Pre-training Deep Neural Networks using 3D Formula-Driven Supervised Learning for COPD Staging

An Accepted Article

Yasumasa Tamura*
Faculty of Information Science and Technology
Hokkaido University
Sapporo, Hokkaido, 060-0814, Japan
ytamura@ist.hokudai.ac.jp

Kohei Harada
Faculty of Information Science and Technology
Hokkaido University
Sapporo, Hokkaido, 060-0814, Japan

Wataru Noguchi
Education and Research Center for Mathematical and Data Science
Hokkaido University
Sapporo, Hokkaido, 060-0812, Japan

Kaoruko Shimizu
Department of Respiratory Medicine
Hokkaido University
Sapporo, Hokkaido, 060-8638 Japan.

Satoshi Konno
Department of Respiratory Medicine
Hokkaido University
Sapporo, Hokkaido, 060-8638 Japan.

Masahito Yamamoto
Faculty of Information Science and Technology
Hokkaido University
Sapporo, Hokkaido, 060-0814, Japan
masahito@ist.hokudai.ac.jp

Received: 14 April, 2025 / Accepted: 14 July, 2025 / Published: 7 August, 2025

Abstract

This study proposes a novel pre-training method employing the 3D formula-driven supervised learning for classifying three-dimensional computed tomography (CT) images, specifically for staging *chronic obstructive pulmonary disease (COPD)*. The proposed method leverages mathematically generated 3D images as a pre-training dataset to enhance feature extraction while addressing the scarcity of labeled training data, primarily caused by ethical concerns inherent in medical imaging. This paper evaluates the effectiveness of the proposed method in two CT image classification tasks: COVID-19 classification and COPD staging. The experimental result shows that the performance of a COVID-19 classification model pre-trained using mathematically generated 3D images is competitive to the existing model pre-trained using numerous CT lung images. The result also reveals the potential of the proposed approach for COPD staging, mitigating the data scarcity issue.

Keywords Pre-training · Formula-driven supervised learning · 3D image classification · Medical informatics

*This study was conducted with the approval of the Institutional Review Board (IRB) of Hokkaido University Hospital (reference number: 018-0394).

1 Introduction

Chronic obstructive pulmonary disease (COPD) is a chronic respiratory disease that leads to impaired lung function with irreversible structural changes, i.e., remodeling, of pulmonary tissues [9]; hence, early diagnosis and staging are crucial. Primary drivers of COPD include prolonged exposure to noxious particles or gases, most notably from tobacco smoking. COPD is a leading global health concern, affecting over 200 million individuals [5]. Characterized by respiratory symptoms and airflow limitations caused by abnormalities in the airways and alveoli, it imposes a significant burden on both patients and healthcare systems.

Medical imaging, particularly computed tomography (CT), has emerged as an clinically indispensable tool in understanding and managing COPD. In contrast to *spirometry*, a traditional plumonary function test that provides a global measurement of lung function, CT imaging provides high-resolution visualization of regional lung structures. Such regional information enables the identification of disease heterogeneity, such as emphysema, bronchial wall thickening, and vascular changes, resulting in obtaining critical insights into disease severity and progression [3, 18].

The promising effectiveness of artificial intelligence (AI), particularly deep neural networks (DNNs), in the field of medical imaging has been reported in automating and augmenting image analysis while also improving precision and reproducibility [4, 7]. High-resolution imaging generates massive data volumes that require intensive manual interpretation, often leading to variability and inefficiencies [20]. This underscores the need for advanced computational methods to help analysis and decision-making. In response to this need, three-dimensional analysis using DNN models demonstrates its potential for providing a more comprehensive understanding of lung structures and disease phenotypes, by incorporating inter-slice information into the analysis [6, 8].

This study aims to enhance the accuracy of a COPD classification model through pre-training using mathematically generated images, known as *formula-driven supervised learning (FDSL)* [15]. One critical challenge in building a DNN-based classification model for CT images—particularly 3D CT images—is the scarcity of labeled training data. This challenge arises not only from the cost of data collection but also from the data privacy and ethical concerns inherent in medical imaging. Our proposed pre-training method mitigates the data scarcity issue by leveraging mathematically generated 3D images that are free from ethical concerns.

Contribution: The primary contribution of this study is suggesting a promising approach to mitigating data scarcity issues in COPD staging with 3D CT images. Specifically, this paper provides

- two mathematically generated 3D image datasets for pre-training: 3D fractal dataset and 3D Perlin noise dataset;
- an evaluation of the impact of pre-training with the generated datasets for the COVID-19 classification task; and
- a demonstration of the effectiveness of pre-training with the 3D fractal dataset for the COPD staging task.

The proposed method may also have a potential to improve accuracy on other classification tasks in medical imaging. One limitation is that the link between a specific task and effective features of pre-training dataset is not yet revealed and remaining as future work.

Extension from the conference version: This paper is an extended version of the study that originally presented at the 30th international symposium on artificial life and robotics (AROB 30th 2025) [10]. The original work only verified the effectiveness of pre-training with the 3D fractal dataset. In this paper, we additionally test the performance of pre-training with the Perlin noise dataset, a mathematically generated pre-training dataset that is not relevant to fractal structures. This extension further highlights the significant effectiveness of pre-training with the 3D fractal dataset.

Paper organization: Section 2 provides a brief review of the literature related to 3D CT image classification and FDSL. Section 3 describes the proposed pre-training method followed by generated two pre-training datasets. Section 4 evaluates and discusses the effectiveness of the proposed pre-training method. Section 5 finally concludes this paper.

2 Related work

This section provides a brief survey on related work to clarify the direction and motive of this study. The first subsection specifically focuses on 3D CT image classification, clarifying its challenges and reviewing recent work. The second subsection introduces the concept of FDSL with related work.

2.1 3D CT image classification

The application of 3D CT image classification in medical imaging has gained increasing attention; however, it still lags behind the extensive studies in two-dimensional imaging modalities, such as the study with chest radiographs (CR) or 2D CT slices [6, 8]. This gap primarily arises from the computational cost associated with processing 3D image data, which is significantly higher than that of processing 2D image data. In addition, although deep learning techniques show their superiority even in medical imaging [4, 7, 20], training models using 3D medical images often faces a scarcity of labeled training data, further complicating the adoption of advanced learning techniques.

To address the scarcity of labeled training data in 3D CT image classification, Y. Huang et al. [13] proposed a promising approach utilizing *autoencoders* for pneumoconiosis staging. Their approach leverages numerous unlabeled 3D lung CT images to train an autoencoder model, enabling the encoder of the trained autoencoder to extract key features from the lung CT images. Using this encoder as a backbone feature extractor, the pneumoconiosis staging model is fine-tuned using labeled training data. The pre-trained encoder showed a significant performance improvement in comparison to models trained from scratch.

One limitation of Huang et al.'s approach is that training the effective encoder still requires large-scale unlabeled 3D lung CT image datasets. The goal of our study is to mitigate this limitation by developing a pre-training method independent of the actual CT images.

2.2 Formula-driven supervised learning

FDSL is a pre-training method using mathematically generated images, proposed by Kataoka et al. [15]. Their study particularly showed the potential of fractal structures, characterized by their self-similarity and intricate geometric patterns, to generate synthetic datasets for training machine learning models. In contrast to manually annotated training data which is commonly used in traditional supervised learning, fractal structures can be mathematically defined and generated in large volumes without extensive human labor. Additionally, as an important advantage of the mathematically generated data, they inherently circumvent ethical concerns.

Pre-training with fractal structures have also demonstrated effectiveness in multi-view image recognition and object recognition on 3D point cloud data, suggesting a capability to effectively capture complex spatial arrangements and enable more accurate identification of objects in unstructured environments [21, 22]. Another interesting suggestion from those studies is that the fractal structures, which are not directly related to the input data in target tasks, still work effectively. This highlights the potential of fractal-based approaches to enhance feature extraction and classification, even for tasks in which the input data does not exhibit fractal characteristics.

These advancements in FDSL strongly motivate our proposed method: pre-training with 3D FDLS. In particular, using 3D fractal images as pre-training dataset has promising potential to be effective in the context of medical imaging, where the scarcity of both labeled and unlabeled training data poses a challenge.

3 Pre-training with 3D formula-driven supervised learning

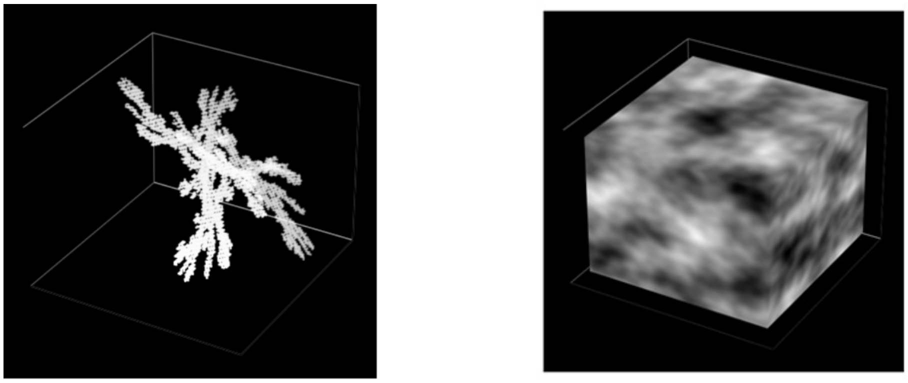
The primary challenge in the 3D CT image recognition is the scarcity of labeled training data, which is inherent in medical imaging, as described in Section 1. This limitation makes training difficult, particularly in extracting key features for recognition or for classification into appropriate classes. To address this issue, the proposed algorithm trains the model based on a large, mathematically generated 3D image dataset, thereby enhancing its feature extraction capability for similar images.

In this section, we first clarify the assumptions in this study regarding the input images. The second subsection outlines the pre-training method and defines two mathematically generated pre-training datasets: a 3D fractal dataset and a 3D Perlin noise dataset.

3.1 Assumptions

Assume that a 3D image is represented as a $W \times H \times D$ voxel volume, consisting of D consecutive slices of images obtained either from a single rule in the context of FDSL, or from a single individual in the context of CT images, where W and H denote the width and height of a single slice respectively. In this study, we specifically assume $W = H = D = 128$ due to the limitation of available computational resources. This assumption does not compromise the generality of the proposed method.

We also assume monochrome 3D images in this study. Each voxel in a monochrome 3D image holds a real value in the range $[0.0, 1.0]$, representing the brightness of the voxel. Under this assumption, a classification model only



(a) 3D fractal image (b) 3D Perlin noise image

Figure 1: Examples of images in generated pre-training datasets: (a) shows a sample in the 3D fractal dataset, whereas (b) shows a sample in the 3D Perlin noise dataset.

Table 1: Summary of generated pre-training datasets

	3D fractal	3D Perlin noise
Classes	1000	1000
Instances per class	145	100
Total instances	145 000	100 000

distinguishes differences in voxel brightness. However, with some effort, the model can be readily extended to deal with multi-valued voxels, such as the RGB values found in colored 3D images. Consequently, this assumption does not excessively limit the applicability of the proposed method.

3.2 Pre-training method and datasets

The proposed pre-training method trains the model in supervised learning manner. Given a set of 3D images $X = \{x_1, x_2, \cdots, x_n\}$, each 3D image $x_i \in X$ is associated with one of class labels $y_j = M(x_i)$, where $M : X \mapsto Y$ denotes the underlying mapping from the image set X to label set $Y = \{y_1, y_2, \cdots, y_k\}$. Pre-training method optimizes the model to approximate the actual mapping M behind the given sample pairs $(x_i, M(x_i)) \in X \times Y$.

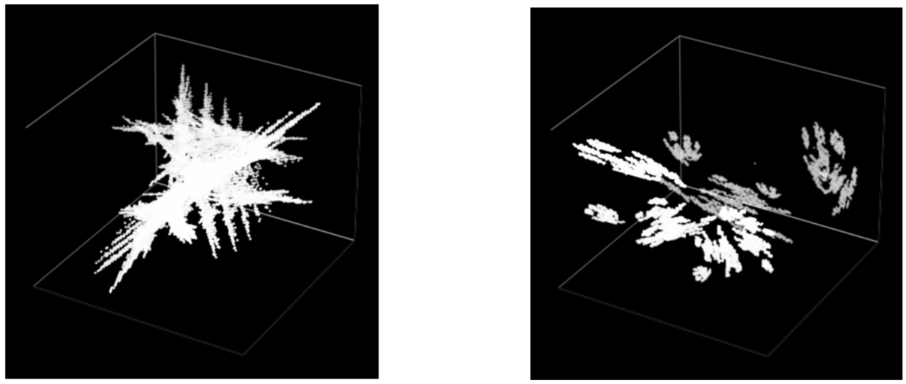
To this end, the pre-training dataset must consist of pairs of input 3D images and their associated class labels. In FDSL, the input 3D images are generated from predefined formulae and classified into multiple classes based on the characteristics of the corresponding formula.

This paper generates and evaluates two types of formula-driven pre-training datasets: a 3D fractal dataset and a 3D Perlin noise dataset. The effectiveness of 3D fractal point clouds as a pre-training dataset for natural image recognition was demonstrated by Yamada et al. [21]; however, their utility in capturing features specific to lung structures remains unclear. In contrast to 3D fractal images with distinct contours, 3D Perlin noise images are composed entirely of smooth grayscale gradations. CT lung images typically contain both characteristics, and we investigate the effectiveness of these two types of 3D images as pre-training datasets.

Examples of generated 3D images from each dataset are shown in Fig. 1. In addition, Table 1 summarizes the two generated pre-training datasets in terms of the number of classes and instances. The rest of this subsection describes the details of two datasets.

3.2.1 3D fractal dataset

3D fractal structures are generated using a mathematical framework called *iterated function systems (IFS)* [14]. An IFS is defined by a set of pairs consisting of an affine transformation and its associated activation probability, provided that every affine function must be a contraction mapping. Given an initial 3D vector representing a point in space, a 3D fractal structure is formed as a point cloud by iteratively applying stochastically selected affine functions to the vector until convergence. Figure 1a shows an example of a 3D fractal structure generated by an well-formed IFS.



(a) Sample 1 (b) Sample 2

Figure 2: Generated 3D fractal structures classified into two distinct classes: (a) appears to consist of one large cluster of points, whereas (b) consists of several small clusters that are widely spread.

The structures generated by an IFS are determined by the parameters of affine transformations. This property allows us to treat structures generated by different parameters as distinct classes. The original work [21] presents a category search method for generating an arbitrary number of transformation matrices for IFS, each producing distinct fractal structure. Based on this method, we generated 1000 classes of fractal structures. By further applying small perturbations to parts of the parameters, we augmented instances while preserving their essential features, thereby creating multiple instances—145 in this case—within each class, resulting in a total of $145\,000$ instances.

To create a single 3D fractal structure, the associated IFS produces $100\,000\,000$ points to form a 3D point cloud. The coordinates of each point are then normalized to the range $[0, 127]$, representing a 3D image of size $128 \times 128 \times 128$. Figure 2 shows two instances arbitrarily picked up from two distinct classes, for example.

3.2.2 3D Perlin noise dataset

Perlin noise [19] is a type of gradient noise widely used in computer graphics (CG) to generate natural textures, patterns, etc. In contrast to fractal structures, a 3D image based on Perlin noise has smooth contour as shown in Fig. 1b. In this sense, the 3D Perlin noise dataset clearly exhibits different characteristics from the 3D fractal dataset.

Although the Perlin noise is typically used to generate a 2D texture, the algorithm can be readily extendable to 3D spaces. In this study, we formulate the 3D Perlin noise generation algorithm to generate the 3D Perlin noise dataset. Given a 3D space of size $W \times H \times D$, where W, H and D are the positive natural numbers, the outline of the generation algorithm is shown as follows.

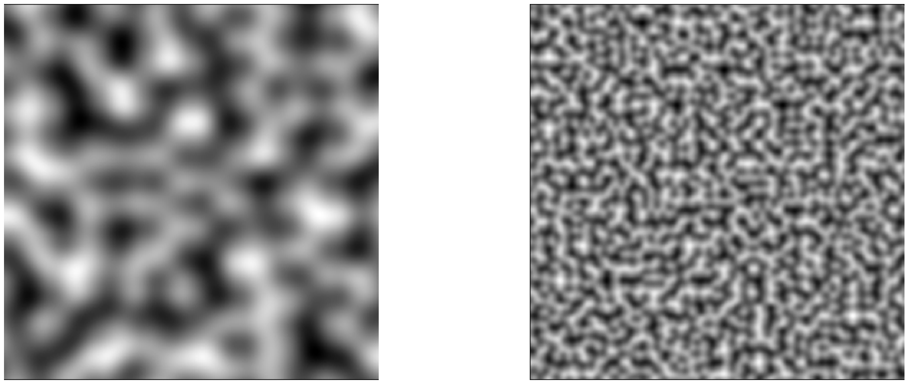
1. Divide the space into $w \times h \times d$ grids, provided that $1 \leq w < W, 1 \leq h < H, 1 \leq d < D$.
2. Assign random gradient vectors to each grid-points.
3. Compute each voxel value based on the inner products of the gradient vectors assigned to the grid-points surrounding the target voxel and the distance vector from the target voxel to those grid-points, with a deterministic smoothing operation applied.

We utilize this algorithm with $W = H = D = 128$ to generate arbitrary numbers of random images of size $128 \times 128 \times 128$.

The generated 3D Perlin noise images can be primarily characterized by the grid size (w, h, d); therefore, we use the grid size to determine the classes of the generated images. In this study, we prepare 1000 grid size patterns as shown in eq. (1), resulting in 1000 distinct classes.

$$\begin{aligned} &(w, h, d) \in Z^3 \\ &\text{where } Z = \{4, 7, 10, 13, 16, 19, 22, 25, 28, 31\} \end{aligned} \tag{1}$$

Within each class corresponding to the grid size (w, h, d), 100 repetitions of the algorithm generate 100 unique instances depending on the random gradient vectors assigned in Step 2. 100 repetitions for each of the 1000 grid size patterns generate a total of $100\,000$ instances.



(a) With 8×8 grids (b) With 32×32 grids

Figure 3: 2D Perlin noise images with different number of grids: (a) shows the 2D Perlin noise generated in 8×8 grids, whereas (b) shows the one in 32×32 grids. The smoothness of contour is a characteristic of Perlin noise.

To illustrate the distinction between Perlin noise images determined by grid size, we show two samples of 2D Perlin noise with different grid sizes: 8×8 and 32×32, in Fig. 3. It is evident that the 2D Perlin noise generated with 32×32 grids has sharper contours and appears more complex than that generated with 8×8 grids. The same characteristics hold for 3D Perlin noise images.

4 Evaluation

In this section, we evaluate the proposed pre-training method in two CT image classification tasks: COVID-19 classification task using open COVID-CT-MT dataset and COPD staging task with COPD dataset. The evaluation on the former task aims to generally assess the impact of the proposed pre-training method in comparison to an existing baseline model, whereas the evaluation on the latter task specifically examines its effectiveness in COPD staging, which is the primary objective of this study. The impact of the proposed pre-training method is evaluated by comparing the performance of models fine-tuned for the target datasets after pre-training with that of models directly trained using the target datasets without pre-training.

4.1 Experimental setup

In the experiment, we determine the hyper-parameters for training models based on preliminary experiments. The models are trained using the Adam optimizer [16] with a learning rate 0.0001, and a training is terminated when the cross-entropy loss has converged.

The specific model architectures, target datasets, and data augmentation techniques used in the experiment are summarized in the following parts.

4.1.1 Model architectures

We compared four DNN models, including two convolutional neural network (CNN) based models: *ResNet* [11] and *DenseNet* [12], and two vision transformer-based models: *ViViT* [2] and *Swin-Transformer* [17]. This selection of architectures was based on their established success in various image and video recognition tasks.

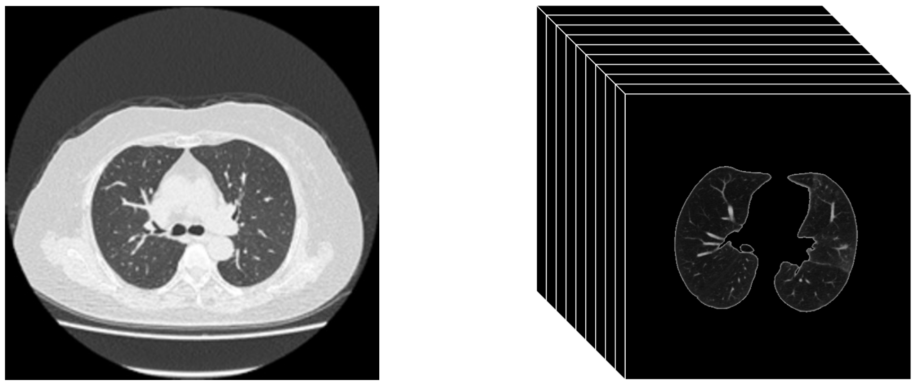
4.1.2 CT image datasets for fine-tuning

COVID-CT-MD dataset [1] is used to generally assess the effectiveness of the proposed pre-training method. This open dataset has been widely utilized in machine learning and deep learning study for COVID-19 classification. It comprises 169 confirmed COVID-19 positive cases, 76 normal cases, and 60 cases of community-acquired pneumonia (CAP). All cases were sourced from the Babak Imaging Center in Tehran, Iran, and labeled by three experienced radiologists.

COPD dataset is provided from the Hokkaido COPD Cohort Study, a multicenter observational prospective cohort study conducted from 2003 to 2015. This dataset includes 91 subjects in total, classified into four stages based on

Table 2: Stages and number of subjects in COPD dataset

Stage	Severity	Subject
I	Mild	17
II	Moderate	47
III	Sever	22
IV	Very sever	5



(a) Original slice of CT image (b) 3D CT image

Figure 4: Lung CT images: (a) shows a slice of the original CT image (512×512), and (b) shows a 3D CT image ($128 \times 128 \times 128$) after preprocessing.

the the Global Initiative for Chronic Obstructive Lung Disease (GOLD)[2] report, as shown in Table 2. The original study was initiated during the 2003–2005 period and completed in 2013–2015. COPD diagnosis was confirmed by respirologists before patient enrollment. During the study, patients with confirmed COPD underwent annual chest CT scans during exacerbation-free stable periods for the first five years. The study was conducted in compliance with the revised Declaration of Helsinki. As stated on the first page, approval was granted by the Institutional Review Board (IRB) of Hokkaido University Hospital (reference number: 018-0394). Written informed consent was obtained from all participants in accordance with the Hokkaido COPD Cohort Protocol. Furthermore, patients were informed of the study protocol through an online platform and were provided an opt-out option as per the IRB guidelines of Hokkaido University Hospital.

Note that the COPD dataset includes significantly more samples from men than from women, due to the common sex disparity in the number of COPD-affected patients. This imbalance, however, does not affect the validity of the proposed pre-training method.

For each dataset, we use $80\,\%$ of the data for training (fine-tuning), $10\,\%$ of the data for validation, and $10\,\%$ of the data for test, ensuring no overlap between the splits while maintaining the balance of class labels. It is also ensured that the test data is never used for training, and only used for evaluation.

4.1.3 Preprocessing and data augmentation

All CT images underwent preprocessing to standardize resolution and intensity values for effective 3D neural network analysis. In this study, the given 3D CT images, originally with varying numbers of 512×512 pixels image slices, are resampled to a uniform voxel data size of $128 \times 128 \times 128$ (Fig. 4).[3] Each voxel contains a real value in the range $[0.0, 1.0]$, where 0.0 and 1.0 respectively correspond to -1000 and 400 in Hounsfield Unit (HU) based on the pixel values of the original CT images. Note that, the HU values under -1000 and over 400 are not relevant for lung tissue analysis. Furthermore, we applied lung segmentation to isolate the regions of interest, ensuring that the neural network focused solely on relevant lung structures. These preprocessing steps have been reported to enhance the efficiency of 3D neural network training by reducing variability and excluding irrelevant features [6].

[2] https://goldcopd.org/
[3] 128 comes from the assumption in Section 3.1

Table 3: Definitions of TP, FP, TN and FN

Set	Entity
TP	positive samples correctly classified as positive
FP	positive samples incorrectly classified as negative
TN	negative samples correctly classified as negative
FN	negative samples incorrectly classified as positive

Through extensive preliminary experiments, we found that data augmentation is highly effective in training models on 3D lung CT images. Random data augmentation techniques help suppress early overfitting and improve model performance, especially when training data is limited in quantity or diversity, as is often the case observed in 3D CT image recognition. The preliminary experiments examined validation accuracy achieved over a fixed number of training epochs, comparing the performance of a model trained without data augmentation to that of a model trained with randomly applied augmentation. As a baseline, the model trained without augmentation achieved $\sim 60\,\%$ validation accuracy. We initially tested six augmentation methods: adding random noise, random scaling, random rotation, random translation, elastic distortion, and vertical or horizontal flipping. On the one hand, adding random noise, random rotation, and random translation each achieved over $90\,\%$ validation accuracy, corresponding to a gain of $\sim 30\,\%\mathrm{pt}$. On the other hand, random scaling, elastic distortion, and flipping each independently improved validation accuracy, but were found to slow convergence when all methods were stochastically combined. Based on these observations, we adopted the following three augmentation methods for the subsequent experiments.

Random noise
Gaussian noise with $\mu = 0.0$ and $\sigma^2 = 0.10$ is added, where μ and σ^2 denote the mean and standard deviation respectively.

Random rotation
The entire image is randomly rotated within the angle range $[0°, 30°]$. This augmentation accounts for orientation variability in CT images.

Random translation
The entire image is randomly translated within the range of $[-6, 6]$ pixels. This augmentation accounts for slight positional aberrations in CT images.

Each of these augmentations was independently applied during training with a probability of $50\,\%$.

4.2 Metrics

We evaluate the performance of models based on four metrics commonly used to assess classification capability: *accuracy*, *recall*, *precision*, and F_1-*score*. Since the target tasks are categorized as multi-class classification, we specifically rely on the macro variant of each metric. The definitions of each metric are briefly introduced below.

In binary classification tasks, the metrics are defined as follows. Let TP, FP, TN and FN be the sets of *true positive*, *false positive*, *true negative*, and *false negative* samples, respectively, as detailed in Table 3. The *accuracy* denotes the proportion of correctly classified samples, defined as eq. (2), where $|\Omega|$ denotes the number of unique entities in the finite set Ω.

$$(\text{Accuracy}) = \frac{|TP| + |TN|}{|TP| + |TN| + |FP| + |FN|} \tag{2}$$

The *recall*, also known as the *true positive rate*, denotes the proportion of $|TP|$ to the total number of actual positive samples, whereas *precision* denotes the proportion of $|TP|$ to the total number of samples classified as positive by the model, defined as eq. (3) and eq. (4), respectively.

$$(\text{Recall}) = \frac{|TP|}{|TP| + |FN|} \tag{3}$$

$$(\text{Precision}) = \frac{|TP|}{|TP| + |FP|} \tag{4}$$

The F_1-*score* is a harmonic mean of the recall and precision, defined as eq. (5).

$$(F_1\text{-score}) = \frac{2}{(\text{Recall})^{-1} + (\text{Precision})^{-1}} \tag{5}$$

Table 4: Classification performance of four models and the baseline model (TBFE) on COVID-CT-MD dataset

Model	Pre-training	Accuracy (%)	Recall (%)	Precision (%)	F_1-score (%)
3D ResNet50	—	93.10	93.23	93.53	93.06
	Fractal	96.55	96.13	**96.75**	**96.45**
	Perlin noise	90.32	88.39	90.52	89.28
3D DenseNet121	—	86.21	81.25	90.28	82.75
	Fractal	93.10	95.83	91.07	92.99
	Perlin noise	86.21	85.52	81.25	81.69
ViViT	—	75.86	81.59	81.94	78.33
	Fractal	79.31	80.56	83.12	79.26
	Perlin noise	76.32	78.52	76.59	76.94
Swin3D	—	82.76	73.02	92.06	76.26
	Fractal	89.66	86.81	91.91	88.08
	Perlin noise	79.31	83.73	77.08	75.67
TBFE [13]	Lung CT images	**96.67**	**98.41**	93.33	95.48

Table 5: Classification performance of 3D ResNet50 on *limited* COVID-CT-MD dataset

Training data	Pre-training	Accuracy (%)	Recall (%)	Precision (%)	F_1-score (%)
100.0 %	—	93.10	93.23	93.53	93.06
	Fractal	**96.55**	**96.13**	**96.75**	**96.45**
75.0 %	—	89.66	86.81	91.91	88.08
	Fractal	**93.10**	**92.36**	**93.75**	**92.66**
50.0 %	—	86.21	82.04	89.68	84.65
	Fractal	**93.10**	**92.36**	**93.75**	**92.66**
25.0 %	—	82.76	73.02	**92.06**	76.26
	Fractal	**89.66**	**86.81**	91.91	**88.08**

In multi-class classification tasks, each metric above is extended to its macro-averaged variant. Let $(\mathrm{Accuracy})_j$ be the accuracy for the class j among K classes, provided that $K > 1$. The macro accuracy, for instance, is defined as eq. (6).

$$(\mathrm{Macro\ accuracy}) = \frac{1}{K}\sum_{j=1}^{K}(\mathrm{Accuracy})_j \tag{6}$$

The other metrics are also extended to their macro variants in the same manner. Hereinafter, the notations of *accuracy*, *recall*, *precision*, and F_1*-score* refer to their macro variants unless otherwise specified.

4.3 Result

We first show the experimental result on the COVID-19 classification task to validate the generalized effectiveness of the proposed pre-training method under the condition that sufficient volume of labeled training dataset are available. Then, we focus on the COPD staging task based on the observations from the COVID-19 classification task.

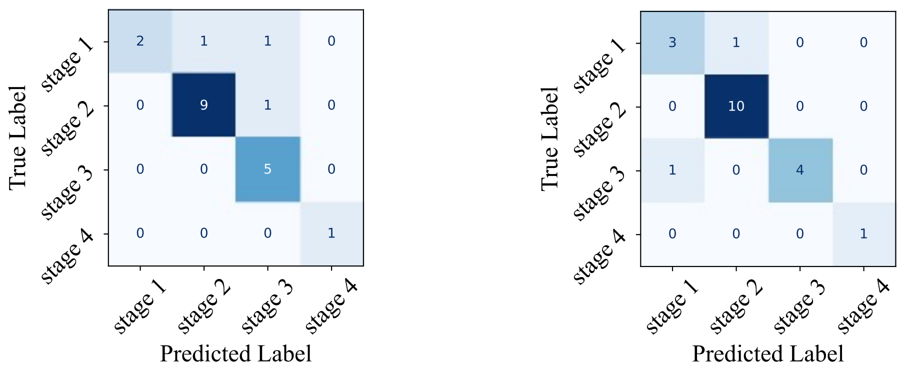
4.3.1 Task 1: COVID-19 classification

Table 4 shows the performance metrics of four DNN models and the baseline model (*TBFE*) from the literature [13]. For each model, except *TBFE*, the table shows the performance comparison across three conditions: without pre-training, with pre-training using the 3D fractal dataset, and with pre-training using the 3D Perlin noise dataset.

According to Table 4, *3D ResNet50* pre-trained using the 3D fractal dataset significantly outperforms in terms of precision, and shows the best performance in terms of F_1-score. Although the baseline model (*TBFE*) shows the best performance in terms of accuracy and recall, the performance of *3D ResNet50* pre-trained using the 3D fractal dataset is also competitive in those metrics. In contrast, the models pre-trained with the 3D Perlin noise dataset generally

Table 6: Classification performance of 3D ResNet50 on COPD dataset

Pre-training	Accuracy (%)	Recall (%)	Precision (%)	F_1-score (%)
—	85.00	85.00	**90.36**	85.00
Fractal	**90.00**	**91.48**	88.75	**89.78**



(a) Without pre-training (b) With pre-training

Figure 5: Confusion matrices of the COPD classification model without and with pre-training. The right figure (b) shows the less incorrect classifications than the left one (a), except for the classification of 'stage 3'.

show lower performance than those pre-trained with the 3D fractal dataset, or even worse than the models without pre-training in some cases.

Table 5 shows a comparison of models' tolerance to varying the volume of training data used for fine-tuning. Based on the result shown in Table 4, this result only compares *3D ResNet50* pre-trained using the 3D fractal dataset and that without pre-training.

The performance gap between the pre-trained model and the model without pre-training becomes more pronounced as the amount of training data decreased, as shown in Table 5. This result suggests the effectiveness of pre-training with a 3D fractal dataset for the COPD staging task, where the dataset size is significantly smaller than that of the COVID-CT-MD dataset.

4.3.2 Task 2: COPD staging

Table 6 shows a performance comparison for the COPD staging task using the COPD dataset. Based on the result observed in Section 4.3.1, and to avoid unnecessary confusion, we omit results with the models other than *3D ResNet50*, and compare its performance with and without pre-training using the 3D fractal dataset.

According to Table 6, the proposed pre-training method using the 3D fractal dataset significantly improves accuracy and recall of the model. Although the pre-training with the 3D fractal dataset affects negatively on precision of the model, the gap in precision values is less significant compered to the improvement in other metrics. It is worth mentioning that this trend is consistent with the result obtained using 25 % of training data in COVID-19 classification task, as shown in Table 5.

Figure 5 shows the confusion matrices for each setting. The classification model without pre-training exhibits misclassifications in 'stage 1' and 'stage 2' as shown in Fig. 5a. Although the model with pre-training introduces a new misclassification in 'stage 3', it improves the classifications in both 'stage 1' and 'stage 2' which correspond to the early stages of COPD.

4.4 Discussion

The experimental results demonstrate that the models pre-trained with the 3D fractal dataset consistently outperform those without pre-training across various architectures and target datasets. This fact suggests that the pre-training with 3D fractal images provides significant advantages, especially in scenarios where the amount of medical imaging data available for training is limited. Furthermore, in the COVID-19 classification task, the performance of certain architectures are competitive to the existing baseline model that leverages actual medical imaging for pre-training. Consequently, the pre-training with 3D fractal images could serve as a cost-effective alternative to traditional pre-training methods that rely on actual medical image datasets.

A couple of limitations should be acknowledged. First, the experiments were conducted only on specific datasets (COVID-CT-MD and COPD), and the effectiveness of the proposed pre-training method for other medical imaging tasks remains to be investigated. Second, although the 3D fractal dataset provides a synthetic and scalable option for pre-training, it inherently does not directly represent the features of real world medical images, potentially limiting its applicability for capturing domain-specific features.

Future work should focus on expanding the application of the pre-training with 3D fractal images to other medical imaging domains, and exploring hybrid approaches that integrate fractal data with real medical data. These efforts are expected to further validate the utility of fractal structures in enhancing the performance and reliability of deep learning models for medical imaging.

Future work also includes the pre-training of autoencoder models instead of classification models. In this study, we designed the pre-training dataset to share similar characteristics with 3D CT images, enabling the model to learn transferable features for 3D CT image classification. Although this approach proved effective, as demonstrated above, the meaningful features in the pre-training and downstream tasks may differ. Pre-training autoencoders using mathematically generated 3D images has the potential to yield more robust feature extraction capabilities for various downstream tasks.

5 Conclusion

This study proposed a pre-training method for the COPD stage classification, utilizing 3D formula-driven supervised learning. To the best of our knowledge, this is the first study to investigate the applicability of datasets composed of mathematically generated 3D images for pre-training in the classification of 3D lung CT images. By leveraging a mathematically driven dataset that avoids ethical concerns, this paper showed the potential of improving the model accuracy in medical imaging tasks, while mitigating the dependency on annotated CT images.

Acknowledgements

This is an accepted version of an article published in Artificial Life and Robotics. The final authenticated version is available online at: `https://doi.org/10.1007/s10015-025-01051-z`.

The authors would like to express our sincere gratitude to Dr. Nobuyasu Wakazono (Hokkaido University) and the members of Emina Study Project—a joint project between the Hokkaido University hospital and the Ebetsu City hospital—for their invaluable support. This study is supported by the grants from Ebetsu-city, Hokkaido, Japan.

References

[1] Afshar, P., Heidarian, S., Enshaei, N., Naderkhani, F., Rafiee, M.J., Oikonomou, A., Fard, F.B., Samimi, K., Plataniotis, K.N., Mohammadi, A.: COVID-CT-MD, COVID-19 computed tomography scan dataset applicable in machine learning and deep learning. Scientific Data **8**(1), 121 (2021). doi:10.1038/s41597-021-00900-3

[2] Arnab, A., Dehghani, M., Heigold, G., Sun, C., Lučić, M., Schmid, C.: ViViT: A video vision transformer. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 6816–6826 (2021). doi:10.1109/ICCV48922.2021.00676

[3] Bhatt, S.P., Washko, G.R., Hoffman, E.A., Newell Jr, J.D., Bodduluri, S., Diaz, A.A., Galban, C.J., Silverman, E.K., Estépar, R.S.J., Lynch, D.A.: Imaging advances in chronic obstructive pulmonary disease. insights from the genetic epidemiology of chronic obstructive pulmonary disease (COPDGene) study. American journal of respiratory and critical care medicine **199**(3), 286–301 (2019). doi:10.1164/rccm.201807-1351SO

[4] Castiglioni, I., Rundo, L., Codari, M., Di Leo, G., Salvatore, C., Interlenghi, M., Gallivanone, F., Cozzi, A., D'Amico, N.C., Sardanelli, F.: AI applications to medical images: From machine learning to deep learning. Physica Medica **83**, 9–24 (2021). doi:10.1016/j.ejmp.2021.02.006

[5] Doke, P.P.: Chronic respiratory disease: a rapidly emerging public health menace. Indian Journal of Public Health **67**(2), 192–196 (2023). doi:10.4103/ijph.ijph_726_23

[6] Fallahpoor, M., Chakraborty, S., Heshejin, M.T., Chegeni, H., Horry, M.J., Pradhan, B.: Generalizability assessment of COVID-19 3D CT data for deep learning-based disease detection. Computers in Biology and Medicine **145**, 105,464 (2022). doi:https://doi.org/10.1016/j.compbiomed.2022.105464

[7] Fourcade, A., Khonsari, R.: Deep learning in medical image analysis: A third eye for doctors. Journal of Stomatology, Oral and Maxillofacial Surgery **120**(4), 279–288 (2019). doi:10.1016/j.jormas.2019.06.002

[8] Gao, X., Khan, M.H.M., Hui, R., Tian, Z., Qian, Y., Gao, A., Baichoo, S.: COVID-VIT: Classification of covid-19 from 3D CT chest images based on vision transformer model. In: 2022 3rd International Conference on Next Generation Computing Applications (NextComp), pp. 1–4 (2022). doi:10.1109/NextComp55567.2022.9932246

[9] GBD 2019 Chronic Respiratory Diseases Collaborators: Global burden of chronic respiratory diseases and risk factors, 1990–2019: an update from the global burden of disease study 2019. eClinicalMedicine **59**, 101,936 (2023). doi:10.1016/j.eclinm.2023.101936

[10] Harada, K., Noguchi, W., Tamura, Y., Shimizu, K., Konno, S., Yamamoto, M.: Pre-training deep neural networks with 3D fractal structures for COPD stage classification. In: 30th International Symposium on Artificial Life and Robotics (AROB 30th 2025), pp. 640–645 (2025)

[11] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016). doi:10.1109/CVPR.2016.90

[12] Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2261–2269 (2017). doi:10.1109/CVPR.2017.243

[13] Huang, Y., Si, Y., Hu, B., Zhang, Y., Wu, S., Wu, D., Wang, Q.: Transformer-based factorized encoder for classification of pneumoconiosis on 3D CT images. Comput. Biol. Med. **150**(C) (2022). doi:10.1016/j.compbiomed.2022.106137

[14] Hutchison, J.E.: Fractals and self similarity. Indian University Mathematics Journal **30**(5), 713–747 (1981)

[15] Kataoka, H., Okayasu, K., Matsumoto, A., Yamagata, E., Yamada, R., Inoue, N., Nakamura, A., Satoh, Y.: Pre-training without natural images. International Journal of Computer Vision (IJCV) **130**, 990–1007 (2022). doi:10.1007/s11263-021-01555-8

[16] Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: 3rd International Conference on Learning Representations, ICLR 2015 (2015). doi:https://doi.org/10.48550/arXiv.1412.6980

[17] Liu, Z., Ning, J., Cao, Y., Wei, Y., Zhang, Z., Lin, S., Hu, H.: Video swin transformer. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3192–3201 (2022). doi:10.1109/CVPR52688.2022.00320

[18] Lynch, D.A., Austin, J.H.M., Hogg, J.C., Grenier, P.A., Kauczor, H.U., Bankier, A.A., Barr, R.G., Colby, T.V., Galvin, J.R., Gevenois, P.A., Coxson, H.O., Hoffman, E.A., Newell Jr, J.D., Pistolesi, M., Silverman, E.K., Crapo, J.D.C.: CT-Definable subtypes of chronic obstructive pulmonary disease: A statement of the fleischner society. Radiology **277**(1), 192–205 (2015). doi:10.1148/radiol.2015141579

[19] Perlin, K.: An image synthesizer. SIGGRAPH Comput. Graph. **19**(3), 287—-296 (1985). doi:10.1145/325165.325247

[20] Williams, L.H., Drew, T.: What do we know about volumetric medical image interpretation?: a review of the basic science and medical image perception literatures. Cognitive Research: Principles and Implications **4**, 21 (2019). doi:10.1186/s41235-019-0171-6

[21] Yamada, R., Kataoka, H., Chiba, N., Domae, Y., Ogata, T.: Point cloud pre-training with natural 3D structures. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 21,283–21,293 (2022). doi:10.1109/CVPR52688.2022.02060

[22] Yamada, R., Takahashi, R., Suzuki, R., Nakamura, A., Yoshiyasu, Y., Sagawa, R., Kataoka, H.: MV-FractalDB: Formula-driven supervised learning for multi-view image recognition. In: 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 2076–2083 (2021). doi:10.1109/IROS51168.2021.9635946