Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

PAPER

Fibro-CoSANet: pulmonary fibrosis prognosis prediction using a convolutional self attention network

RECEIVED
12 May 2021

REVISED
23 October 2021

ACCEPTED FOR PUBLICATION
4 November 2021

PUBLISHED
18 November 2021

Zabir Al Nazi[1], **Fazla Rabbi Mashrur**[2], **Md Amirul Islam**[3] and **Shumit Saha**[4,*]

[1] Brainekt AI Lab Dhaka, Bangladesh
[2] Department of Biomedical Engineering, Khulna University of Engineering & Technology, Khulna, Bangladesh
[3] Department of CS, Ryerson University; and Vector Institute for AI, Toronto, Canada
[4] Institute of Health Policy, Management, and Evaluation, Dalla Lana School of Public Health, University of Toronto and Centre for Global eHealth Innovation, Techna Institute, University Health Network, Toronto, Canada
[*] Author to whom any correspondence should be addressed.

E-mail: zabiralnazi@yahoo.com, rabbi.mashrur@gmail.com, amirul@cs.ryerson.ca and shumit.saha@mail.utoronto.ca

Keywords: pulmonary fibrosis, computed tomography (CT), convolutional neural network, self-attention, computer-aided diagnosis

Supplementary material for this article is available online

Abstract

Idiopathic pulmonary fibrosis (IPF) is a restrictive interstitial lung disease that causes lung function decline by lung tissue scarring. Although lung function decline is assessed by the forced vital capacity (FVC), determining the accurate progression of IPF remains a challenge. To address this challenge, we proposed Fibro-CoSANet, a novel end-to-end multi-modal learning based approach, to predict the FVC decline. Fibro-CoSANet utilized computed tomography images and demographic information in convolutional neural network frameworks with a stacked attention layer. Extensive experiments on the OSIC Pulmonary Fibrosis Progression Dataset demonstrated the superiority of our proposed Fibro-CoSANet by achieving new state-of-the-art modified Laplace log-likelihood score of −6.68. This network may benefit research areas concerned with designing networks to improve the prognostic accuracy of IPF. The source-code for Fibro-CoSANet is available at: https://github.com/zabir-nabil/Fibro-CoSANet.

1. Introduction

Idiopathic pulmonary fibrosis (IPF) is a chronic lung disease which is caused by forming scar tissue within the lungs (Paolo *et al* 2015). IPF leads to a gradual, irreversible deterioration of lung function by replacing the healthy lung tissues with scar tissue over time. IPF can potentially lead to rapid deterioration from long-term stability, which results in complete pulmonary dysfunction (Ganesh *et al* 2018). Due to the high variability in deterioration speed, management of pulmonary fibrosis relies on the decline in the lung function progression. Therefore, an accurate estimation of the lung function progression decline would lead to better management of IPF.

The current guideline for IPF diagnosis follows several procedures, such as surgical or transbronchial lung biopsy (Ganesh *et al* 2018). After the diagnosis, physicians often assess the decline of lung function by force vital capacity (FVC) using spirometry tests to monitor the prognosis of IPF. FVC measures the total amount of air exhaled after breathing in as deeply as possible (Zappala *et al* 2010). To assess the lung function, observing the FVC at intervals of six to twelve months is recommended (Ganesh *et al* 2018). While the FVC provides a general understanding of the prognosis of the IPF (Flaherty *et al* 2006), there are no widely used techniques to estimate the IPF progression. As such, due to the heterogeneous course of IPF, imaging modalities may provide valuable information regarding IPF prognosis.

Computed tomography (CT) images of the chest can be effectively used to assess the lung function decline from pulmonary fibrosis as the CT scans contain several visual signs essential for assessment by radiologists. Shi *et al* (2019) developed a voxel-wise radio-logical model using high-resolution CT scans and achieved

82.1% accuracy in predicting the progression of IPF. Furthermore, Salisbury *et al* (2016) utilized CT scans of IPF patients to predict the survival and FVC decline for 12 months with a significant correlation value of 0.6 between visual and predicted measurement. These studies have demonstrated the effectiveness of utilizing CT imaging as an important modality to predict the progression of pulmonary fibrosis. However, precisely predicting the progression of IPF from CT images remains challenging due to the high variability.

The recent advancements of artificial intelligence (e.g. convolutional neural networks (CNNs) (He *et al* 2016)) and the Kaggle: OSIC Pulmonary Fibrosis Progression Challenge (Kaggle) have significantly inspired to develop CT image based machine learning systems to obtain computer-aided clinical decision for IPF prognosis. In particular, Wong *et al* (2021) recently proposed Fibrosis-Net based on deep CNNs for predicting pulmonary fibrosis progression from chest CT images. Fibrosis-Net utilized the chest CT scans of a patient along with spirometry measurement and clinical metadata to predict the FVC of a patient at a specific time-point in the future (Wong *et al* 2021). While the existing CNNs based approaches have a higher capacity to predict pulmonary fibrosis progression from chest CT images, we strongly believe there is still room for improvement in terms of overall correctness. In this work, we argue that extracted convolutional features from chest CT scans along with patient's clinical or demographic features are not discriminative enough to correctly predict the FVC of a patient in cases where the network requires to focus on a specific region of the lung. To address this issue, we proposed a simple and efficient end-to-end multi-modal network, termed as Fibro-CoSANet, which utilized both the chest CT scan images and demographic information, such as sex, age, smoking history to predict the FVC of a patient at a specific time-point. Our proposed Fibro-CoSANet used a convolutional self-attention network that extracted features from a randomly selected CT image which are merged with the normalized demographics features. The merged features were then passed through a one-layer perceptron to obtain the predicted FVC. While the Fibrosis-Net (Wong *et al* 2021) utilized the multiple CT slices to generate convolutional features, we introduced an efficient formulation of the IPF prognosis task where we randomly selected a single CT image from multiple scans to extract convolutional features. However, we used the approximated lung volume information from all the available scans as a shallow feature which was merged with the convolutional features. In addition, we predicted the *slope* of FVC based on a *linear prior assumption* to reduce the computational overhead, while Fibrosis-Net (Wong *et al* 2021) used an elastic net to obtain the local FVCs.

We summarize our main contributions as follows:

- To the best of our knowledge, this is the first study that proposed a simple and efficient end-to-end multi-modal based convolutional self-attention network to predict the progress of IPF by utilizing the deeper CT and shallower demographic features.
- We introduced an intuitive and efficient way to apply a stacked self-attention layer on top of extracted convolutional CT features for further refinement and the advantages of this module are demonstrated with extensive experiments.
- We further introduced a unique formulation for FVC measurement of a patient where the goal of the proposed network was to predict the *slope* of the FVC trend.
- We showed, through extensive quantitative experiments under different settings, that our proposed Fibro-CoSANet achieved lower modified Laplace Log-Likelihood score than existing works on the publicly available Kaggle: OSIC Pulmonary Fibrosis Progression dataset.

2. Data preparation

In this section, we discuss the number of preprocessing steps conducted to prepare the training inputs and labels. We used the recently introduced Kaggle: OSIC Pulmonary Fibrosis Progression Challenge Benchmark Dataset which consists of CT scans, FVC measurements, and associated demographic features, such as age, sex, smoking status (Kaggle). As the main goal of our method was to predict the *slope* of the FVC trend of a patient, we first calculated the initial *slope* of FVC values using singular value decomposition (SVD) (Golub and Reinsch 1971) which were used as pseudo-labels in our proposed model (see figure 1 (3)). Then, we estimated the lung volume from the CT scans (Figure 1 (1)) followed by the extraction and normalization of demographics, such as age, sex, and smoking status (figure 1 (2)). Note that, we trained our proposed Fibro-CoSANet using a random CT image, estimated lung volume, age, sex, and smoking status.

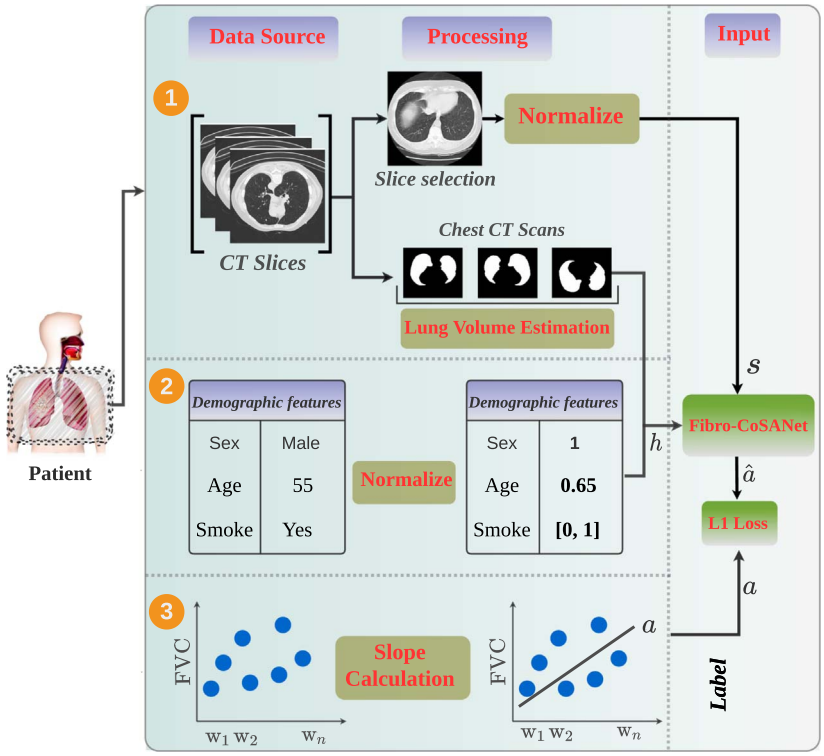

Figure 1. Illustration of the process of dataset preparation (①, ②) and ground-truth generation (③) for training Fibro-CoSANet. We randomly selected a CT slice and passed through the Fibro-CoSANet. We further normalized the demographic features and estimated the lung volume to fuse with the convolutional CT features. We calculated the *slope* of FVC using SVD least squares method explained in section 2.1 which is used as ground-truth to measure the correctness of the predicted FVC *slope*.

2.1. FVC formulation

We introduced a unique formulation to predict the *slope* of the FVCs by using the calculated initial *slope* of FVC values as ground-truth. First, we pre-processed the CT scans, $C \in \{c_i, c_{i+1}, \ldots, c_n\}$, where n refers to the number of patients. For each CT scan, $c_i \in \{s_1, s_2, \ldots, s_{m_i}\}$, we randomly selected a slice, s_k, from c_i for extracting features, where m_i is the number of slices in c_i and k is the index of the selected slice. The selected slice, s_k, was fed to a self-attention driven CNN model to extract pulmonary CT features. In parallel, we generated the demographics and volumetric feature sets, H, where $H \in \{h_1, h_2, \ldots, h_n\}$. Finally, these two sets of features were concatenated to predict the *slope* of FVC, $Z \in \{z_i, z_{i+1}, \ldots, z_n\}$, based on a linear priori assumption. Each patient data had FVC values, $z_i \in \{v_{w_1}, v_{w_2}, \ldots, v_{w_{t_i}}\}$, where t_i and w refers to the number of FVC values and representation of the corresponding weeks, respectively. We can formalize the FVC value of the ith patient in jth week as follows:

$$V_{w_j} = a_i \times w_j + V_{b_i}, \tag{1}$$

where, V_{b_i} is the base FVC and a_i is the *slope* of the ith patient. We can further extend equation (1) by expanding FVC along the week, upto j, as follows:

$$w_1 a_i + V_{b_i} = V_{w_1}, \qquad w_2 a_i + V_{bi} = V_{w_2}, \ldots\ldots, w_j a_i + v_{bi} = V_{w_j}. \tag{2}$$

For ease of presentation, we vectorized equation (2) as follows:

$$Ax = b \quad \text{where} \quad A = \begin{bmatrix} w_1 & 1 \\ w_2 & 1 \\ \vdots & \vdots \\ w_j & 1 \end{bmatrix}, \quad x = \begin{bmatrix} a_i & V_{b_i} \end{bmatrix}, \quad b = \begin{bmatrix} V_{w_1} \\ V_{w_2} \\ \vdots \\ V_{w_j} \end{bmatrix}. \tag{3}$$

Next, we decomposed the matrix $A_{j\times 2}$ into singular value form as:

$$A_{j\times 2} = U \Sigma V^{\top}, \tag{4}$$

where U and Σ refer to $j \times j$ orthogonal matrix and $j \times 2$ diagonal matrix, respectively. $V^{\top}$ is a 2×2 orthogonal matrix. We replaced A in equation (3) with (4) to achieve our desired least square solution (Golub and Reinsch 1971). The replacement operations can be formalized and solved using SVD (Golub and Reinsch 1971) as follows:

$$Ax = b, \qquad \tilde{x} = A^{+}b_{j\times 2} \tag{5}$$

where A^{+} is Moore–Penrose inverse of the matrix $A_{j\times 2}$ and $\tilde{x}$ minimizes our desired least square solution, $\|A\tilde{x} - b\|_2$. Thus, we calculated the *slope*, a_i, for a patient i, and used it as a pseudo ground-truth slope to train our network.

2.2. CT pre-processing and lung volume extraction

Contrary to natural images, CT scans consist of inconsistent and high-dimensional redundant information (Park *et al* 2020) which is computationally expensive to process and can result in poor performance. Therefore, to achieve a better signal-to-noise ratio, it is imperative to pre-process the CT scan images before feeding them to the CNN model. We applied the following pre-processing steps to resolve the issues.

Slice selection. Each patients' CT scan, c_i, contains many CT slices which represent the depth information of the lung. To reduce the computational complexity, we selected one slice per CT scan by the following operations: (i) we first truncated the first and last 15% of the slices as these slices contain minimal volume information. (ii) Then, we *randomly* selected one CT slice, s_k, from the middle to feed into the CNN.

For CT slice selection, first, we ran exploratory data analysis to visualize the slices and we observed that the most information (variance) was available in the slices in the selected range. Second, the order of the slices was not important as we randomly selected a slice from the middle to fed into the CNN. The intuition for selecting slices from the middle is that the slices are ordered, and the middle slices contain larger lung regions. Note that, we used all the slices per patient in the pre-processing step for volume calculation, while a single slice was used for CNN feature extraction. In addition, CNN is likely to learn from different slices as we randomly select a slice in each iteration during training.

Resizing and normalizing CT slices. We resized the randomly chosen CT slice based on the input specification (512×512) of the backbone CNN model. Further, to mitigate the inconsistency in tissue intensities across different scanners and improve convergence of the model, we normalized the pixel values using, $s' = (s - \lambda_b)/(\lambda_a - \lambda_b)$, where $\lambda_a = 2048$ and $\lambda_b = 0$ for any CT slice, s.

Note that for volume calculation, all of the slices were converted to the Hounsfield unit before lung mask generation. We applied the watershed algorithm on the masks to calculate the approximated lung volume for all the slices. The CT slices are stored in a 12 bit signed integer array (-2047 to 2048 range) (Regression with ct + tabular features [pytorch]). Also, we chose the normalization parameters based on the original range of the DICOM image.

Lung volume estimation from the CT slices. Along with demographics, we calculated the lung volume from CT images for each patient. The main motivation of using the lung volume estimation was to incorporate the approximated volume in the feature set, as we did not include all the CT slices to extract the volumetric features due to the computational complexity. Given a CT slice, s, we applied the *watershed* algorithm (Beucher 1979) to extract a segmentation map, p, of the lung for slice, s. The generated map, p, is a binary map between $\{0, 1\}$, where 1 and 0 indicate if a pixel belongs to the lung or not. Then, we generated the segmented lung image by simply multiplying the binary segmentation map with the original CT image. The procedure of generating the lung volume, v_i, for ith patient can be formalized as follows:

$$v_i = \sum_{j=1}^{m_i} p_j \times \delta_x \times \delta_y \times \delta_z, \tag{6}$$

where δ_x and δ_y are the pixel spacing in x and y directions respectively, and δ_z denotes the thickness of the slice.

2.3. Extracting demographic features

As IPF is associated with demographics, such as baseline FVC (Ganesh *et al* 2018, Ley *et al* 2012), age (GarcÃa-Sancho *et al* 2011, Ley *et al* 2012), gender (GarcÃa-Sancho *et al* 2011), smoking status (GarcÃa-Sancho *et al* 2011), we take inspiration from (Wong *et al* 2021) to incorporate these features along with CT image to improve the performance of our proposed method. We normalized the estimated lung volume, age, sex, and smoking status features using: $x' = \frac{x - \bar{x}}{\sigma}$ where, x is the raw numeric feature, $\bar{x}$ is the arithmetic mean of x, and σ is the standard deviation of x.

3. Network architecture of Fibro-CoSANet

We proposed a novel multi-modal convolutional self-attention network, Fibro-CoSANet, to predict the *slope* of FVC in an end-to-end manner. The overall pipeline of our proposed Fibro-CoSANet is illustrated in figure 2. Our proposed training framework consists of two key steps: (i) extraction of the deep features from the normalized CT image using a CNN with self-attention module (section 3.1), (ii) fusing the deep features

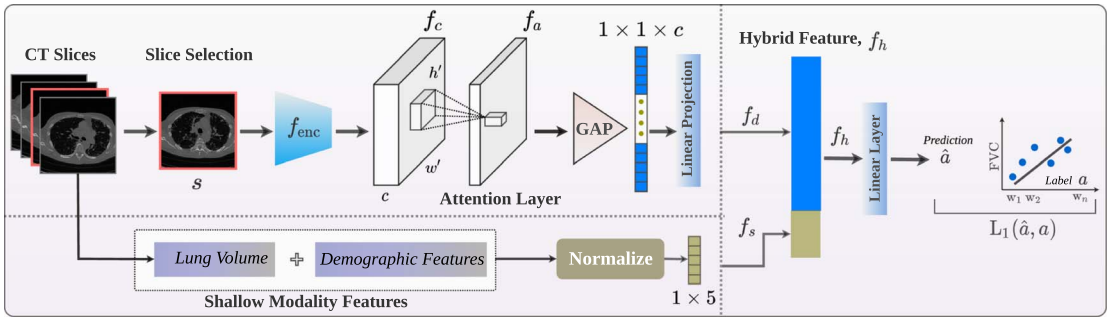

Figure 2. Illustration of the dual-stream pipeline of our proposed Fibro-CoSANet. We passed the randomly selected CT image to a pre-trained CNNs, f_{enc}, (e.g. ResNet18 (He *et al* 2016)) which produced a CT feature map, f_c. Then, we integrated a stacked self-attention module on top of the last convolutional layer of CNN to allow the network to focus on specific regions. The resultant output feature map, f_a, from the stacked self-attention module is passed through the global average pooling to produce the final representation, f_d, of the CT image. In parallel, we merged the demographics and estimated lung volume features, f_s with f_d which is further passed through a linear layer to predict the *slope* of FVC. We computed $L1$ loss between the predicted *slope*, $\hat{a}$ and pseudo ground-truth *slope*, a.

extracted from a CNN with shallow lung volume and demographic features followed by a fully-connected layer which predicts the *slope* of the FVC (section 3.2).

3.1. Convolutional self-attention network

In this section, we present our proposed convolutional self-attention network for extracting features from CT scan images with the ultimate goal of predicting the *slope* of FVC. Our proposed deep CT feature extractor network consists of two key components, (i) a CNN-based feature extractor network and (ii) a self-attention module which further refined the convolutional features extracted from the CNN. We first discuss the convolutional feature extractor network (section 3.1.1) followed by the self-attention module (section 3.1.2).

3.1.1. Deep CNN for CT feature extraction

In recent years, CNNs have been widely adopted for processing medical images (e.g. CT scans) (Sarvamangala and Kulkarni 2021). In general, CNN-based networks on medical imaging can be characterized as generic feature extractor networks which are termed encoder networks. The encoder network is simply a CNN that extracts features from a given input image. However, one downside of training CNNs is that it requires a huge amount of labeled training data to learn the millions of parameters involved in the network. This crucial issue limits the adoption of CNNs on medical image-based tasks as the majority of the datasets have a small volume of training data. To address this limitation, inspired by the existing works (Sajja *et al* 2019, Wang*et al* 2020), we fine-tuned the feature extractor CNN on CT scan images rather than training from scratch with random initialization. Let $s \in \mathbb{R}^{h \times w \times 1}$ be the input CT scan image (where h, w are the spatial dimensions). Given the input CT image, s, we adopted a CNN, f_{enc}, to extract a feature representation from the last convolutional layer of the CNN. Let $f_c \in \mathbb{R}^{h' \times w' \times c}$ be the extracted feature map which has smaller spatial dimensions than the original CT image, s. We used *ResNet* (He *et al* 2016), *ResNeXt* (Xie *et al* 2016), and *EfficientNet* (Tan and Le 2019) based architectures to build encoder networks in our study. We formalize the key operations as follows:

$$f_c = f_{\text{enc}}(\mathbf{W}_a * s), \tag{7}$$

where $\mathbf{W}_a$ denotes weights of the CNN model and $*$ denotes the convolutional operation. The extracted feature map, f_c, was fed to a self-attention module which further refines the feature representations before combining them with the demographics and lung volume features.

3.1.2. Self-attention module

The extracted feature map, f_c, was likely to capture high-level semantics of the CT images; however, allowing the network to focus on the specific region of the CT feature map was important to accurately predict the progression of IPF. Since IPF can result in honeycomb cysts in the lungs (Gruden 2016), these regions in the CT image require more attention than others. To focus on these regions of interest, we took inspiration from the existing works (Ramachandran *et al* 2019, Zhang *et al* 2019) and applied a self-attention module on top of the CNN feature extractor.

In the self attention module (figure 3), we first rearranged the extracted convolutional feature map, f_c, resulting in a feature map, $f_c' \in \mathbb{R}^{c' \times N}$ where c' is the number of channels and N is the product of h' and w'.

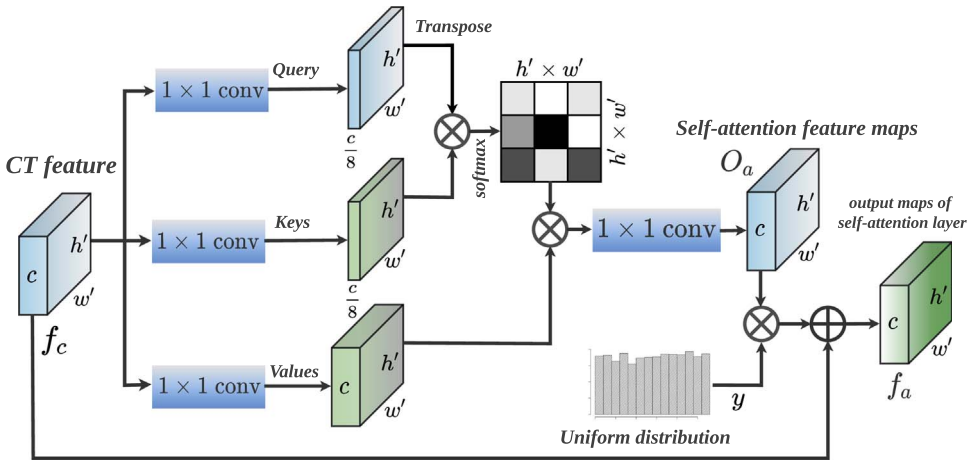

Figure 3. Overview of the proposed self-attention module for Fibro-CoSANet. Note that the stacking factor is one here.

Then we fed the feature map, f_c', to the self-attention layer (Zhang *et al* 2019), and obtained an attention map, $o_a \in \mathbb{R}^{c' \times N}$, with same dimension.

We further multiplied the output of the attention layer, o_a, by a scaling parameter, γ, and added with the input feature map, f_c' to obatin the self-attention map, f_a. We formalized the operations as follows:

$$f_a = \gamma^* o_a + f_c'. \tag{8}$$

Note that γ is a learnable scalar that is initialized from a uniform distribution in our work. The main advantage of learning γ was that it enabled the network to first focus on the local neighborhood indicators since it was easier. Then it eventually tried to assign more weight to the non-local region. Thus the module learned simpler tasks first to improve convergence (Zhang *et al* 2019). Finally, f_a was passed through an adaptive global average pooling operation (GAP), followed by a linear layer to obtain the deep CT feature set, f_d (see figure 2). We considered f_d as our final feature representation extracted from the CT scan image.

We augmented the CNN with the self-attention layer for realizing a richer effective receptive field and learning better feature representation as the recent work (Ramachandran *et al* 2019) has shown the advantages of applying a self-attention layer on top of convolutional feature representation. Unlike the previous self-attention-based works (Ramachandran *et al* 2019, Zhang *et al* 2019), we extended the existing idea by stacking l self-attention layers just before applying the GAP operation. In our implementation, l is the stacking factor and we considered l as a hyper-parameter. Note that, we placed the attention module between the last convolutional layer of CNNs and the pooling layer as convolutions were likely to better capture the low-level features while stand-alone attention layers may integrate global information by modeling long range pixel interactions (Ramachandran *et al* 2019). Furthermore, this placement reduced the computational complexity as the attention module was applied on a relatively low dimensional convolutional feature map.

3.2. Hybrid fusion of convolutional and shallow modality feature

Finally, we concatenated both the deep CNN features, f_d, extracted by the self-attention driven CNN backbone from the CT modality and the shallow modality features representation, f_s, to generate a hybrid multi-modal feature representation, f_h. The resultant feature representation, f_h, was passed to a fully connected layer to obtain the *slope*, *a*, of FVC, which was used to predict the patient's progression curve. We computed FVC from the predicted *slope*, *a* along the timeline, *w*, as follows:

$$\mathrm{FVC}(w_j) = a \times w_j + \mathrm{FVC}_b, \tag{9}$$

where FVC_b is the baseline force vital capacity and j is the week index.[5]

4. Experiments

We evaluated the effectiveness of our proposed approach for predicting the progression of pulmonary fibrosis and demonstrated the efficacy of the method under different settings. First, we showed the superiority of our proposed multi-modal learning pipeline followed by a comparison with recent approaches. Then, we evaluated

[5] Patient index i is used with variables when specifying a particular subject, for general case, the index is omitted for simplicity.

Table 1. Performance comparison between different modalities. Our proposed multi-modality based Fibro-CoSANet outperforms the single modalities (e.g. CT and demographics).

Mode	$LLL_m \downarrow$	$RMSE \downarrow$
Multi modality	$\mathbf{-6.68 \pm 0.31}$	$\mathbf{181.5 \pm 25.88}$
CT modality	-6.69 ± 0.28	184.16 ± 22.84
Demographics + Lung Volume	-6.75 ± 0.33	185.52 ± 22.89

our approach to generating baseline results under different backbone and metric settings to show generalizability and consistency. We further conducted an ablation study to investigate the necessity of each component of our proposed approach. Finally, we provided a comparison between different variants of our approach in terms of computational complexity, inference time, and total memory.

Dataset. In this study, we used the publicly available Kaggle: OSIC Pulmonary Fibrosis Progression challenge benchmark dataset provided by Open Source Imaging Consortium (OSIC) (Kaggle). The dataset consists of chest CT scans and associated demographics about fibrosis diagnosed patients. It contains 176 unique patients with a total of 1576 demographic information (multiple from the same patients) collected from numerous follow-up visits over the course of approximately 1–2 years. The demographics include the patient's *ID, weeks, FVC, percent, age, sex, and smoking status*. Note that the *weeks* represent the relative number of weeks pre or post from the baseline CT scan for each patient and we determined the time series of the weeks of a specific patient based on the patient's ID. For each patient, CT scan images (varies between 10 and 180) are provided in *DICOM* format files that contained meta-data about the patients and the scan. We used 5 fold cross-validation scheme to validate the best performance model. In the cross-validation setting, we carefully restrict to have no overlapping between the train and test splits of different subjects. The test set includes a baseline CT scan with *only* the initial FVC measurement for each patient.

Evaluation metric. We used a modified Laplace Log-Likelihood (LLL_m) and root mean square error ($RMSE$) metrics to report the performance of our proposed model. We choose LLL_m to evaluate a model's confidence in its decisions as it is designed to reflect both the accuracy and certainty of each prediction. For each true FVC measurement, we calculated the FVC and confidence measure as follows (Kaggle):

$$\begin{aligned} \sigma_{\text{clipped}} &= \max(\sigma, 70) \\ \Delta &= \min\left(|FVC_{\text{true}} - FVC_{\text{predicted}}|, 1000\right) \\ LLL_m &= -\frac{\sqrt{2}\,\Delta}{\sigma_{\text{clipped}}} - \ln\left(\sqrt{2}\,\sigma_{\text{clipped}}\right), \end{aligned} \tag{10}$$

where σ is the standard deviation and we threshold the error at 1000 ml to avoid the adverse penalty due to large errors. The confidence values were clipped at 70 ml to reflect the approximate measurement uncertainty in FVC. We calculated the final score by averaging the metric across all *weeks*. Note that, the calculated value of the metric was always negative, and *lower is better*.

Implementation details. We used publicly available PyTorch (Paszke *et al* 2019) framework to implement our proposed Fibro-CoSANet and an Intel(R) Xeon(R) Gold 5118 CPU with 187GB physical ram and an Nvidia Tesla V100 SXM2 (32GB) GPU to run our experiments. We trained models for 40 epochs using Adam optimizer with a decoupled weight decay regularization of 0.01. We initialized the backbone CNN by the ImageNet pre-trained model and optimized the network to minimize the $L1$ loss.

4.1. Results of proposed Fibro-CoSANet

We first conducted an ablation study to analyze the effectiveness of the multi-modality feature fusion technique by comparing it with other available modes. To demonstrate the superiority of our proposed *Multi-modal* training pipeline, we conducted experiments under three different *modes*: (i) **Multi modality**: convolutional features from CT images + shallow features (demographics + lung volume), (ii) **CT modality**: convolutional features from only CT modality, (iii) **Shallow Modality**: only lung volume and demographic features were used to train our model without any CNN backbone. We found that the multi-modality modes achieved higher performance than standalone CT modality and shallow modality in terms of LLL_m and $RMSE$ (table 1). These results suggested that demographics with lung volume or CT scans independently achieve reasonable performance while combing these two modalities improved the overall performance. Note that, the reported experimental results in the following sections are based on only multi modalities.

Table 2. Performance comparison of Fibro-CoSANet with recent works in terms of LLL_m and *RMSE*. Our proposed Fibro-CoSANet outperforms the existing state-of-the-art works on predicting the progression of pulmonary fibrosis.

Work	Regression type	$LLL_m \downarrow$	*RMSE* ↓
Fibrosis-Net (Wong *et al* 2021)	Elastic Net	−6.82	—
Mandal *et al* (Mandal *et al* 2020)	Quantile	−6.92	—
	Ridge	−6.81	—
	Elastic Net	−6.72	—
Fibro-CoSANet (**Ours**)	EfficientNet-b2	**−6.68 ± 0.31**	**181.5 ± 25.88**
	ResNet-50	**−6.68 ± 0.31**	**181.6 ± 22.89**
	EfficientNet-b3	**−6.68 ± 0.28**	182.58 ± 24.04
	EfficientNet-b1	**−6.68 ± 0.28**	183.96 ± 22.89

Table 3. Fibro-CoSANet results under different CNN backbone[a].

Backbone	$LLL_m \downarrow$ (CV)	*RMSE* ↓ (CV)
ResNet-18	−6.70 ± 0.29	183.68 ± 23.52
ResNet-34	−6.72 ± 0.28	185.18 ± 22.71
ResNet-50	−6.72 ± 0.27	186.52 ± 21.03
ResNet-101	−6.71 ± 0.25	188.92 ± 19.94
ResNet-152	−6.73 ± 0.28	186.19 ± 21.75
ResNeXt-50	−6.72 ± 0.27	186.39 ± 24.64
ResNeXt-101	−6.70 ± 0.26	184.04 ± 22.62
EfficientNet-b0	−6.70 ± 0.29	183.00 ± 23.60
EfficientNet-b1	−6.72 ± 0.31	183.22 ± 23.35
EfficientNet-b2	−6.74 ± 0.34	184.17 ± 22.89
EfficientNet-b3	−6.74 ± 0.34	184.17 ± 22.89
EfficientNet-b4	−6.70 ± 0.30	183.00 ± 22.42

[a] LLL_m = Modified Laplace Log Likelihood, RMSE = Root mean square error, CV = Cross valitaion (5-fold).

4.2. Comparison with recent approaches

We compared the overall performance of our proposed method with recent state-of-the-art approaches which predict the progression of pulmonary fibrosis (table 2). Mandal *et al* used *Multiple Quantile Regression*, *Ridge Regression*, and *Elastic Net Regression* to predict the progression, while Elastic Net Regression achieved the best result, achieving $-LLL_m$ of −6.72 (Mandal *et al* 2020). Wong *et al* achieved $-LLL_m$ of −6.82 (Wong *et al* 2021). Our proposed algorithm with *EfficientNet-b2* performed better than (Mandal *et al* 2020, Wong *et al* 2021), resulting in $-LLm$ of –6.68 and *RMSE* of 181 ± 25.88 (table 2). Also, the approximate complexity of our model was linear with respect to the number of patients as we processed all the information of a patient in a single mini-batch.

4.3. Baseline analysis of Fibro-CoSANet

We conducted an extensive experimental evaluation using widely-used CNNs, including *ResNet*, *ResNeXt*, and *EfficientNet* to show the consistency and generalizability of our proposed approach. We reported experimental results under two key variants of our proposed pipeline as follows:

Baseline model without self-attention module. We implemented the base model under different network backbones without any self-attention layer. To show the consistency and generalizability of our approach, we used 12 different CNNs architectures (five of ResNets, two of ResNeXts and five of EfficientNets) as the feature extractor for our proposed Fibro-CoSANet. Table 3 presents the baseline results in terms of LLL_m and *RMSE*. Interestingly, ResNets with lighter architecture (e.g. *ResNet-18-*LLL_m: −6.70 RMSE: 183.68) and *ResNeXt-101* achieved lower LLL_m and *RMSE* compared to the deeper ResNets. Furthermore, *EfficientNet-b0*, and *EfficientNet-b4* achieved comparative performance ($LLL_m \approx -6.70$ and *RMSE* ≈ 183.00) to *ResNet-18*. These results altogether suggested that our proposed Fibro-CoSANet with various backbones had the ability to predict FVC slope. The heavier models were prone to over-fitting as the size of the dataset was relatively smaller.

Baseline model with self-attention module. To further improve the overall performance, we introduced a stacked self-attention module (section 3.1.2) on top of the each CNN backbone (table 4). Here, we used fixed

Table 4. Performance of Fibro-CoSANet under different backbone with respect to attention module hyper-parameters. FS and SF refer to filter size and stacking factor, respectively. It is clear that stacking the self-attention module improves the overall performance.

Backbone	FS	SF	$LLL_m \downarrow$	$RMSE \downarrow$
EfficientNet-B0	32	3	-6.7 ± 0.29	183.7 ± 23.55
	32	5	-6.77 ± 0.31	185.63 ± 21.33
	64	1	-6.72 ± 0.34	182.13 ± 22.63
	128	3	-6.73 ± 0.33	183.67 ± 24.56
	128	5	-6.74 ± 0.36	183.57 ± 22.55
EfficientNet-B1	32	3	$\mathbf{-6.68 \pm 0.28}$	183.96 ± 22.89
	32	5	-6.7 ± 0.28	185.64 ± 24.25
	64	1	-6.71 ± 0.29	184.16 ± 24.78
	128	3	-6.79 ± 0.38	186.31 ± 24.79
	128	5	-6.69 ± 0.31	183.12 ± 22.05
EfficientNet-B2	32	3	$\mathbf{-6.68 \pm 0.31}$	$\mathbf{181.5 \pm 25.88}$
	32	5	-6.69 ± 0.3	183.39 ± 21.98
	64	1	-6.73 ± 0.28	184.71 ± 20.74
	128	3	-6.77 ± 0.33	187.13 ± 21.03
	128	5	-6.75 ± 0.33	18-6.03 $\pm$ 23.14
EfficientNet-B3	32	3	-6.72 ± 0.34	183.28 ± 22.87
	32	5	-6.74 ± 0.31	184.68 ± 21.05
	64	1	-6.71 ± 0.34	183.34 ± 22.57
	128	3	$\mathbf{-6.68 \pm 0.28}$	182.58 ± 24.04
	128	5	-6.72 ± 0.33	184.01 ± 24.4
EfficientNet-B4	32	3	-6.73 ± 0.3	184.86 ± 22.6
	32	5	$\mathbf{-6.68 \pm 0.3}$	183.45 ± 23.19
	64	1	-6.73 ± 0.39	182.29 ± 23.11
	128	3	-6.74 ± 0.32	184.35 ± 22.04
	128	5	-6.71 ± 0.28	184.06 ± 23.57
ResNet-18	32	3	-6.73 ± 0.32	184.71 ± 23.79
	32	5	-6.71 ± 0.31	183.79 ± 21.39
	64	1	-6.69 ± 0.28	183.84 ± 21.9
	128	3	-6.72 ± 0.27	184.96 ± 22.54
	128	5	-6.73 ± 0.35	185.29 ± 24.12
ResNet-34	32	3	-6.73 ± 0.31	184.79 ± 21.45
	32	5	-6.72 ± 0.28	185.4 ± 21.98
	64	1	-6.71 ± 0.31	183.3 ± 22.79
	128	3	-6.7 ± 0.29	183.32 ± 23.83
	128	5	-6.72 ± 0.28	184.33 ± 21.54
ResNet-50	32	3	-6.72 ± 0.31	184.07 ± 21.74
	32	5	-6.7 ± 0.28	184.94 ± 22.52
	64	1	-6.73 ± 0.27	185.46 ± 22.6
	128	3	$\mathbf{-6.68 \pm 0.31}$	$\mathbf{181.6 \pm 22.89}$
	128	5	-6.74 ± 0.34	185.06 ± 24.97
ResNet-101	32	3	-6.71 ± 0.3	183.13 ± 24.87
	32	5	-6.69 ± 0.28	184.17 ± 23.38
	64	1	-6.73 ± 0.28	185.33 ± 23.05
	128	3	-6.72 ± 0.3	185.53 ± 22.72
	128	5	-6.7 ± 0.3	183.63 ± 22.39
ResNet-152	32	3	-6.7 ± 0.29	183.89 ± 22.25
	32	5	-6.7 ± 0.3	183.64 ± 23.6
	64	1	-6.72 ± 0.33	183.05 ± 22.54
	128	3	-6.73 ± 0.35	184.15 ± 23.15
	128	5	-6.72 ± 0.33	182.92 ± 23.9
ResNeXt-50	32	3	-6.7 ± 0.27	184.75 ± 22.75
	32	5	-6.71 ± 0.3	184.61 ± 23.26
	64	1	-6.71 ± 0.32	183.84 ± 20.58
	128	3	-6.73 ± 0.28	185.22 ± 22.57
	128	5	-6.73 ± 0.32	184 ± 22.88
ResNeXt-101	32	3	-6.72 ± 0.28	184.01 ± 23.38
	32	5	-6.7 ± 0.3	183.46 ± 23.37
	64	1	-6.7 ± 0.31	182.5 ± 24.82
	128	3	-6.72 ± 0.3	184.9 ± 22.38
	128	5	-6.7 ± 0.28	183.57 ± 22.21

Table 5. Comparison of different variants of our proposed Fibro-CoSANet in terms of Macs (G), parameters (M), inference time (s), LLL_m, and *RMSE*.

Backbone	Macs (G)	Params (M)	Infer	$LLL_m \downarrow$	*RMSE* $\downarrow$
EfficientNet-B0	**0.07**	**4.05**	0.67	-6.7 ± 0.29	183.7 ± 23.55
EfficientNet-B1	0.1	6.65	0.7	$\mathbf{-6.68 \pm 0.28}$	183.96 ± 22.89
EfficientNet-B2	0.1	7.75	0.73	$\mathbf{-6.68 \pm 0.31}$	$\mathbf{181.5 \pm 25.88}$
EfficientNet-B3	0.14	10.75	0.7	-6.72 ± 0.34	183.28 ± 22.87
EfficientNet-B4	0.18	17.61	0.71	-6.73 ± 0.3	184.86 ± 22.6
ResNet-18	9.11	11.19	0.67	-6.73 ± 0.32	184.71 ± 23.79
ResNet-34	18.79	21.3	0.63	-6.73 ± 0.31	184.79 ± 21.45
ResNet-50	21.13	23.57	0.69	-6.7 ± 0.27	184.75 ± 22.75
ResNet-101	40.61	42.56	0.69	-6.71 ± 0.3	183.13 ± 24.87
ResNet-152	60.1	58.21	0.7	-6.7 ± 0.29	183.89 ± 22.25
ResNeXt-50	21.92	23.04	0.68	-6.7 ± 0.27	184.75 ± 22.75
ResNeXt-101	85.84	8-6.81	0.7	-6.72 ± 0.28	184.01 ± 23.38

output *channel* dimension (32) of CNN backbones with several attention filter sizes, such as 32, 64, and 128 along with a different number of stacking factors, such as 1, 3, and 5 to empirically identify the best combination that achieved superior performance. As shown in table 4, the overall performance was improved for most of the models with the addition of the self-attention layer. For instance, *EfficientNet*-B1, B2, B3, B4, and *ResNet-50* improved the overall performance by a considerable margin, resulting in ≈ -6.68 LLL_m. *EfficientNet*-B2 and *RestNet-50* achieved better score than other models in terms of *RMSE*(≈ 181). Comparing the results of different variants under various design choices, *EfficientNet-B2* achieved overall best performance (LLL_m: -6.68 ± 0.31 and *RMSE*: 181.5 ± 25.88) followed by *RestNet-50* (LLL_m: -6.68 ± 0.31 and *RMSE*: 181.6 ± 22.89) compared to other variants. We empirically found that *EfficientNet-B2* and *Reset-50* achieved the best performance with the attention filter size of 32 and 128, respectively, and three attention layers. ResNeXt-101 results were further not improved with the addition of self-attention later.

4.4. Performance analysis

We analyzed the overall performance of our proposed approach under two key aspects: (i) efficiency and (ii) computational complexity.

Efficiency. One of the important aspects of the high-volume biomedical data analysis is the latency or inference speed of the system. Our approach used a single CT image and shallow modality features to calculate the prognosis line from a single scalar, *a*. This simple *linear priori* assumption made the training and inference much faster, making our pipeline much efficient in handling a large amount of data. Note that, the training complexity depends on the number of patients.

Computational complexity of CNNs. Table 5 presents the comparison results of different baselines models in terms of the total number of parameters, inference time, and memory. Note that, we reported the best result for each CNN used in our experiments. *EfficientNet-B0* achieved the lowest computational complexity (0.07 GMacs, 4.05 million parameters, 0.67 s inference) compared to other CNNs backbones; however, failed to achieve superior performance. This could be due to the fact that the *EfficientNet-B0* architectures were relatively light-weight compared to ResNets. As *EfficientNet-B2* achieved the best result with relatively lower computational complexity (0.1 GMacs, 7.75 million parameters, 0.73 s inference), we termed *EfficientNet-B2* as the best network for FVC slope prediction.

One of the key motivations of our work was to optimize the model. Using more slices makes the network heavier, it also makes the network prone to overfitting. However, we experimented with more than one random slice while training the EfficientNet-b2 model. The results are reported in table 6.

The performance was not improved using more slices. Furthermore, when using seven slices, the network achieved significantly lower score than the best-reported score. We discuss the reason behind the performance degradation as follows:

- For three slices, we obtained similar performance as our best score (even though slightly higher RMSE), but for a higher number of slices, we did not observe any improvement in the score. EfficientNet is trained with three channels on natural images (Tan and Le 2019). In the original training setup, the channels contain similar information but in different color domains. We used these models, as the pre-trained weight already had better feature extraction capabilities. However, it has been suggested in the literature that such 2D convolutional models might not be ideal for large 3D volumetric data (Yang *et al* 2021). So, an increasing number of channels might not result in likewise performance. Note that, we could not experiment with 3D

Table 6. Performance of EfficientNet-b2 with self-attention under different number of slices.

Backbone	Num. of slices	LLR	RMSE
EfficientNet-b2 + Self-Attention	3	-6.68 ± 0.34	183.51 ± 23.79
EfficientNet-b2 +Self-Attention	5	-6.69 ± 0.32	185.25 ± 24.37
EfficientNet-b2 + Self-Attention	7	-6.74 ± 0.30	184.20 ± 25.04

ConvNets as our work depends on 2D ConvNets. Comparing 2D and 3D ConvNets would be extremely challenging in this context.

- The dataset for the experiment is not large, so it is easy for the models to get over-fitted on the training samples. We followed a strict testing policy so that none of the patients in the test contained in the training data. So, the network must generalize over the patients to perform well on the test data. As we used a single CT scan for each patient, increasing the number of slices may force the network to focus too much on the CT data (multiple similar slices) at a time resulting in overfitting.

5. Discussion and conclusion

We proposed a novel multi-modal convolutional self-attention-based learning pipeline to predict the prognosis of IPF. To the best of our knowledge, this work was one of the earliest attempts that incorporated both CT scan and demographic information in an end-to-end manner. Furthermore, we integrated a self-attention layer on top of the CNNs to further refine the convolutional features by allowing the network to focus on a specific region of the CT scan image. Moreover, we predicted the slope of the FVC trend of a patient based on a simple linear prior assumption. Extensive experiments demonstrated the superiority of our proposed approach over the recent models tested on the same dataset (Mandal *et al* 2020, Wong *et al* 2021).

We would like emphasise the fact that, for each patient, there is only single CT scan included in the dataset. The CT scan was recorded in a random week during study period. Therefore, the FVC decline must be predicted with a single CT scan which includes a few slices. As mentioned earlier, we used all of the slices in the pre-processing step to calculate the volume from the CT slices. We also included other features calculated from all of the slices (mean, skew, kurtosis); however, when we incorporated these features with the ConvNet, it achieved poor results. This is why we only used volume which contains information from all of the slices of the CT scan from a patient. Even though we used all the CT slices in the pre-processing stage to calculate the volume and used it as a feature in the neural network training, we used a single random slice for training the ConvNet. There are multiple reasons for this setup.

- The number of slices in the patients varied significantly in the dataset, and we observed low variance in the adjacent slices. So, it was hard to choose a fixed optimal number of slices for training the CNN.
- Using all the slices from the CT scan was also not technically possible as there were a different number of slices per patient and it significantly differs. We used a 2D CNN (as training 3D CNN with this limited data would be extremely challenging) which required to use the slices channel-wise. Moreover, we experimented with multiple slices channel-wise, and results showed that the model achieves lower performance as the number of slices increases due to overfitting.

We have reported results under two metrics, Laplace Log-Likelihood Score and the RMSE. The metrics reported in the paper are based on the public data. We have used 20% of the patients for testing with 5-fold cross-validation and reported the average of metrics on the five test folds. The private test data is not publicly available and requires an end to end notebook to submit in the competition. Therefore, we could not perform the pre-processing steps for volume calculation and RMSE calculation. Thus, the metric from the private leaderboard could not be reported. Furthermore, a direct comparison in terms of parameters and inference time is not possible as the code of the baseline models are not publicly available. However, we strongly believe our pipeline is much simpler than the baselines as we only considered one random CT slice during training the CNN. We claim the efficacy of our proposed pipeline from two aspect. First of all, we avoided the high-volume biomedical data analysis which generally increase the latency or inference speed of the system. In our case we used a single CT image. Secondly, our prior linear assumption made the training and inference much faster, making our pipeline much efficient in handling a large amount of data. Finally, we compared the number of

parameters and inference time within the framework of multiple CNN backbones and attention blocks to demonstrate the complexity of our proposed pipeline.

Despite the impressive performance, one major limitation of our proposed approach was that the prognosis of pulmonary fibrosis was of linear nature. This assumption limited us to predict the actual FVC values at each temporal point. We used the linear assumption to regularize our model and to avoid overfitting. As we have discussed, while testing, for each patient, only a single CT scan is provided with a baseline FVC. It is important to make sure that the model is not overfitted on the training data. Besides, it is extremely challenging to predict the local FVC for each week from a single CT scan. So, we simplified our model by predicting only a single slope from the CNN, which suggested the overall progress or decline of the FVC. Our model also suggests a simple metric, the slope to denote the patients' condition. For example, if the slope is positive, it suggests an improvement in FVC (progress), whereas a negative slope suggests an FVC decline. Furthermore, Fibro-CoSANet failed to produce better performance with deeper architectures, which could be due to the relatively small sample size. Finally, throughout our experiments, we used a fixed set of hyper-parameters. The overall performance for each backbone can be further improved by a careful selection of the best possible hyper-parameters.

In conclusion, we aimed to provide a framework to the research community that can be used on a larger dataset and clinical trial in the future. As accurate progression prediction of IPF patients is crucial for the effective treatment and IPF based datasets are rarely available, our proposed algorithm could shed light on the new approaches to build trustworthy algorithms for IPF prognosis.

ORCID iDs

Fazla Rabbi Mashrur https://orcid.org/0000-0002-5832-7044
Shumit Saha https://orcid.org/0000-0003-2650-084X

References

Paolo S, Nicola S, Giulio R, Alberto C, Argyris T, Bruno C and Carlo V 2015 Idiopathic pulmonary fibrosis: an update *Ann. Med.* **47** 15–27
Ganesh R *et al* 2018 Diagnosis of idiopathic pulmonary fibrosis. an official ats/ers/jrs/alat clinical practice guideline *Am. J. Respiratory Crit. Care Med.* **198** e44–68
Flaherty K R *et al* 2006 Idiopathic pulmonary fibrosis: prognostic value of changes in physiology and six-minute-walk test *Am. J. Respiratory Crit. Care Med.* **174** 803–9
Shi Y *et al* 2019 Prediction of progression in idiopathic pulmonary fibrosis using CT scans at baseline: A quantum particle swarm optimization-Random forest approach *Artif. Intell. Med.* **100** 101709–18
Salisbury M L *et al* 2016 Idiopathic pulmonary fibrosis: the association between the adaptive multiple features method and fibrosis outcomes *Am. J. Respiratory Crit. Care Med.* **195** 921–9
He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition* pp 770–8
Kaggle OSIC Pulmonary Fibrosis Progression
Wong A, Lu J, Dorfman A, McInnis P, Famouri M, Manary D, Lee J R H and Lynch M 2021 Fibrosis-Net: a tailored deep convolutional neural network design for prediction of pulmonary fibrosis progression from chest CT images arXiv:2103.04008 [cs, eess]
Golub G H and Reinsch C 1971 Singular value decomposition and least squares solutions *Handbook for Automatic Computation: Volume II: Linear Algebra, Die Grundlehren der mathematischen Wissenschaften* ed J H Wilkinson *et al* (Berlin: Springer) pp 134–51
Park S *et al* 2020 Annotated normal ct data of the abdomen for deep learning: Challenges and strategies for implementation *Diagn. Interventional Imaging* **101** 35–44
Regression with ct + tabular features [pytorch] (https://kaggle.com/furcifer/q-regression-with-ct-tabular-features-pytorch)
Beucher S 1979 Use of watersheds in contour detection *Proc. of the Int. Workshop on Image Processing. CCETT*
Ley B *et al* 2012 A multidimensional index and staging system for idiopathic pulmonary fibrosis *Ann. Intern. Med.* **156** 684–91
GarcÃa-Sancho C *et al* 2011 Familial pulmonary fibrosis is the strongest risk factor for idiopathic pulmonary fibrosis *Respiratory Med.* **105** 1902–7
He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition *Proc. IEEE Conf on Comput Vision and Pattern Recognit* pp 770–8
Sarvamangala D R and Kulkarni R V 2021 Convolutional neural networks in medical image understanding: a survey *Evol. Intell.* **14** 1–22
Sajja T, Devarapalli R and Kalluri H 2019 Lung cancer detection based on ct scan images by using deep transfer learning *Trait. Signal* **36** 339–44
Wang S, Dong L, Wang X and Wang X 2020 Classification of pathological types of lung cancer from ct images by deep residual neural networks with transfer learning strategy *Open Med.* **15** 190–7
Xie S, Girshick R, Dollàr P, Tu Z and He K 2016 Aggregated residual transformations for deep neural networks *Proceedings of the IEEE conference on computer vision and pattern recognition (Honolulu, HI, 21–26 July 2017)* (Picastaway, NJ: IEEE) pp 1492–500
Tan M and Le Q 2019 Efficientnet: Rethinking model scaling for convolutional neural networks *Int. Conf. on Mach. Learn.* pp 6105–14 PMLR
Gruden J F 2016 CT in idiopathic pulmonary fibrosis: diagnosis and beyond *Am. J. Roentgenol.* **206** 495–507
Ramachandran P, Parmar N, Vaswani A, Bello I, Levskaya A and Shlens J 2019 Stand-alone self-attention in vision models arXiv:1906.05909
Zhang H, Goodfellow I, Metaxas D and Odena A 2019 Self-Attention Generative Adversarial Networks *International conference on machine learning* (PMLR) 7354–7363

Paszke A *et al* 2019 Pytorch: An imperative style, high-performance deep learning library *Advances in Neural Information Processing Systems* ed H Wallach *et al* 32 (Red Hook, New York: Curran Associates, Inc.) pp 8024–35

Mandal S, Balas V E, Shaw R N and Ghosh A 2020 Prediction analysis of idiopathic pulmonary fibrosis progression from osic dataset *IEEE International Conference on Computing, Power and Communication Technologies (GUCON)* (Greater Noida, India: IEEE) pp 861–5

Yang J, Huang X, He Y, Xu J, Yang C, Xu G and Ni B 2021 Reinventing 2d convolutions for 3d images *IEEE J. Biomed. Health Inform.* **25** 3009–18

Zappala C J, Latsi P I, Nicholson A G, Colby T V, Cramer D, Renzoni E A and Hansell D M 2010 RM Du Bois, and AU Wells. Marginal decline in forced vital capacity is associated with a poor outcome in idiopathic pulmonary fibrosis *Eur. Respiratory J.* **35** 830–6