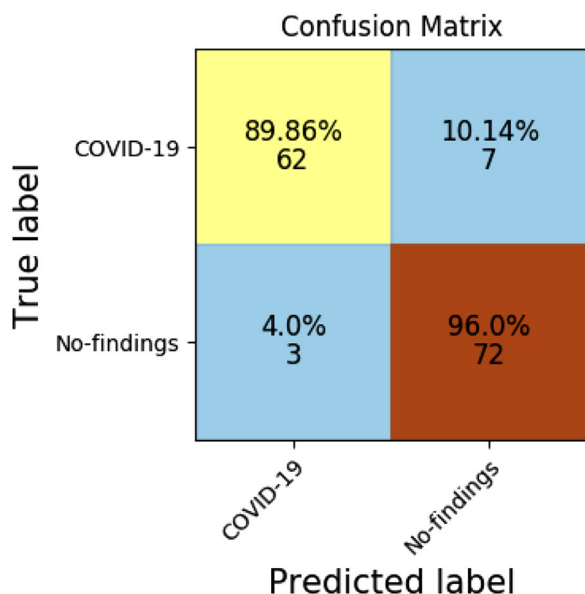
FIGURE 10 The confusion matrix of COVID-19 detection on CT dataset

samples, 89 ordinary pneumonia samples, and 110 healthy subjects samples. Among them, ten COVID-19 samples were all detected, while the samples of common pneumonia and healthy subjects had mutual detection errors, indicating that the proposed model was sensitive to the COVID-19 image recognition in this dataset. The probability of correct identification of all test samples is 90.43%, and the average accuracy is 93.58%.

4.3.2 | COVID-19 detection on CT dataset

In this experiment, COVID-19 (349) and non COVID-19 (379) are recognized for COVID-19 detection. Since all datasets are augmented, the number of images in each class is 3000 to avoid over-fitting.

Ablation study on the composition module
In order to judge the performance of the proposed model components on the CT dataset, we compared the four evaluation

accuracies of seven combinations of the six components, and the best result was obtained by the fusion model, which is shown in bold in the Table 3. By comparing all fusion modules, the effect of single model is obviously inferior to the accuracy of multi model fusion output, for example, Fisher Vector + Conv + CA (90.33%) >Fisher Vector (87.46%) / Conv + CA (89.75%), Multi-Feature Extractor + Multi-Level Capsules (91.01%) >Multi-Feature Extractor (90.33%) / Multi-Level Capsules (89.49%). Finally, the performance of the fusion model DCACorrCapsNet of the three modules achieves the highest accuracy.

Comparison with state-of-the-arts
To verify the superiority of DCACorrCapsNet, we compared the recognition accuracy of six machine learning classifiers and 12 deep learning methods. The performance is demonstrated in Table 4. The evidence shows that among all traditional classifiers, RF has the highest detection rate for COVID-19, and its excellent classification ability exceeds nine deep learning methods. In the deep learning model, our model and Regnet have the best classification ability. In contrast, the recognition rate of healthy subjects is 96%, and there is no over-fitting phenomenon, which verifies its superiority.

The ROC curves of all the compared methods on the CT dataset are demonstrated in Figure 9. The coloured dotted line represents the ROC curve of the traditional classifier, and the coloured solid line represents the ROC curve of the deep learning algorithm. The area AUC below them indicates the classification performance. It can be seen that many models have crossed ROC lines on the CT dataset, and RF as the best performing classifier outperforms some deep learning models, that is, DenseNet121, MobileNetV3 and EfficientNet-B0. The Swin Transformer and EfficientNet-V2 in the deep models are ranked second and third, respectively, and their performance is only inferior to our proposed DCACorrCapsNet. The AUC value of LR is less than 70%, because the treatment of nonlinear problems depends on features, cross features must be combined to discretize the features. It is proved that it is not suitable for this nonlinear separable dataset.

The confusion matrix of our proposed scheme is shown in Figure 10. COVID-19 samples and 75 healthy subjects were collected from 144 test samples. The samples were 69 COVID-19
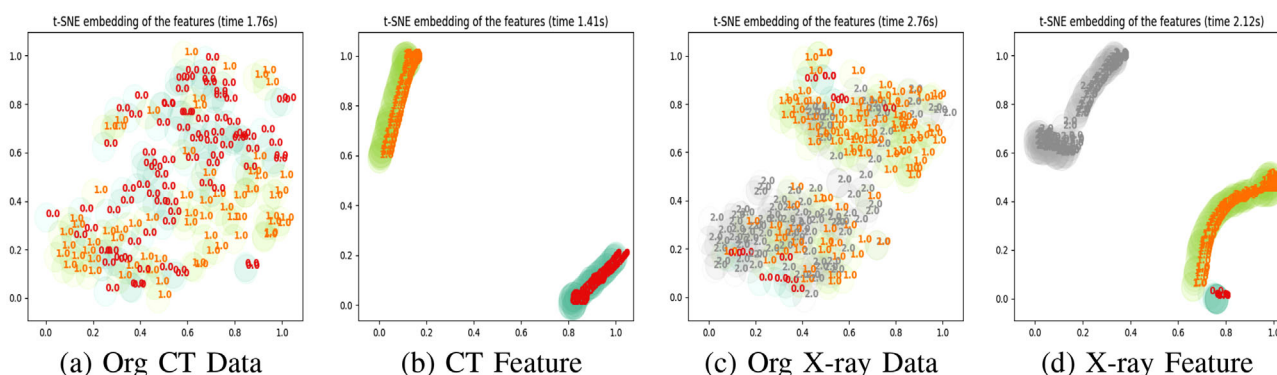

(a) Org CT Data (b) CT Feature (c) Org X-ray Data (d) X-ray Feature

FIGURE 11 Dimension reduction comparison between extracted features and original data by using t-SNE

samples. COVID-19 matrix shows us that there are 134 correct samples for model detection, and ten wrong samples include six COVID-19 samples and three healthy subjects. The recognition accuracy is 93.06%, and the average recognition rate for the two classifications is (89.86% + 96%) / 2 = 92.93%. By comparison, the detection rate of the proposed model for COVID-19 is lower than the recognition rate for healthy subjects, which is related to the insignificant lung imaging performance of patients with mild COVID-19. The verification of this experiment also requires the supervision of medical experts.

4.3.3 | Classification ability evaluation

We utilize t-SNE (t-distributed stochastic neighbour embedding) to show the dispersion of the original data and extracted features after dimensionality reduction to verify that the proposed method can make the original samples more separable. As shown in Figure 11, (a) and (b) are the comparison between the original data in CT dataset and the extracted feature data after dimensionality reduction, and (c) and (d) are the comparison between the original data in X-ray dataset and the extracted feature data after dimensionality reduction. The evidence shows that (b) and (d) have strong separability, while the sample distribution of (a) and (c) is scattered, and the labelled similar data are not clustered together.

5 | CONCLUSION

We developed a DCACorrCapsNet for COVID-19 detection and recognition in this paper. This model constructs features from both spatial, structural, and correlative dimensions by performing multi-level networks. The developed deep architecture generates correlative information from adjacent capsules, extracts attention-based spatial feature, and fuses time- and frequency-feature as a Fisher vector. The final feature representation is computed by combining information from these outputs. We evaluated the DCACorrCapsNet model using the X-ray and the CT medical datasets. Results show that the DCACorrCapsNet model outperforms compared methods on the data, demonstrating its superior performance in real-world environments.

In this work, the DCACorrCapsNet is considered for image recognition. There are also other deep architectures, such as the variants of the transformer, which achieve promising performance on image classification. It would be interesting to extend such models for medical detection and diagnosis. Furthermore, we will explore the unsupervised training of DCACorrCapsNet in the future.

AUTHOR CONTRIBUTIONS
Aite Zhao: Supervision, writing - original draft, writing - review and editing. Huimin Wu: Methodology, software, writing - original draft. Ming Chen: Formal analysis, methodology, software, visualization. Nana Wang: Methodology, software, writing - original draft.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT
Data available on request from the authors.

ORCID
Aite Zhao https://orcid.org/0000-0003-3494-175X

REFERENCES
1. Telenti, A., Arvin, A., Corey, L., Corti, D., Diamond, M.S., García-Sastre, A., Garry, R.F., Holmes, E.C., Pang, P.S., Virgin, H.W.: After the pandemic: Perspectives on the future trajectory of Covid-19. Nature 596(7873), 495–504 (2021)
2. Fan, D.-P., Zhou, T., Ji, G.-P., Zhou, Y., Chen, G., Fu, H., Shen, J., Shao, L.: INF-NET: Automatic Covid-19 lung infection segmentation from CT images. IEEE Trans. Med. Imaging 39(8), 2626–2637 (2020)
3. Kraemer, M.U., Yang, C.-H., Gutierrez, B., Wu, C.-H., Klein, B., Pigott, D.M., Open COVID-19 Data Working Group, du Plessis, L., Faria, N.R., Li, R., et al.: The effect of human mobility and control measures on the Covid-19 epidemic in China. Science 368(6490), 493–497 (2020)
4. Oh, Y., Park, S., Ye, J.C.: Deep learning Covid-19 features on CXR using limited training data sets. IEEE Trans. Med. Imaging 39(8), 2688–2700 (2020)
5. Demir, F.: Deepcoronet: A deep LSTM approach for automated detection of Covid-19 cases from chest X-ray images. Appl. Soft Comput. 103, 107160 (2021)
6. Al-Antari, M.A., Hua, C.H., Bang, J., Lee, S.: Fast deep learning computer-aided diagnosis of Covid-19 based on digital chest X-ray images. Appl. Intell. 51(5), 2890–2907 (2021)
7. Hussain, E., Hasan, M., Rahman, M.A., Lee, I., Tamanna, T., Parvez, M.Z.: Corodet: A deep learning based classification for Covid-19 detection using chest X-ray images. Chaos Solitons Fractals 142, 110495 (2021)
8. Hira, S., Bai, A., Hira, S.: An automatic approach based on CNN architecture to detect Covid-19 disease from chest X-ray images. Appl. Intell. 51(5), 2864–2889 (2021)
9. Nayak, S.R., Nayak, D.R., Sinha, U., Arora, V., Pachori, R.B.: Application of deep learning techniques for detection of Covid-19 cases using chest X-ray images: A comprehensive study. Biomed. Signal Process. Control 64, 102365 (2021)
10. Kollias, D., Arsenos, A., Soukissian, L., Kollias, S.: Mia-cov19d: Covid-19 detection through 3-D chest CT image analysis. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 537–544. IEEE, Piscataway, NJ (2021)
11. Alshazly, H., Linse, C., Barth, E., Martinetz, T.: Explainable Covid-19 detection using chest CT scans and deep learning. Sensors 21(2), 455 (2021)
12. Aslan, M.F., Unlersen, M.F., Sabanci, K., Durdu, A.: CNN-based transfer learning-BILSTM network: A novel approach for Covid-19 infection detection. Appl. Soft Comput. 98, 106912 (2021)
13. Castiglione, A., Vijayakumar, P., Nappi, M., Sadiq, S., Umer, M.: Covid-19: Automatic detection of the novel coronavirus disease from CT images using an optimized convolutional neural network. IEEE Trans. Ind. Inf. 17(9), 6480–6488 (2021)
14. Panigrahy, C., Seal, A., Mahato, N.K.: MRI and SPECT image fusion using a weighted parameter adaptive dual channel PCNN. IEEE Signal Process. Lett. 27, 690–694 (2020)

15. Shastri, S., Singh, K., Kumar, S., Kour, P., Mansotra, V.: Deep-LSTM ensemble framework to forecast Covid-19: An insight to the global pandemic. Int. J. Inf. Technol. Manage. 13(4), 1291–1301 (2021)

16. Dastider, A.G., Sadik, F., Fattah, S.A.: An integrated autoencoder-based hybrid CNN-LSTM model for Covid-19 severity prediction from lung ultrasound. Comput. Biol. Med 132, 104296 (2021)

17. Zhao, A., Li, J., Ahmed, M.: SpiderNet: A spiderweb graph neural network for multi-view gait recognition. Knowledge Based Syst. 206, 106273 (2020)

18. Wang, Y., Lv, Z., Sheng, Z., Sun, H., Zhao, A.: A deep spatio-temporal meta-learning model for urban traffic revitalization index prediction in the Covid-19 pandemic. Adv. Eng. Inf. 53, 101 678–101 678 (2022)

19. Zhao, A., Li, J.: Two-channel LSTM for severity rating of parkinson's disease using 3D trajectory of hand motion. Multimedia Tools Appl. 81, 33851–33866 (2022)

20. Hu, R., Gan, J., Zhu, X., Liu, T., Shi, X.: Multi-task multi-modality svm for early Covid-19 diagnosis using chest CT data. Inf. Process. Manage. 59(1), 102782 (2022)

21. Sen, S., Saha, S., Chatterjee, S., Mirjalili, S., Sarkar, R.: A bi-stage feature selection approach for Covid-19 prediction using chest CT images. Appl. Intell. 51, 8985–9000 (2021)

22. Le, D.-N., Parvathy, V.S., Gupta, D., Khanna, A., Rodrigues, J.J., Shankar, K.: IOT enabled depthwise separable convolution neural network with deep support vector machine for Covid-19 diagnosis and classification. Int. J. Mach. Learn. Cybern. 12, 3235–3248 (2021)

23. Mansour, N.A., Saleh, A.I., Badawy, M., Ali, H.A.: Accurate detection of Covid-19 patients based on feature correlated naïve Bayes (FCNB) classification strategy. J. Ambient Intell. Hum. Comput. 13, 41–73 (2021)

24. Sultana, J., et al.: Predicting indian sentiments of Covid-19 using MLP and ADABOOST. Turkish J. Comput. Math. Educ. 12(10), 706–714 (2021)

25. Gök, E.C., Olgun, M.O.: SMOTE-NC and gradient boosting imputation based random forest classifier for predicting severity level of Covid-19 patients with blood samples. Neural Computing Appl. 33(22), 15693–15707 (2021)

26. Du, W., Chen, H., Liao, P., Yang, H., Wang, G., Zhang, Y.: Visual attention network for low-dose CT. IEEE Signal Process. Lett. 26(8), 1152–1156 (2019)

27. Dubey, S.R., Singh, S.K., Singh, R. K.: Local diagonal extrema pattern: A new and efficient feature descriptor for CT image retrieval. IEEE Signal Process. Lett. 22(9), 1215–1219 (2015)

28. Zhao, A., Qi, L., Dong, J., Yu, H.: Dual channel LSTM based multi-feature extraction in gait for diagnosis of neurodegenerative diseases. Knowledge Based Syst. 145, 91–97 (2018)

29. Hussain, A.A., Bouachir, O., Al-Turjman, F., Aloqaily, M.: Notice of Retraction: AI techniques for Covid-19. IEEE Access. 8, 128776–128795 (2020)

30. Singh, V., Poonia, R.C., Kumar, S., Dass, P., Agarwal, P., Bhatnagar, V., Raja, L.: Prediction of Covid-19 corona virus pandemic based on time series data using support vector machine. J. Discrete Math. Sci. Cryptogr. 23(8), 1583–1597 (2020)

31. Novitasari, D.C.R., Hendradi, R., Caraka, R.E., Rachmawati, Y., Fanani, N.Z., Syarifudin, A., Toharudin, T., Chen, R.C.: Detection of Covid-19 chest X-ray using support vector machine and convolutional neural network. Commun. Math. Biol. Neurosci. 2020, 42 (2020)

32. Tang, Z., Zhao, W., Xie, X., Zhong, Z., Shi, F., Liu, J., Shen, D.: Severity assessment of Coronavirus disease 2019 (Covid-19) using quantitative features from chest CT images. Phys. Med. Biol. 66(3), 035015 (2020)

33. Dai, Q., Zhu, X.: Age prediction of death in india' Covid-19 pandemic using machine learning methods. In: 2021 IEEE 3rd International Conference on Frontiers Technology of Information and Computer (ICFTIC), pp. 242–245. IEEE, Piscataway, NJ (2021)

34. Nasiri, H., Hasani, S.: Automated detection of Covid-19 cases from chest X-ray images using deep neural network and XGBOOST. 28(3), 732–738 (2022)

35. Gayathri, J., Abraham, B., Sujarani, M., Nair, M.S.: A computer-aided diagnosis system for the classification of Covid-19 and non-Covid-19 pneumonia on chest X-ray images by integrating CNN with sparse autoencoder and feed forward neural network. Comput. Biol. Med. 141, 105134 (2022)

36. Krishnaswamy Rangarajan, A., Ramachandran, H.K.: A fused lightweight cnn model for the diagnosis of Covid-19 using CT scan images. Automatika 63(1), 171–184 (2022)

37. de Moura, J., Novo, J., Ortega, M.: Fully automatic deep convolutional approaches for the analysis of Covid-19 using chest X-ray images. Appl. Soft Comput. 115, 108190 (2022)

38. Guarrasi, V., D'Amico, N.C., Sicilia, R., Cordelli, E., Soda, P.: Pareto optimization of deep networks for covid-19 diagnosis from chest x-rays. Pattern Recognit. 121, 108242 (2022)

39. Woo, S., Park, J., Lee, J.Y., Kweon, I.S.: Cbam: Convolutional block attention module. In: European Conference on Computer Vision. Lecture Notes in Computer Science, Vol. 11211, pp 3–19. Springer, Cham (2018)

40. Ke, G., Xu, Z., Zhang, J., Bian, J., Liu, T.Y.: Deepgbm: A deep learning framework distilled by GBDT for online prediction tasks. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 384–394. Association for Computing Machinery, New York, NY (2019)

41. Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: 2017 IEEE conference on computer vision and pattern recognition (CVPR), pp. 2261–2269. IEEE, Piscataway, NJ (2018)

42. Tan, M., Le, Q.V.: EfficientNet: Rethinking model scaling for convolutional neural networks. arXiv:1905.11946, (2020)

43. Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q.V., Adam, H.: Searching for MobileNetv3. arXiv:1905.02244, (2019)

44. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16 × 16 words: Transformers for image recognition at scale. arXiv:2010.11929, (2020)

45. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. arXiv:2103.14030, (2021). https://arxiv.org/abs/2103.14030

46. Radosavovic, I., Kosaraju, R.P., Girshick, R., He, K., Dollár, P.: Designing network design spaces. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10425–10 433. IEEE, Piscataway, NJ (2020)

47. Touvron, H., Vedaldi, A., Douze, M., Jégou, H.: Fixing the train-test resolution discrepancy: FixefficientNet. arXiv:2003.08237, (2020)

48. Tan, M., Le, Q.V.: EfficientNetV2: Smaller models and faster training. arXiv:2104.00298, (2021)

49. Cohen, J.P., Morrison, P., Dao, L.: Covid-19 image data collection. arXiv:2003.11597, (2020)

50. Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R.M.: ChestX-Ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3462–3471. IEEE, Piscataway, NJ (2017)