# DABC-Net for robust pneumonia segmentation and prediction of COVID-19 progression on chest CT scans

**Xiao-Yong Zhang** ( xiaoyong_zhang@fudan.edu.cn )
Fudan University https://orcid.org/0000-0001-8965-1077
**Ziqi Yu**
Fudan University
**Xiaoyang Han**
Fudan University https://orcid.org/0000-0002-3007-6079
**Botao Zhao**
Fudan University
**Yaoyao Zhuo**
Fudan University
**Yan Ren**
Fudan University
**Xiangyang Xue**
Fudan University
**Lorenz Lamm**
Helmholtz Zentrum Munich
**Jianfeng Feng**
Fudan University
**Carsten Marr**
Institute of Computational Biology, Helmholtz Zentrum München-German Research Center for Environmental Health, Neuherberg https://orcid.org/0000-0003-2154-4552
**Feng Shan**
Fudan University
**Tingying Peng**
Helmholtz Center Munich

**Research Article**

**Keywords:** COVID-19, Deep learning, Model uncertainty, Pneumonia segmentation, Progression prediction

**Posted Date:** December 11th, 2020

**DOI:** https://doi.org/10.21203/rs.3.rs-114267/v1

# DABC-Net for robust pneumonia segmentation and prediction of COVID-19 progression on chest CT scans

Ziqi Yu[1,2 #], Xiaoyang Han[1,2 #], Botao Zhao[1,2], Yaoyao Zhuo[3], Yan Ren[4], Xiangyang Xue[5], Lorenz Lamm[6,7], Jianfeng Feng[1,2], Carsten Marr[6], Fei Shan[3 *], Tingying Peng[6,7 *], Xiao-Yong Zhang[1,2 *]

[1]Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, 200433, P. R. China

[2]Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence, Ministry of Education, Shanghai 200433, P. R. China

[3]Department of Radiology, Shanghai Public Health Clinical Center, Fudan University, Shanghai 201508, China.

[4] Department of Radiology, Huashan Hospital, Fudan University, Shanghai 200433, China

[5] Shanghai Key Laboratory of Intelligent Information Processing, School of Computer Science, Fudan University, Shanghai 200433, China.

[6] Institute of Computational Biology, Helmholtz Zentrum München, Ingolstädter Landstraße 1, D-85764

[7] Helmholtz AI, Helmholtz Zentrum München, Ingolstädter Landstraße 1, D-85764 Neuherberg, Germany

[#]These authors contributed equally.

*Correspondence should be addressed to X.Y.Z. (xiaoyong_zhang@fudan.edu.cn) or T.P. (tingying.peng@tum.de) or F.S. (shanfei@shphc.org.cn)

## Abstract

Currently, reliable, robust and ready-to-use CT-based tools for prediction of COVID-19 progression are still lacking. To address this problem, we present DABC-Net, a novel deep learning (DL) tool that combines a 2D U-net for intra-slice spatial information processing, and a recurrent LSTM network to leverage inter-slice context, for automatic volumetric segmentation of lung and pneumonia lesions. We evaluate DABC-Net on more than 10,000 radiologists-labeled CT slices from four different cohorts. Compared to state-of-the-art segmentation tools, DABC-Net is much faster, more robust, and able to estimate segmentation uncertainty. Based only on the first two CT scans within 3 days after admission from 656 longitudinal CT scans, the AUC of our DBAC-Net for disease progression prediction reaches 93%. We release our tool as a GUI for patient-specific prediction of pneumonia progression, to provide clinicians with additional assistance to triage patients at early days after the diagnosis and to optimize the assignment of limited medical resources, which is of particular importance in current critical COVID-19 pandemic.

Keywords: COVID-19, Deep learning, Model uncertainty, Pneumonia segmentation, Progression prediction

## Introduction

The Coronavirus Disease 2019 (COVID-19) has infected more than 45 million people worldwide (as of 31th October 2020) and caused more than 1.1 million deaths. In practice, most people infected with COVID-19 have mild cold-like symptoms, while others may evolve into serious illnesses that require intensive medical treatment or even lead to death. To maximize the distribution of available medical resources and save lives, it is vital to predict the progression of COVID-19 for each individual at the early days following diagnosis.

Chest computed tomography (CT) plays an important role for evaluating COVID-19 patients by showing specific image features such as ground-glass opacification and consolidation[1]. So far, deep learning (DL) based analyses on chest CT are mostly concentrated on COVID-19 diagnosis or classification, e.g. differentiating COVID-19 positive patients from patients with normal pneumonia (see[2] for an overview). CT-based diagnosis was essential at the beginning of the pandemic, when RT-PCR, the gold-standard of COVID-19 diagnosis, was relatively slow and often in short supply. Yet the implementation of large-scale RT-PCR test reduces the need for using radiology images for screening purposes. Nevertheless, radiology images remain vital to detect pulmonary involvement in COVID-19 patients. Particularly, it has been suggested that CT assessment of lesions can be used as an imaging surrogate for disease burden and to identify severe patients in need of hospital admission[3,4]. Yet the lesion quantification in both these studies relies on laborious manual examination of CT images by experienced radiologists and is hence difficult to be integrated into a standard clinical workflow, in particular as a rapid outbreak can bring enormous pressure on radiologists in terms of speed and number of examinations.

To address these problems, we propose a novel DL based tool to process and quantify COVID-19 induced pneumonia progression using chest CT images, thereby producing a progression score to assist clinicians in triaging patients (Fig. 1a). The core of our tool is a Dual spatial and channel Attention Bidirectional ConvLSTM Net (DABC-Net, Fig. 1b), for automatic segmentation of regions of interest, i.e. lung and pneumonia lesions. Compared to the state-of-the-art 3D U-Net that uses 3D convolutions to process the spatial context in an isotropic fashion, DABC-Net uses a 2D U-Net for intra-slice information and a ConvLSTM for inter-slice context (see Online Methods for details), which is more robust for anisotropic CT images where the slice thickness (z-axis resolution) does not match the intra-slice (x-y) resolution. Furthermore, we add a spatial attention mechanism and a channel attention mechanism, which reduces the number of ConvLSTM units and hence the overall parameter size of DABC-Net, making it faster and more resistant to overfitting. To demonstrate model generalisability, we evaluate DABC-Net on four different cohorts, which is more comprehensive than previous studies where models are trained and evaluated on data that originates from the same hospital [2,5]. We also assess the model uncertainty using Monte Carlo dropout[6] and suggest expert recheck on samples associated with high uncertainty to prevent the misuse of our model. Last but not least, we release our DABC-Net as an open-source and ready-to-use toolkit, which leads to real-time CT segmentation of COVID lesions and prediction of disease progression.

## Results

*DABC-Net is faster and more robust than other methods in lung and COVID-19 pneumonia segmentation*

Since accurate segmentation is the key step for disease progression prediction, our first motivation is to evaluate segmentation performance of DABC-Net along with four state-of-the-art segmentation methods: DeepLabV3[7], Sensor3D[8], nnUNet2D and nnUNet3D[9] (a short comparison of all methods is provided in Table S1; see Online Methods for details of all methods) on four datasets:

- Coronacases: 2581 slices of 10 COVID patients from Wenzhou Medical University, obtained from \url{https://coronacases.org}
- Radiopedia: 1389 slices of 19 patients from multiple hospitals, obtained from \url{https://radiopaedia.org}
- Wuhan: 3805 slices of 27 patients from Wuhan hospitals
- Shanghai: 1450 slices of 23 patients from Shanghai Public Health Clinical Center.

A summary of the complete data information is shown in Table S2. Annotations of lung and COVID-19 lesions of the Coronacases dataset and parts of the Radiopedia dataset are publicly available[10], which we use to train our model and to evaluate it with a five-fold cross-validation. We annotate the remaining data by two experienced radiologists and use it for testing only. In all four datasets, DABC-Net achieves better Dice scores than other methods for both lung and lesion segmentation (Fig. 2a,b, Fig. S1), suggesting that DABC-Net is robust for variations in lesion size, CT image intensities, and slice thickness. By contrast, the performance of a full 3D convolution deteriorates with increasing slice thickness (Fig. S2) and could completely fail when the information of slice thickness is missing and the interpolation along the z-axis is not properly done (Fig. 2b, nnUNet3D). Moreover, compared to other methods, DABC-Net is three times faster than the second fastest method, Sensor3D, and hundreds times faster than nnUNet (Fig. 2c). This is an important advantage, as local hospitals generally face enormous time pressure when a COVID-19 wave hits a region.

*DABC-Net estimates segmentation uncertainties*

As any other neural network, our DABC-Net can result in bias for lesion segmentation when there is a domain shift between training and testing data. Hence, we estimate the uncertainty of DABC-Net predictions in a pixel-wise fashion (Fig. 2d), using an approximate Bayesian inference by Monte Carlo dropout[11]. More specifically, we consider two types of uncertainty measures here, i.e. epistemic uncertainty (also known as model uncertainty) that is raised when the input sample is outside of the training distribution, and aleatory uncertainty (also known as data uncertainty) that is caused by the intrinsic randomness of the real data generating process (Fig. S3, the derivation of both uncertainty measures can be found in Online Methods). We find a strong negative correlation between uncertainty and prediction accuracy, which suggests that the uncertainty estimation can be used to infer prediction accuracy in the real testing scenario when the ground-truth segmentation is not available (Fig. 2e, Fig. S4).

*Based on DABC-Net segmentation, we can predict the disease progression*

From Shanghai Public Health Clinical Center, we obtain 656 longitudinally measured CT scans from 117 patients (usually one CT scan per three days is taken, see Table S3 and Fig. S5 for more details). Each scan is classified according to the clinical diagnosis as mild vs. severe status. It should be noted that these labels are not based solely on the CT appearance, but rather indicate the severeness of the clinical symptoms of the patient. Based on the segmentation results of DABC-Net, we can easily visualise the pneumonia progression for individual patients plotting their lesion volume ratios (lesion volume/lung volume) over time (Fig. 3a). Despite large individual variabilities, patients who progress into the severe status generally show a sharp increase of lesion volume ratio in the first week of hospitalisation, reaching a peak around day 7-10 before slowly recovering in the following days (Fig. 3b). By contrast, patients who only show mild symptoms during their hospitalisation have a consistently lower lesion volume ratio (Fig. 3b, Fig. S6).

Based on DABC-Net segmentation results, we aim to predict disease progression, i.e. whether one patient will develop into a severe status, using extracted features from CT images. Besides lung and lesion regions outputted by DABC-Net, we further delineate the area of consolidations by intensity thresholding and morphological operations as additional features (see Methods). Consolidations were found more common in patients >50 years old[12] and could be a warning sign of a severe progression. Using ensemble learning of multiple classifiers on features extracted from the first two CT scans, we achieve an accurate prediction with an area under the receiver operating characteristic (ROC) curve (AUC) of 0.93 (Fig. 3c). We can further improve our prediction by adding additional scans and reach an AUC of 0.97 with the first three scans (Fig. S6). Even with only the first CT scan, we can still achieve an AUC of 0.84 (Fig. S7). It should be noted that the prediction accuracy is strongly affected by CT segmentation quality. For example, when we use DeepLabv3+ instead of DABC-Net for lung and lesion segmentation, we get AUCs of only 0.76, 0.88 and 0.90 for the first scan, first two-scan and three-scan predictions, respectively, which correspond to a 6-8% drop in performance (Figs 3e, S6, S7, Table S4). Practically, we can use the prediction of disease progression as guidance for hospital triage to distribute the limited hospital beds to patients who are predicted to have a severe disease progression. By moving along the ROC curve, we can adapt the trade-off between sensitivity and specificity to the number of available hospital beds. For example, we can aim at more sensitivity if we have an abundant number of beds, but have to focus on more specificity if we have a restricted number. For explainability, we identify the 10 most features that contribute to the prediction (Fig. 3d). In addition to the CT-related features, we also find age to be a key factor that leads to different disease progression patterns, consistent with previous studies [13,14].

## Discussion

How to optimise the triage of COVID-19 patients is a critical issue in clinics during this world-wide pandemic. So far, several machine learning studies have addressed this issue and attempt to assess the risk of critical illness for COVID-19 patients at the hospital admission [4,15,16]. Although these studies are very promising to identify patients at high risk, they usually rely on clinical

measurements from laboratories for their prediction, which can be laborious and delayed in time. By contrast, our triage system is based on chest CT examination, which is widely available in many clinics, relatively fast and can be easily integrated into a routine examination workflow for COVID-19 patients. Therefore, our tool is fast and complementary to existing AI tools for COVID-19 triage and prognosis.

We found that our prediction accuracy increased from 84% to 97% as we increased the number of included CT scans from the first single scan to the first three scans (Figs 3e, S6, S7). This demonstrates the necessity of continuous monitoring of pneumonia progression as some patients can have rapid progression in a short time. E.g. patient 1 in Fig. 3a shows a fast enlargement of lesion volume ratio from less than 2.5% (scan I and II) to 15% (scan III) within 3 days. Such a dynamic progression will make it difficult for prediction based on only a single CT scan at hospital admission, e.g. other studies that use single CT scans for severity prediction achieve a 75-85% AUC [4,17]. As a comparison, by measuring a second CT scan with a three-day interval, we can achieve a much more reliable prediction with 93% AUC.

We also found that our new volumetric segmentation algorithm, DABC-Net, increases the prediction accuracy by 6~8% over the current state-of-the-art method, DeepLabV3[5,18]. Yet DeepLabV3 is a pure 2D segmentation algorithm that neglects the rich inter-slice contexts in CT scans. By contrast, DABC-Net is a hybrid 2D-3D segmentation network that combines classical 2D intra-slice feature extraction and a bidirectional ConvLSTM network for inter-slice feature learning, yielding an improved performance. On the other hand, compared to full 3D volumetric segmentation such as 3D U-Net, DABC-Net is several orders of magnitude faster, lower in memory consumption, and robust with respect to thick-slice CT as well as thin-slice CT. Hence it is more generalisable in clinical settings where different CT scanners and computing resources are used.

In clinics, although chest X-ray examination is the predominant method for screening lung diseases because of easy access with low-cost, it can only provide 2-D images and can not show the exact location and the volume of lesions within the lung. As a tomography method with high resolution and the ability to generate 3-D views, CT scans, substituted for X-rays, can produce quantitative information for both lungs and lesions, which play an important role in diagnosing and monitoring COVID-19 patients. Therefore, our CT-based tool would facilitate the clinical workflow to fight COVID-19.

We release our segmentation and prediction tools as open-source code (https://robin970822.github.io/DABC-Net-for-COVID-19/ ) as well as a GUI shown in Fig. S8 (https://github.com/Robin970822/DABC-Net-for-COVID-19/tree/master/APP） that facilities the usage without a deep learning and coding background. In the future, we will explore the combination of CT features and other clinical features to further improve the prediction accuracy. Another potential usage of DABC-Net based CT quantification is to evaluate the efficiency of different treatments.

## Methods

### *DABC-Net Architecture*

DABC-Net is a hybrid 2D-3D network, which combines a U-shaped network (UNet) with shared-weighted encoder and decoder to process in-plane context and a DABC-Module that uses dual attention bidirectional convolutional LSTM to integrate cross-plane context. We explain both components in detail below.

### *U-Net with share-weighted encoder and decoder to process in-plane context*

2D U-Net has been successfully applied in many medical image segmentation tasks[19]. It has an encoder that contracts the original input and a symmetrical decoder to restore the compressed feature maps into the initial size. In our DABC-Net, we use a share-weighted convolution operator in the encoding and decoding paths of 2D U-Net to extract intra-slice features. More specifically, each convolution and transposed convolution block consists of two $3 \times 3$ shared convolution filters followed by a $2 \times 2$ max pooling layer and ReLU function. These parameters are shared between all input slices.

### *DABC-Module uses bidirectional convolutional LSTM to integrate cross-plan context*

Although the 2D U-Net structure works well on processing in-plane context, it does not account for the context along the z-axis that could be also useful for semantic segmentation. Alternatively, 3D U-Net was proposed to process 3D volumetric data [20]. However, due to the memory limits, 3D U-Net usually uses a small patch size, which only covers a very small fraction of an image and thus cannot sufficiently capture large-scale context. In this case, the original high-resolution xy in-plane information becomes fragmented in 3D U-Net. Therefore, in our approach, we use a convolutional long-short term memory (C-LSTM) [21] to integrate in-plane features extracted by 2D U-Net, with methodological details given below:

Unlike the standard LSTM that considers the input information as vectors, C-LSTM keeps abundant spatial semantic features by replacing the matrix multiplication with the convolution operator in input and recurrent transformations. More specifically, C-LSTM consists of a memory cell $c_t$, an input gate $i_t$, a forget gate $f_t$, and an output gate $o_t$ and can be expressed as:

$$
\begin{aligned}
i_t &= \sigma\left(x_t * W_{xi} + h_{t-1} * W_{hi} + b_i\right) \\
f_t &= \sigma\left(x_t * W_{xf} + h_{t-1} * W_{hf} + b_f\right) \\
c_t &= c_{t-1} \circ f_t + i_t \circ \tanh\left(x_t * W_{xc} + h_{t-1} * W_{ho} + b_o\right) \\
o_t &= \sigma\left(x_t * W_{xo} + h_{t-1} * W_{ho} + bo\right) \\
h_t &= o_t \circ \tanh(c_t)
\end{aligned}
$$

where $\sigma$ is the sigmoid function, $x_t$ stands for the input that is passed from the previous layer, $W_x$ and $W_h$ are 2D convolution filters, $\circ$ denotes the Hadamard product and $*$ denotes convolution operator. $h_t$ is the hidden state preserving the information to the next unit and determined by the

cell state and the output of the current unit. It can be observed that C-LSTM is able to retain spatial features as well as encoding temporal dependency (in our case, z-axis dependency). Unlike temporal sequential data (e.g., video clips), where information flows in only a forward direction, structural CT scans have two orientations that need to be considered. Hence, we use Bidirectional C-LSTM (BC-LSTM) to model both forward and backward information transfer.

*Dual attention mechanism in DABC-Module*

Due to the large channel size of the input feature map, the conventional BC-LSTM module has a large parameter size and hence a high computational cost. To solve this problem, we propose here a dual attention mechanism, namely, a spatial attention module (SABC-Module) and a channel attention module (CABC-Module).

SABC-Module is motivated by the fact that the adjacent slices have similar saliency maps. As shown in Fig. 1b, SABC-Module gets the concatenation of two different level feature maps, $F_{\text{in}} \in \mathbb{R}^{H \times W \times S \times C}$, as input. Note that those feature maps only contain in-plane xy context because they are encoded by the 2D U-Net encoder that independently processes each 2D CT slice. Since the high dimension channels of the feature map often have redundant information, we apply here a channel-dimension squeeze procedure with a depthwise convolution followed by $1 \times 1$ kernel to distill information. Consequently, we acquire a feature map $F_s = [f_1, f_2, \ldots, f_s] \in \mathbb{R}^{H \times W \times S}$ which represents refined details in intra-slice context consisting of local feature response $f_i \in \mathbb{R}^{H \times W}$ for each slice. We generate a slice-specific spatial attention map $M_s \in \mathbb{R}^{H \times W \times S}$ by feeding $F_s$ to BC-LSTM units to capture inter-slice context. Generally, similar spatial feature maps of adjacent slices contribute to higher attention scores among them. Formally, the spatial attention is formulated as:

$$M_s = \sigma\left(f_L^{3\times3}\left(f^{1\times1}\left(f_{DW}^{3\times3}\left(F_{\text{in}}\right)\right)\right)\right)$$

where $\sigma$ denotes the sigmoid activation function, $f_{DW}^{3\times3}$, $f^{1\times1}$ and $f_L^{3\times3}$ stand for the $3 \times 3$ depthwise convolution kernel, $1 \times 1$ normal convolution kernel, $3 \times 3$ convolution kernel in BC-LSTM, respectively.

Besides SABC-Module that highlights spatial relationships between adjacent slides, we also use a channel attention mechanism, CABC-Module, to highlight important channels for image segmentation. The channel-wise information is complementary to spatial saliency maps captured by SABC-Module and hence improves the understanding of overall representation.
As shown in Fig. 1b, we first use a global average pooling to generate channel-wise maps $F_c \in \mathbb{R}^{S \times C}$ from original features $F_{\text{in}} \in \mathbb{R}^{H \times W \times S \times C}$. Note that those channel-wise maps already contain multi-slice context, which represents spatial statistics in consecutive slices. Formally, CABC-Module is reformulated as:

$$M_c = \sigma\left(f_L^{5\times5}\left(P_{\text{avg}}(F_{\text{in}})\right)\right)$$

where $\sigma$, same as above, denotes the sigmoid activation function, $f_L^{5\times5}$ stand for the $5 \times 5$ convolution kernel in BC-LSTM.

It should be noted that our CABC-Module differs from the squeeze-and-excitation network proposed in previously work[22], which has only one-dimension channel map $M_{1D} \in \mathbb{R}^C$ whilst our network can utilize receptive fields in higher dimension space and hence is more efficient in capturing target response across different slices. Empirically, the attention of information from multiple slices leads to improved performance as compared to using each channel map independently. With such a channel attention mechanism, the memory cell in BC-LSTM is superior in modelling spatial-temporal interdependence between adjacent slices than fully connected layers or normal convolution layers [8] [23].

Finally, we add both spatial attention map $M_s$ and channel attention map $M_c$ to the input $F_{\text{in}}$ to generate output $F_{\text{out}} \in \mathbb{R}^{H \times W \times S \times C}$ with integrated multiple slice context.

*Uncertainty quantification of DABC-Net segmentation*

Like any neural network, predictions of DABC-Net are not always reliable, particularly when testing samples are out of training distribution, or corrupted with noise. Uncertainty estimation can measure model robustness on a particular testing sample, hence providing valuable insights of model performance to clinicians or medical experts. In DABC-Net, we approximate Bayesian inference using DropBlock [6], a form of Monte Carlo dropout. Standard dropout where random units are dropped independently does not work well for convolutional layers as spatial features are highly correlated and hence information can still be sent to the next layer via neighbouring pixels. By contrast, DropBlock drops units in a contiguous region of a feature map together and hence allows for more feature variability of convolution layers.

There are two principal types of uncertainty that can be quantified in neural networks[24]: Aleatoric uncertainty captures potential inherent noise of the input data, which means it remains comparatively constant with even more given data. The prototypical example of aleatoric uncertainty is coin flipping where the data-generating process is completely stochastic and cannot be reduced by any additional information. As opposed to this, epistemic uncertainty describes uncertainty in the model parameters, which represent the lack of knowledge of the best model. In deep learning models, epistemic uncertainty can be caused by the lack of training data in certain areas of the input domain and will decrease when training data is large and diverse. Quantification of epistemic uncertainty is particularly important for safety-critical systems such as clinical applications, where operators are sensitive to the errors in model prediction, hence epistemic uncertainty can be used to measure qualities of model outputs.

In our DABC-Net, we estimate both uncertainties with multiple inference results via Monte Carlo dropout, as in [25]:

$$\text{Uncertainty} = \underbrace{\frac{1}{T}\sum_{t=1}^{T} p_t(1-p_t)}_{\text{aleatoric}} + \underbrace{\frac{1}{T}\sum_{t=1}^{T}(p_t-\bar{p})^2}_{\text{epistemic}}$$

Where $p_t$ is the softmax outputs in the final layer at $t$th inference and $\bar{p}$ is the average of the entire $T$ inferences. In this work, we choose $T = 10$, which is a good trade-off between the reliability of uncertainty estimation and temporal consumption made by Monte Carlo sampling.

*Comparison to other state-of-the-art segmentation methods*

We compared DABC-Net with four state-of-the-art medical segmentation methods, namely DeepLabV3+, Sensor3D, nnUNet 2D & nnUNet 3D. DeepLabV3+ [7] is a popular network that is primarily used for semantic segmentation in computer vision tasks and also achieved promising performance in COVID-19 segmentation [5]. nnUNet [9] is an integrated framework for biomedical image segmentation, which contains 2D U-Net and 3D U-Net as backbone and includes extensive preprocessing and postprocessing steps, hence can be used as an out-of-the-box tool. Sensor3D [8] is a novel 3D segmentation method which also combines BC-LSTM and U-Net, yet it has a high computational cost due to the lack of an attention mechanism. Furthermore, there is no open-source Sensor3D implementation, which makes it difficult to be used in practice. In this study, we implement Sensor3D [8] on our own using Keras/Tensorflow framework. Table s1 summarizes individual features of all five methods including DABC-Net in terms of model architecture, parameter size, inference speed, code availability.

*Segmentation of lung, pneumonia lesion and consolidation region*

In this work, we train two DABC-Nets, one for lung segmentation and the other for lesion segmentation. We further multiply the output of lesion DABC-Net with the corresponding output of lung DABC-Net to remove possible false positive lesions outside the lung organ. Within the lesion region, we further outline the consolidation region thresholding at 0.5 on normalised intensity and denoising with open and close morphology operation.

*Feature extraction*

We quantify the lung volume, lesion volume and consolidation volume for left and right lung, respectively, and obtain the consolidation/lesion volume ratio by dividing by the corresponding lung volume. Additionally, we calculate the weighted volume from the inner product of the lesion and the intensity, determine the center of lesion in the z-axis (z-position of lesions is also suggested to be important for prognosis [26]). Together with non-image based features such as age and gender, we obtain 14 features per CT scan.

*Ensemble learning for prediction of disease progression*

We implemented an ensemble learning method to classify the mild vs. severe status. We use SVM, KNN, naive bayes, MLP, random forest, gradient boost, logistic regression, adaboost, and xgboost as base learners, and calculated the averaged-output from all base learners as the final output. We attempt to use the CT scans in a very early stage of patients to predict the disease progression[27]. We predict the occurrence of severe illness within a 35-day follow-up using features of the first scan, first two scans and first three scans via ensemble learning (each scan is usually obtained in a three-days interval). Additional to prediction, we also highlight important

features that mostly contribute to our prediction using random forest, gradient boost, adaboost, and xgboost. Moreover, through multiple lung/lesion segmentation with Monte Carlo dropout, we obtain multiple sets of features for each scan, which is then used to calculate the predictive entropy [25, 28].


## Acknowledgements

This work was supported by Shanghai Municipal Science and Technology Major Project (No.2018SHZDZX01), ZJLab, Shanghai Center for Brain Science and Brain-Inspired Technology, and National Key R&D Program of China (No.2019YFA0709502). C.M. acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (No. 866411).


References:


1. Wong, H. Y. F. *et al.* Frequency and Distribution of Chest Radiographic Findings in Patients Positive for COVID-19. *Radiology* **296**, E72–E78 (2020).
2. Shi, F. *et al.* Review of Artificial Intelligence Techniques in Imaging Data Acquisition, Segmentation and Diagnosis for COVID-19. *IEEE Rev. Biomed. Eng.* **PP**, (2020).
3. Yang, R. *et al.* Chest CT Severity Score: An Imaging Tool for Assessing Severe COVID-19. *Radiology: Cardiothoracic Imaging* vol. 2 e200047 (2020).
4. Wu, G. *et al.* Development of a Clinical Decision Support System for Severity Risk Prediction and Triage of COVID-19 Patients at Hospital Admission: an International Multicenter Study. doi:10.1101/2020.05.01.20053413.
5. Zhang, K. *et al.* Clinically Applicable AI System for Accurate Diagnosis, Quantitative Measurements, and Prognosis of COVID-19 Pneumonia Using Computed Tomography. *Cell* **182**, 1360 (2020).
6. Ghiasi, G., Lin, T.-Y. & Le, Q. V. DropBlock: A regularization method for convolutional networks. in *Advances in Neural Information Processing Systems 31* (eds. Bengio, S. et al.) 10727–10737 (Curran Associates, Inc., 2018).

7. Chen, L.-C., Papandreou, G., Schroff, F. & Adam, H. Rethinking Atrous Convolution for Semantic Image Segmentation. *arXiv [cs.CV]* (2017).

8. Novikov, A. A., Major, D., Wimmer, M., Lenis, D. & Buhler, K. Deep Sequential Segmentation of Organs in Volumetric Medical Scans. *IEEE Trans. Med. Imaging* **38**, 1207–1215 (2019).

9. Isensee, F., Jäger, P. F., Kohl, S. A. A., Petersen, J. & Maier-Hein, K. H. Automated Design of Deep Learning Methods for Biomedical Image Segmentation. *arXiv [cs.CV]* (2019).

10. Ma, J. *et al.* Towards Efficient COVID-19 CT Annotation: A Benchmark for Lung and Infection Segmentation. *arXiv [eess.IV]* (2020).

11. Gal, Y. & Ghahramani, Z. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. in *International Conference on Machine Learning* 1050–1059 (2016).

12. Poyiadji, N. *et al.* Acute Pulmonary Embolism and COVID-19. *Radiology* 201955 (2020).

13. Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497–506 (2020).

14. Wang, D. *et al.* Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus–Infected Pneumonia in Wuhan, China. *JAMA* **323**, 1061–1069 (2020).

15. Liang, W. *et al.* Early triage of critically ill COVID-19 patients using deep learning. *Nat. Commun.* **11**, 3543 (2020).

16. Vaid, A. *et al.* Machine Learning to Predict Mortality and Critical Events in COVID-19 Positive New York City Patients. *medRxiv* (2020).

17. Zhu, X. *et al.* Joint prediction and time estimation of COVID-19 developing severe symptoms using chest CT scan. *Med. Image Anal.* **67**, 101824 (2020).

18. Ankalaki, S., Majumdar, D. R. J. & Rakesh, H. A SEMI-SUPERVISED APPROACH TO SEMANTIC SEGMENTATION OF CHEST X-RAY IMAGES USING DEEPLABV3 FOR COVID19 DETECTION. *J. Toxicol. Environ. Health B Crit. Rev.* **7**, 3205–3212 (2020).

19. Falk, T. *et al.* U-Net: deep learning for cell counting, detection, and morphometry. *Nat. Methods* **16**, 67–70 (2019).

20. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. & Ronneberger, O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016* 424–432 (2016) doi:10.1007/978-3-319-46723-8_49.

21. Shi, X. *et al.* Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. in *Advances in Neural Information Processing Systems 28* (eds. Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M. & Garnett, R.) 802–810 (Curran Associates, Inc., 2015).

22. Hu, J., Shen, L. & Sun, G. Squeeze-and-excitation networks. in *Proceedings of the IEEE conference on computer vision and pattern recognition* 7132–7141 (2018).

23. Azad, R., Asadi-Aghbolaghi, M., Fathy, M. & Escalera, S. Bi-directional ConvLSTM U-net with Densley connected convolutions. in *Proceedings of the IEEE International Conference on Computer Vision Workshops* 0–0 (2019).

24. Kendall, A. & Gal, Y. What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? in *Advances in Neural Information Processing Systems 30* (eds. Guyon, I. et al.) 5574–5584 (Curran Associates, Inc., 2017).

25. Kwon, Y., Won, J.-H., Kim, B. J. & Paik, M. C. Uncertainty quantification using Bayesian neural networks in classification: Application to biomedical image segmentation. *Computational Statistics & Data Analysis* vol. 142 106816 (2020).

26. Yu, Q. *et al.* Multicenter cohort study demonstrates more consolidation in upper lungs on initial CT increases the risk of adverse clinical outcome in COVID-19 patients. *Theranostics* **10**, 5641–5648 (2020).

27. Liu, F. *et al.* CT quantification of pneumonia lesions in early days predicts progression to severe illness in a cohort of COVID-19 patients. *Theranostics* **10**, 5613–5622 (2020).

28. Nair, T., Precup, D., Arnold, D. L. & Arbel, T. Exploring uncertainty measures in deep networks for Multiple sclerosis lesion detection and segmentation. *Med. Image Anal.* **59**, 101557 (2020).

## Figures and caption

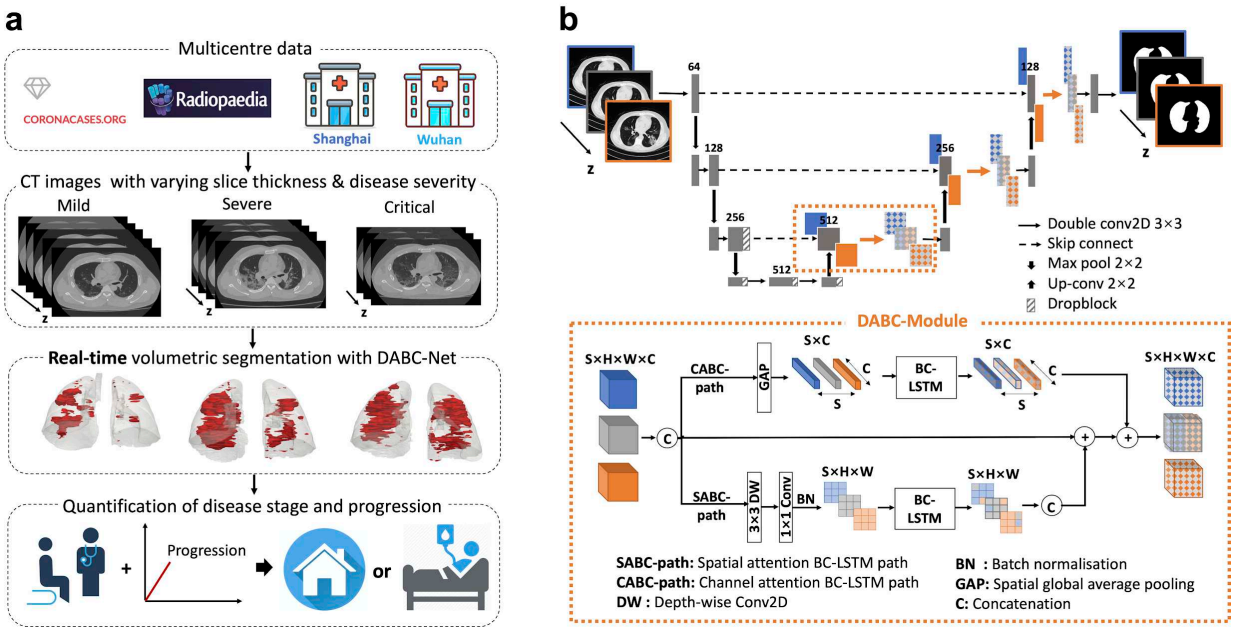



Fig. 1 | **A robust and real-time AI-empowered tool for quantitative and confident COVID-19 CT image analysis, based on a DABC-Net for automatic volumetric segmentation of lung and pneumonia lesions. a,** Workflow of our study. From top to bottom: Our multicentre data comprises CT images from four different sources. Due to variation in CT scanners and imaging protocols used in different hospitals, these images vary in terms of quality, intensity distribution, and slice thickness. Despite this variation, we achieve a robust and real-time volumetric segmentation of lung (transparent) and COVID-19 lesion (red) with DABC-Net. With an accurate lesion volume quantification, we derive a pneumonia progression score for each patient to predict whether the patient will progress into severe status. **b**, Our DABC-Net combines a 2D U-net to process intra-slice spatial information with an LSTM to leverage inter-slice context. DABC-Net uses the share-weighted 2D convolution in both encoding and decoding paths, avoiding computational expensive 3D convolution. Instead, it uses a DABC-Module (bottom) to combine inter-slice context from multiple CT slices. The DABC-module consists of two paths: i) a spatial attention bidirectional convolutional LSTM (BC-LSTM) path that uses a depthwise 2D convolution and a 1 x 1 convolution to aggregate *C* channels into a single channel resulting in only *S* x *H* x *W* BC-LSTM units, and ii) a channel attention BC-LSTM path that uses a global average pooling to eliminate spatial information, resulting in *S x C* BC-LSTM units. For a normal four-level 2D U-net, the channel number in the bottom level is 512 (*C = 512*), so our DABC-module reduces the number of BC-LSTM units by more than two orders of magnitude. Additionally, we add Dropblock modules at the end of convolution operations to allow for uncertainty assessment with Monte Carlo dropout.

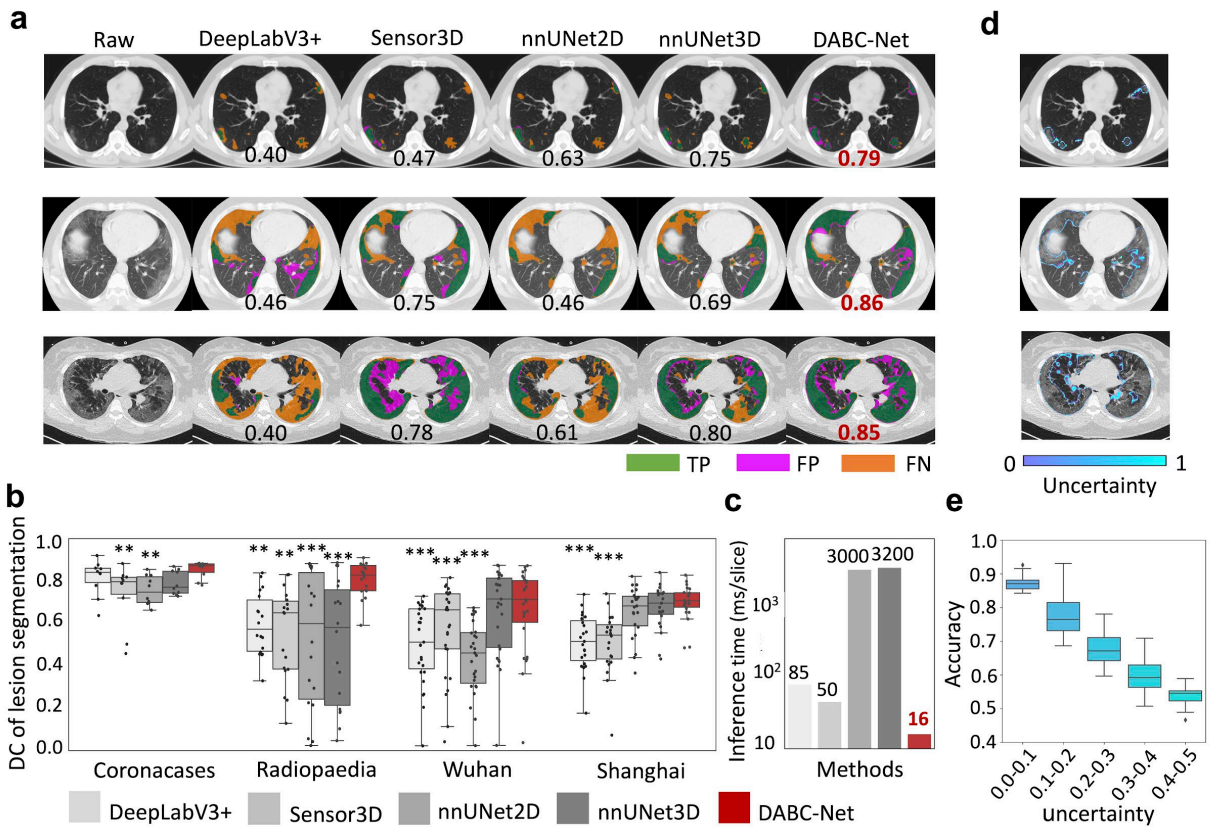



Fig. 2 | **DABC-Net outperforms state-of-the-art methods in terms of accuracy and inference time and estimates segmentation uncertainty. a,** Representative CT images of three COVID-19 patients in mild (top), severe (middle) and critical (bottom) stages and corresponding segmentation results of five methods. DABC-Net achieves highest dice coefficient (DC) in all three cases (bold) and robustly segments small lesions in the mild stage, which are often missed by other methods (see false negative (FN) regions marked by orange). **b**, DABC-Net achieves significantly higher DC than other methods in all three datasets (**p<0.01, ***p<0.001, ns: p>0.05, Friedmann test with adjusted significance level). Note the superior segmentation performance on Coronacases dataset as compared to the other two datasets. This occurs since Coronacases represents intra-center evaluation whilst the remaining three datasets, Radiopaedia, Wuhan and Shanghai hospitals are cross-center evaluation: The cases in the training set come from different centers as the test set cases. **c**, A particular highlight of DABC-Net is its fast speed. With an average inference time of 16 ms/slice, it needs less than 5 seconds to segment a conventional CT scan with about 300 slices in clinics. In comparison, nnUnet3D, due to its computational expensive interpolation preprocessing step, needs almost 20 minutes to process the same CT image. **d**, Besides, being faster and more accurate, DABC-Net also allows for estimating an uncertainty map, which highlights the region where the network is unsure of its segmentation. Indeed, those regions overlap with the regions where DABC-Net makes mistakes (see **a** right column). **e**, We divide image pixels into different intervals according to their estimated uncertainty and calculate the average pixel-wise prediction accuracy through bootstrapping. The negative correlation between uncertainty and prediction accuracy suggests that we can use uncertainty estimation to infer prediction accuracy in the testing phase when the ground-truth segmentation is not available.

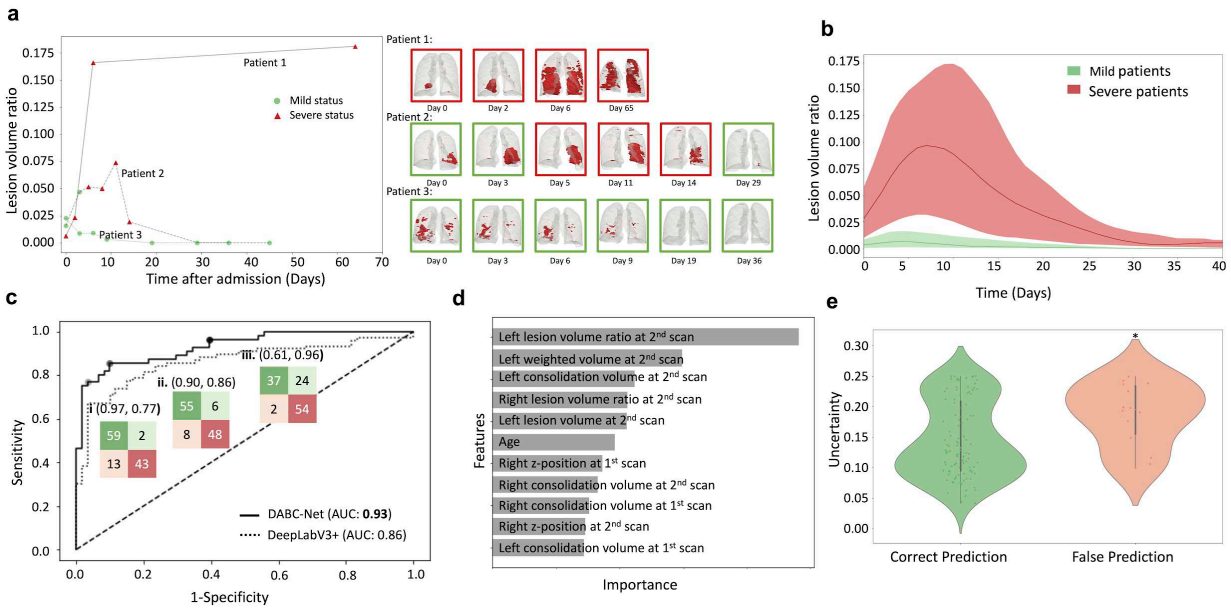



**Fig. 3 | Based on DABC-Net segmentation of longitudinal CT scans of a patient, we can quantify the development of lesion volume over time, and predict disease progression using the first two scans. a,** Temporal trajectories of lesion volume of individual patients illustrates strong variability in the COVID-triggered pneumonia progression. Severe patients tend to have larger lesion volume as compared to mild ones. Note that severe patients also include patients who only show mild symptoms when admitted to hospital (marked by green triangles) but develop severe symptoms (red triangles) during hospitalisation. **b**, Averaged trajectories of mild vs. severe patients (shaded area represents 25%-75% percentiles). Severe patients generally show a more acute disease progression than mild ones. **c**, By segmenting lesion volume from the first two CT scans with DABC-Net, we can predict whether a patient will develop severe symptoms during hospitalisation with an AUC (area under the receiver operating characteristic, ROC curve) score of 0.93. As a comparison, a less accurate segmentation method, e.g. with DeepLabV3+, would only lead to an AUC of 0.86, though we use the same features and classifier. By moving along the ROC curve, we can reach different (sensitivity, specificity) pairs (examples marked by red, green and blue dots), which adapts the availability of hospital beds and optimises the assignment. **d**, Top 10 important features selected by our classifier to distinguish severe patients from mild ones, with the most important feature being the consolidation volume of the second scan. **e**, Patients who are predicted wrongly are associated with a higher uncertainty than patients who are predicted correctly (*p<0.05, Wilcoxon ranksum test).

## Supplementary information:

| Method | Multi-slice fusion | Parameter | Speed (ms/slice) | Dice(%) for lesion | Dice(%) for Lung | Additional preprocessing |
|---|---|---|---|---|---|---|
| DeepLabV3+ | N/A | 41,252,497 | 85 | 0.7271±0.1598 | 0.9783±0.0096 | N/A |
| Sensor3D | ConvLSTM | 18,657,857 | 50 | 0.6529±0.2231 | 0.9682±0.0940 | N/A |
| nnUNet2D | N/A | **17,802,945** | 3000 | 0.7800±0.1147 | 0.9640±0.0285 | Interpolation |
| nnUNet3D | 3D convolution kernel | 30,350,177 | 3200 | 0.8163±0.0640 | 0.9164±0.1162 | Interpolation |
| DABC-Net | DABC block | 19,507,640 | **16** | **0.8401±0.0565** | **0.9821±0.0071** | N/A |

**Table S1. Comparison of different network architectures and segmentation performances**. In addition to DABC-Net, we tested four state-of-the-art segmentation methods, including DeepLabV3+, Sensor3D, nnUNet2D and nnUNet3D. Among all these methods, DeepLabV3+ and nnUNet2D are purely 2D segmentation methods; nnUNet3D is a full 3D volumetric segmentation method with 3D convolution kernels; both Sensor3D and DABC-Net are hybrid 2D-3D methods, whilst Sensor3D fuses 2D segmentation with convolutional LSTM (ConvLSTM). Unlike these methods, our DABC-Net uses a DABC module with an additional dual spatial and channel attention mechanism. Compared to other methods, DABC-Net achieves the highest dice score for both lung and lesion segmentation as well as the fastest inference speed.

| | Patients | Scans | Slices | Labeled scans | Labeled slices | | Slice thickness (mm) |
|---|---|---|---|---|---|---|---|
| | | | | | Lesion | Lung | |
| Coronacase.org* | 10* | 10 | 2581 | 10 | 2581 | 2581 | 1.0-1.5 |
| Radiopaedia.org | 19 | 19 | 1342 | 19 | 1342 | 1342 | 4.0-6.0 |
| Wuhan | 27 | 27 | 3805 | 27 | 3805 | - | 3.0 |
| Shanghai | 146 | 679 | 293349 | 43 | 2717 | 10007 | 1.0-6.0 |
| Total | 202 | 735 | 301077 | 99 | 10445 | 13930 | N/A |

**Table S2. A short summary of our multicentre datasets used to develop and evaluate our image segmentation algorithm, DABC-Net**. Four datasets, Coronacases, Radiopedia, Wuhan and Shanghai were collected independently, uploaded by individual community users, Wuhan Tongji hospital and Shanghai Public Health Clinical Center (Coronacases and Radiopedia are publicly available). Among these datasets, over 14,000 slices labelled by two radiologists were used to train and test DABC-Net and other competitive segmentation models. Moreover, the Shanghai dataset contains longitudinal scans of the same patients which reflects the progression of corona-triggered pneumonia over time. Built upon DABC-net acquired segmentation, we extract features such as lesion volume ratio (lesion volume/lung volume) and train a classification model to predict different progression patterns between mild vs severe patients.

| | Mild patients | Severe patients | Total |
|---|---|---|---|
| Male | 27 | 31 | 58 |
| Female | 34 | 25 | 59 |
| Total | 61 | 56 | 117 |

**Table S3. COVID-19 patients with longitudinal CT scans used for disease progression prediction**. All 117 patients are from Shanghai Public Health Clinical Center. Most patients receive CT examinations every three days during their hospitalisation, with exception of critically-ill patients who are in ICU and may have difficulty receiving CT examinations. Note that clinicians annotate individual scans by mild/severe/critical not only based on CT appearance, but based on patients' clinical symptoms.

| Task | Method | ROC-AUC | Sensitivity | Specificity |
|---|---|---|---|---|
| Prediction of mild status vs. severe status from single scan | DeepLabV3+ | 0.895 | 0.696 | 0.900 |
| | DABC-Net | **0.924** | **0.785** | **0.946** |
| Prediction of patient label from first scan | DeepLabV3+ | 0.759 | 0.694 | 0.703 |
| | DABC-Net | **0.840** | **0.696** | **0.820** |
| Prediction of patient label from first two scans | DeepLabV3+ | 0.881 | 0.842 | 0.833 |
| | DABC-Net | **0.931** | **0.857** | **0.918** |
| Prediction of patient label from first three scans | DeepLabV3+ | 0.902 | 0.825 | 0.850 |
| | DABC-Net | **0.967** | **0.865** | **0.951** |

**Table S4. Performance comparison of ensemble learning on DABC-Net and DeepLabV3+ segmentation.** In all three experiments, DABC-Net enhances the classification accuracy by 3 ~ 10% as compared to DeepLabV3+ segmentation, which may warrant the robustness of DABC-Net in subsequent prediction and analysis.

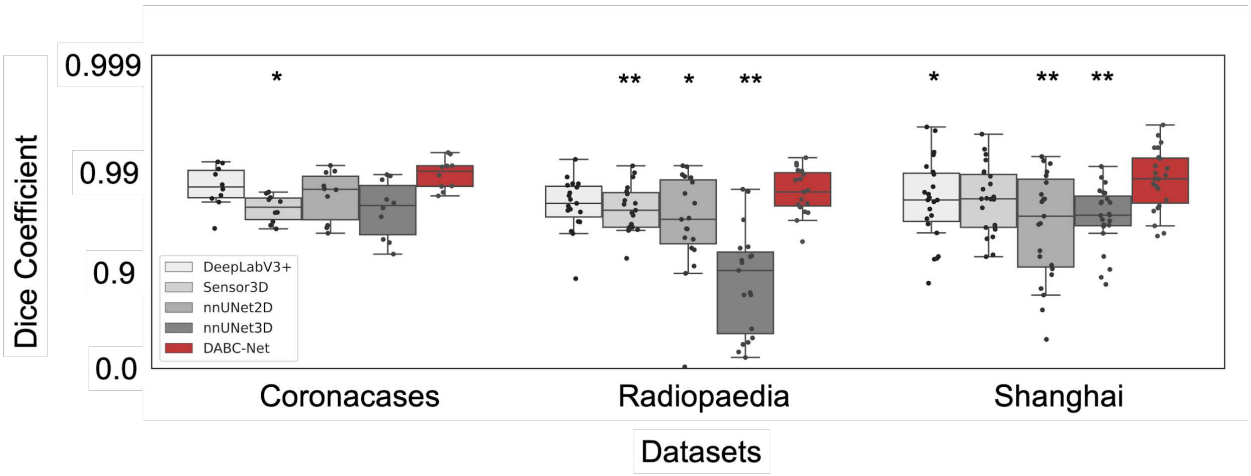



**Fig S1. The box plots present the performances of different methods for lung segmentation on multi-center datasets**. DABC-Net achieves higher dice coefficient than other methods in all three datasets (**p<0.01, ***p<0.001, Friedmann test with adjusted significance level). Note that DABC-Net achieves a robust performance of 0.95+ dice coefficient in all datasets. By contrast, nnUNet3D could completely fail in cases of Radiopaedia where the corresponding image header files are not available.

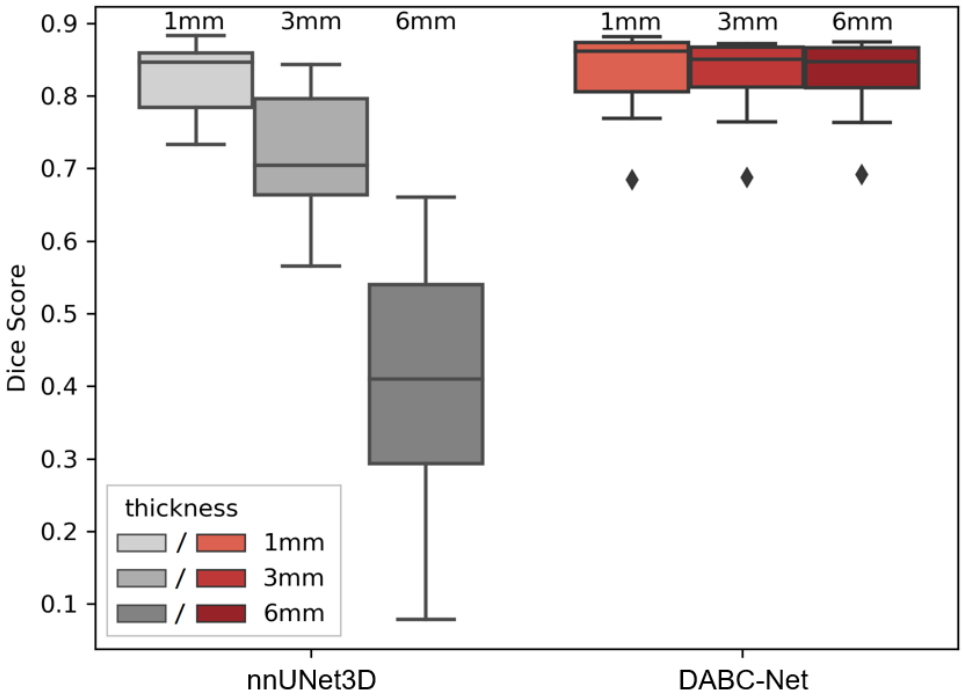



**Fig S2. Compared to nnUNet3D, DABC-Net is more robust to various CT thickness.** To evaluate the effects of CT volumes thickness, we test 1mm, 3mm and 6mm CT scans with different methods. 3mm and 6mm volumes are uniformly-spaced sampling from 1mm volumes. The performance of nnUNet3D remarkably suffers from the increasing slice thickness, while the performance of our DABC-Net remains stable, suggesting its robustness to different slice thickness.

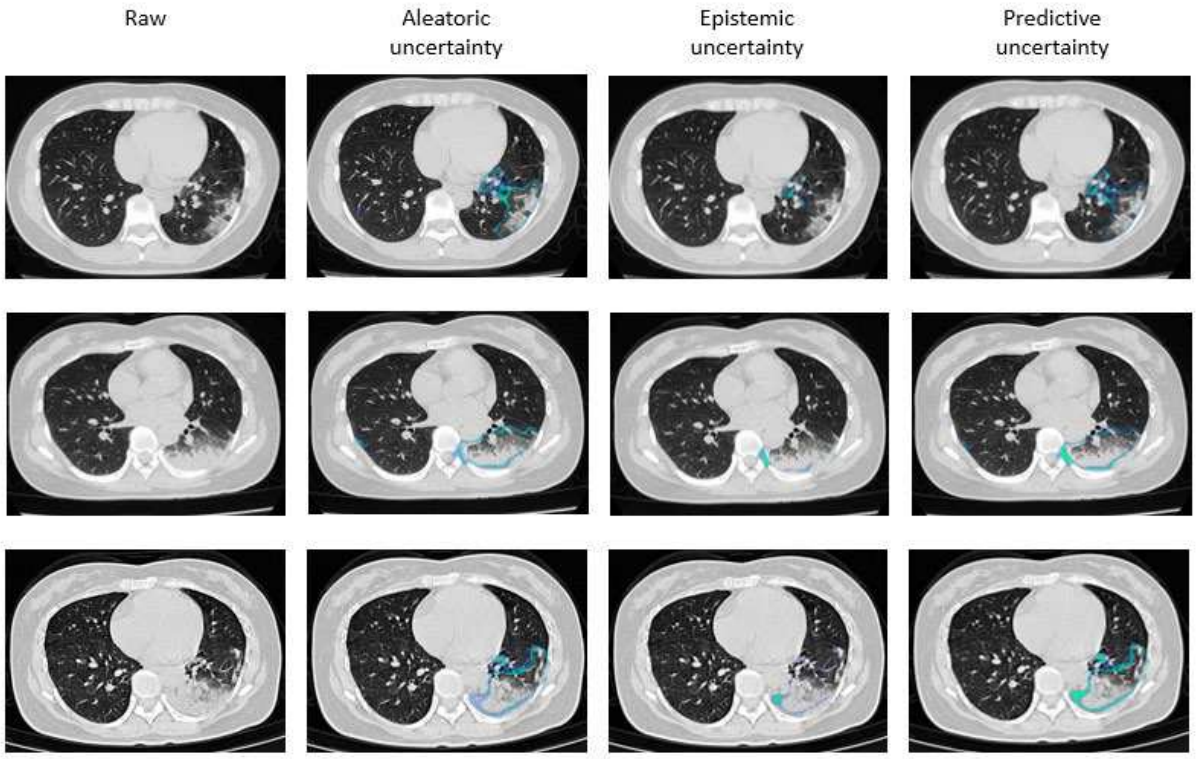



**Fig S3. Visual comparisons of uncertainty distributions generated via DABC-Net.** Aleatoric uncertainty captures potential inherent noise of the input data, while epistemic uncertainty describes uncertainty in the model parameters. Predictive uncertainty is composed of aleatoric and epistemic uncertainty.

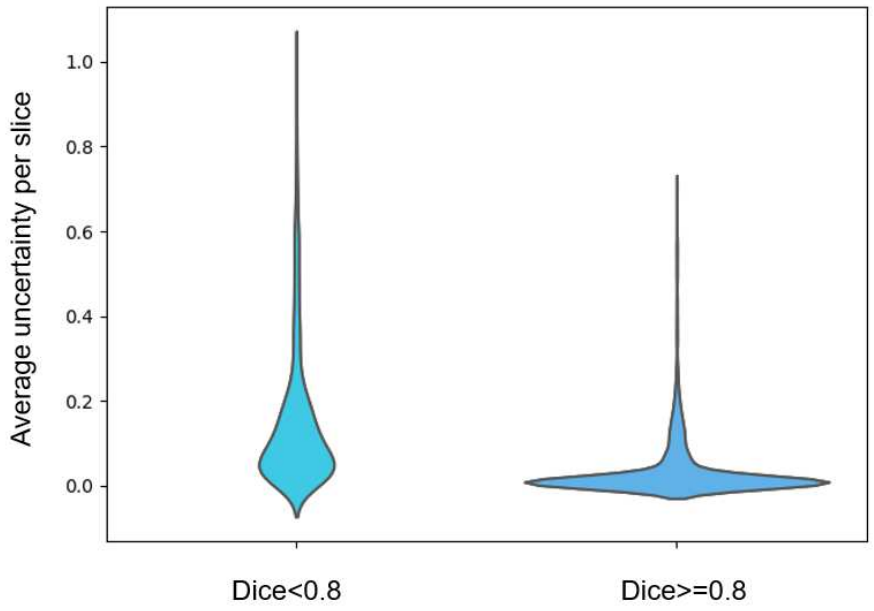



**Fig S4. The violin plot of uncertainty distribution of slices with high (dice>=0.8) vs. low (dice<0.8)**. Segmentation accuracy demonstrates that slices with higher segmentation accuracy have lower uncertainty as compared to those with less accuracy in segmentation. Uncertainty is quantified by stimulating Monte Carlo dropblock during inference in DABC-Net.

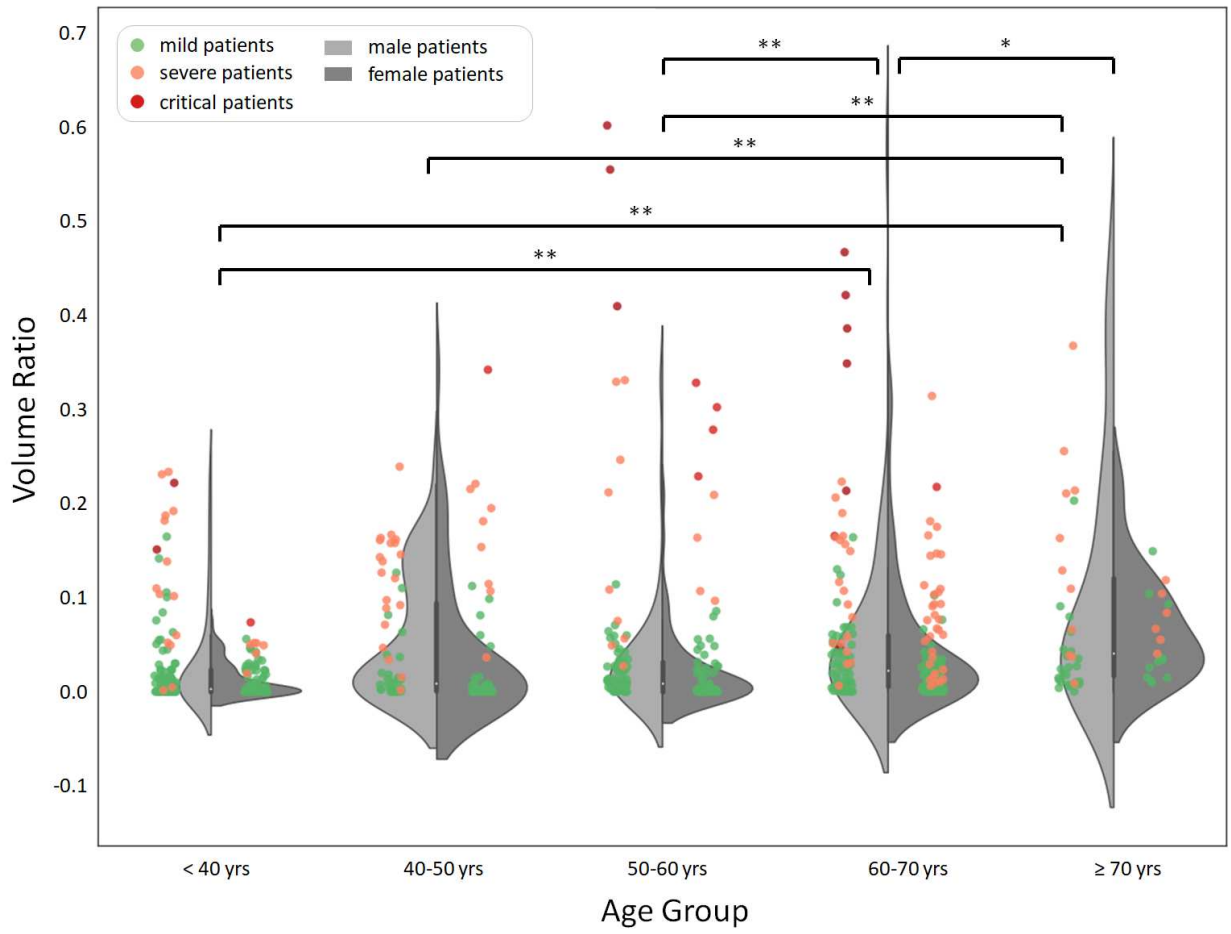



**Fig S5. Comparison of segmented lesion volume ratio between different gender and age groups.** Elderly patients tend to have a significantly larger lesion volume radio as compared to younger patients (*p<0.01 **p<0.001). In addition to age, lesion volume ratios of male patients are significantly higher than those of female patients.

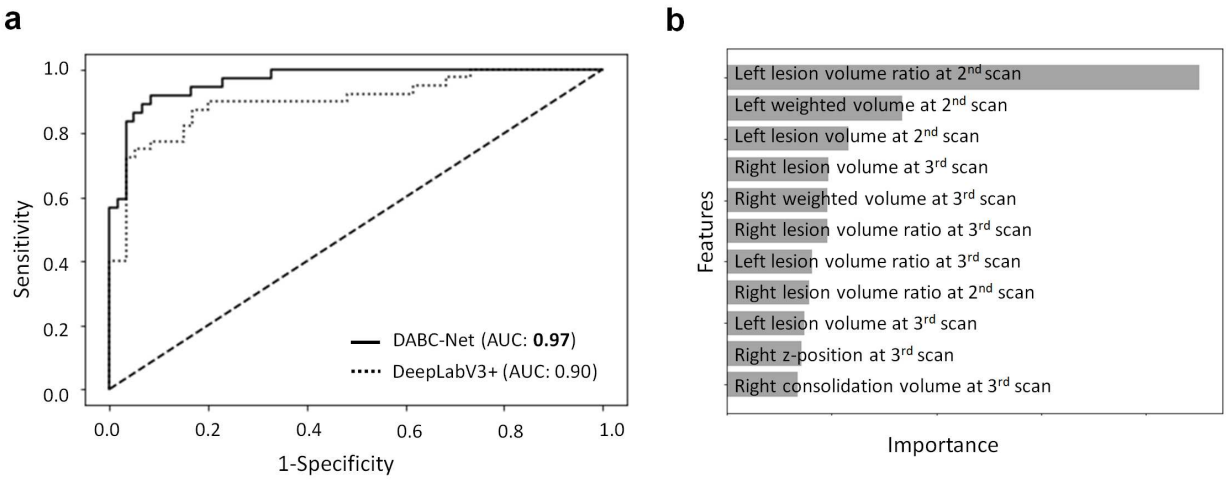



**Fig S6. We predict disease progression using the first three scans with an AUC of 0.97. a**, ROC curves of our classifiers based on segmentation of DABC-Net (AUC = 0.97) and segmentation of DeepLabV3+ (AUC = 0.90). **b**, Top 10 important features selected by our classifier based on DABC-Net segmentation to distinguish severe patients from mild ones, with the most important feature being the left lesion volume at second scan.

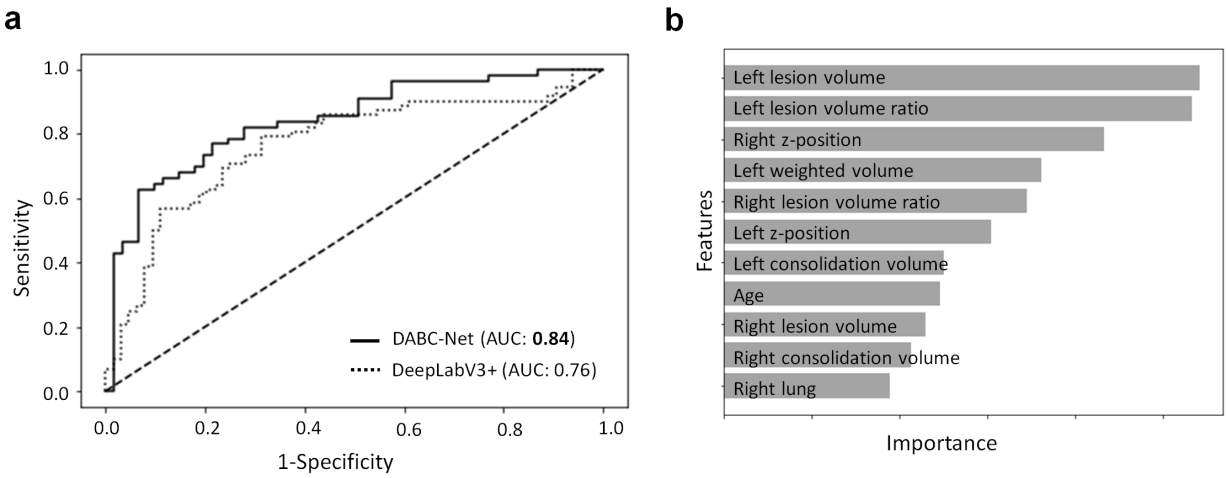



**Fig S7. We predict disease progression using the first scan with an AUC of 0.84**. **A**, ROC curves of our classifiers based on segmentation of DABC-Net (AUC = 0.84) and segmentation of DeepLabV3+ (AUC = 0.76). **b**, Top 10 important features selected by our classifier based on DABC-Net segmentation to distinguish severe patients from mild ones, with the most important feature being the left lesion volume.

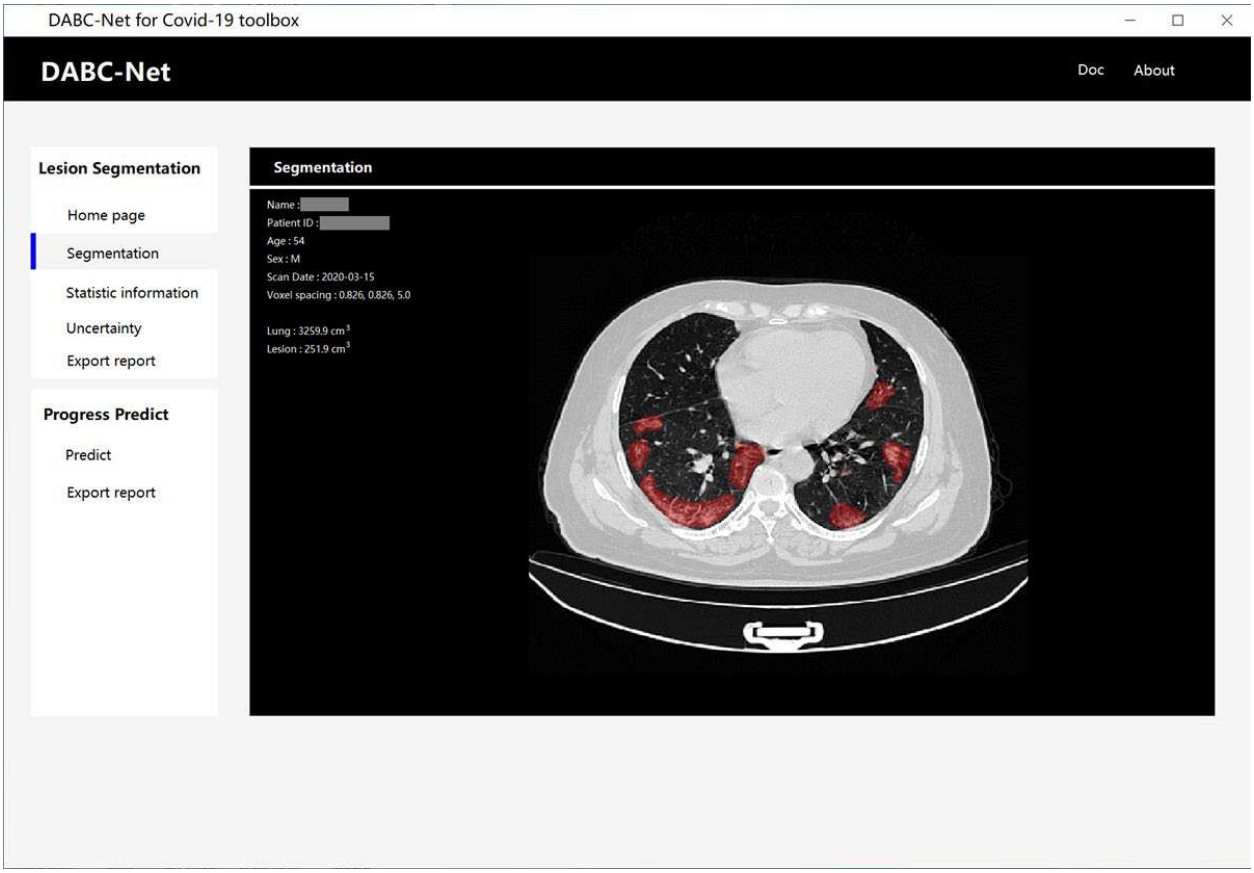



**Fig S8. The graphical user interface (GUI) of toolbox is publicly available** (download link: https://github.com/Robin970822/DABC-Net-for-COVID-19/tree/master/APP), which is also friendly for users without an AI background. Our toolbox provides lung lesion segmentation with uncertainty quantification, and prediction of disease progression.

# Figures

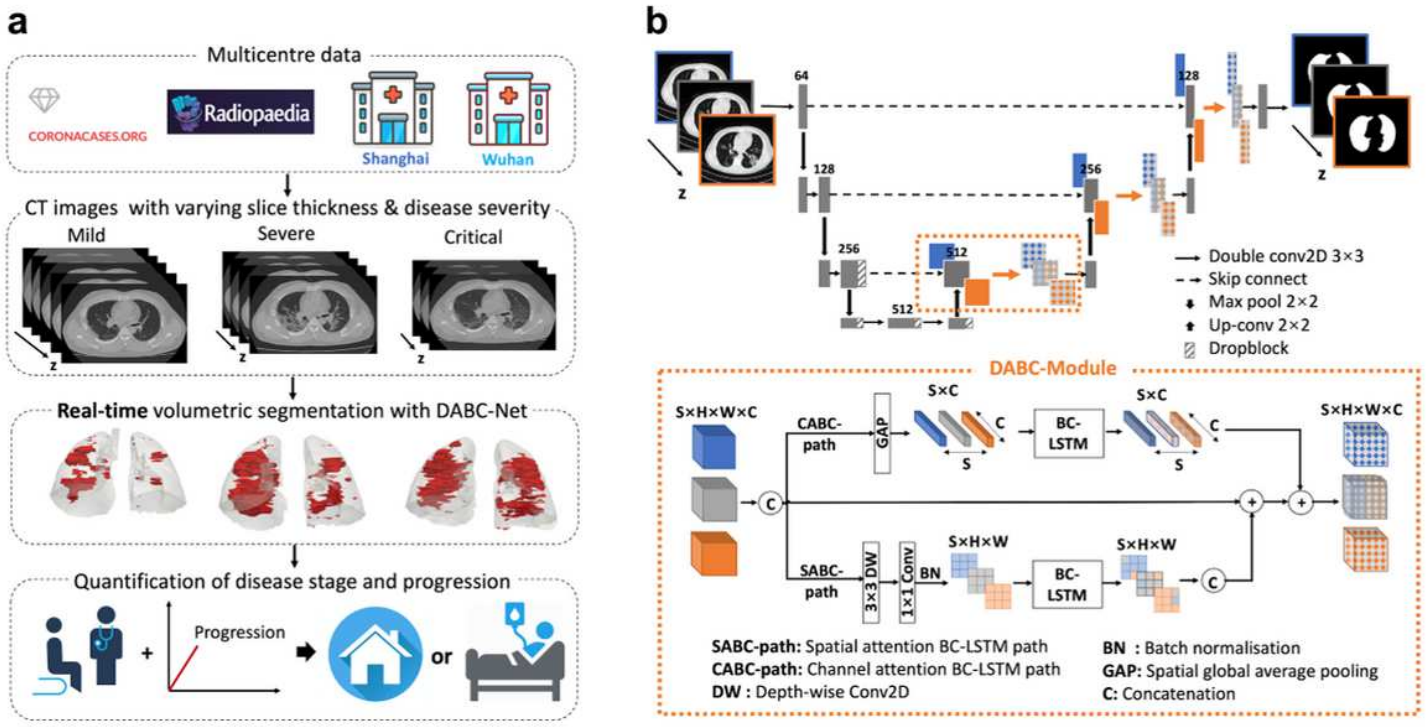
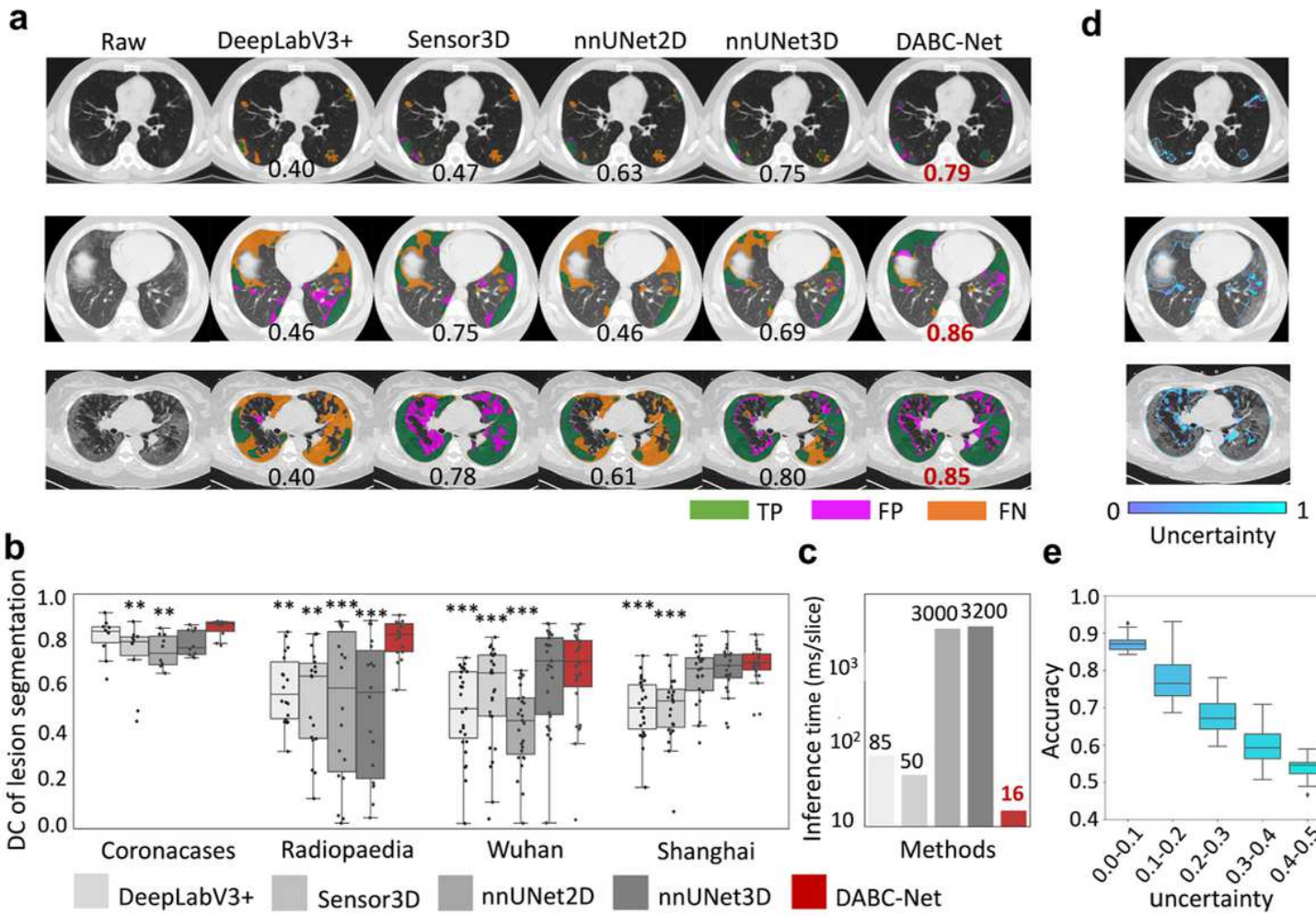



**Figure 1**

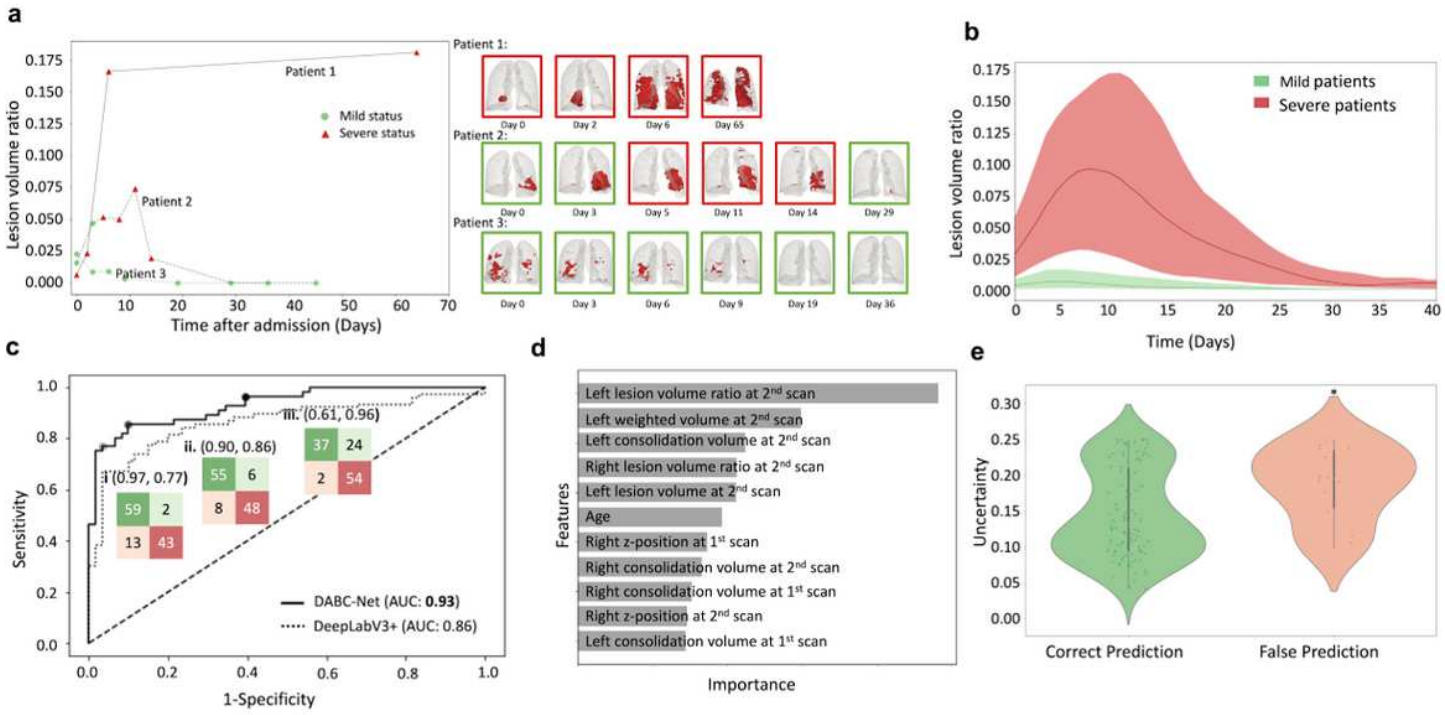
A robust and real-time AI-empowered tool for quantitative and confident COVID-19 CT image analysis, based on a DABC-Net for automatic volumetric segmentation of lung and pneumonia lesions. a, Workflow of our study. From top to bottom: Our multicentre data comprises CT images from four different sources. Due to variation in CT scanners and imaging protocols used in different hospitals, these images vary in terms of quality, intensity distribution, and slice thickness. Despite this variation, we achieve a robust and real-time volumetric segmentation of lung (transparent) and COVID-19 lesion (red) with DABC-Net. With an accurate lesion volume quantification, we derive a pneumonia progression score for each patient to predict whether the patient will progress into severe status. b, Our DABC-Net combines a 2D U-net to process intra-slice spatial information with an LSTM to leverage inter-slice context. DABC-Net uses the share-weighted 2D convolution in both encoding and decoding paths, avoiding computational expensive 3D convolution. Instead, it uses a DABC-Module (bottom) to combine inter-slice context from multiple CT slices. The DABC-module consists of two paths: i) a spatial attention bidirectional convolutional LSTM (BC-LSTM) path that uses a depthwise 2D convolution and a 1 x 1 convolution to aggregate C channels into a single channel resulting in only S x H x W BC-LSTM units, and ii) a channel attention BC-LSTM path that uses a global average pooling to eliminate spatial information, resulting in S x C BC-LSTM units. For a normal four-level 2D U-net, the channel number in the bottom level is 512 (C = 512), so our DABC-module reduces the number of BC-LSTM units by more than two orders of magnitude. Additionally, we add Dropblock modules at the end of convolution operations to allow for uncertainty assessment with Monte Carlo dropout.




## Figure 2

DABC-Net outperforms state-of-the-art methods in terms of accuracy and inference time and estimates segmentation uncertainty. a, Representative CT images of three COVID-19 patients in mild (top), severe (middle) and critical (bottom) stages and corresponding segmentation results of five methods. DABC-Net achieves highest dice coefficient (DC) in all three cases (bold) and robustly segments small lesions in the mild stage, which are often missed by other methods (see false negative (FN) regions marked by orange). b, DABC-Net achieves significantly higher DC than other methods in all three datasets (**p<0.01, ***p<0.001, ns: p>0.05, Friedmann test with adjusted significance level). Note the superior segmentation performance on Coronacases dataset as compared to the other two datasets. This occurs since Coronacases represents intra-center evaluation whilst the remaining three datasets, Radiopaedia, Wuhan and Shanghai hospitals are cross-center evaluation: The cases in the training set come from different centers as the test set cases. c, A particular highlight of DABC-Net is its fast speed. With an average inference time of 16 ms/slice, it needs less than 5 seconds to segment a conventional CT scan with about 300 slices in clinics. In comparison, nnUnet3D, due to its computational expensive interpolation preprocessing step, needs almost 20 minutes to process the same CT image. d, Besides, being faster and more accurate, DABC-Net also allows for estimating an uncertainty map, which highlights the region where the network is unsure of its segmentation. Indeed, those regions overlap with the regions where

DABC-Net makes mistakes (see a right column). e, We divide image pixels into different intervals according to their estimated uncertainty and calculate the average pixel-wise prediction accuracy through bootstrapping. The negative correlation between uncertainty and prediction accuracy suggests that we can use uncertainty estimation to infer prediction accuracy in the testing phase when the ground-truth segmentation is not available.




## Figure 3

Based on DABC-Net segmentation of longitudinal CT scans of a patient, we can quantify the development of lesion volume over time, and predict disease progression using the first two scans. a, Temporal trajectories of lesion volume of individual patients illustrates strong variability in the COVID-triggered pneumonia progression. Severe patients tend to have larger lesion volume as compared to mild ones. Note that severe patients also include patients who only show mild symptoms when admitted to hospital (marked by green triangles) but develop severe symptoms (red triangles) during hospitalisation. b, Averaged trajectories of mild vs. severe patients (shaded area represents 25%-75% percentiles). Severe patients generally show a more acute disease progression than mild ones. c, By segmenting lesion volume from the first two CT scans with DABC-Net, we can predict whether a patient will develop severe symptoms during hospitalisation with an AUC (area under the receiver operating characteristic, ROC curve) score of 0.93. As a comparison, a less accurate segmentation method, e.g. with DeepLabV3+, would only lead to an AUC of 0.86, though we use the same features and classifier. By moving along the ROC curve, we can reach different (sensitivity, specificity) pairs (examples marked by red, green and blue dots), which adapts the availability of hospital beds and optimises the assignment. d, Top 10 important features selected by our classifier to distinguish severe patients from mild ones, with the most important feature being the consolidation volume of the second scan. e, Patients who are predicted wrongly are

associated with a higher uncertainty than patients who are predicted correctly (*p<0.05, Wilcoxon ranksum test).