
Supplementary information

Open resource of clinical data from patients with pneumonia for the prediction of COVID-19 outcomes via deep learning

In the format provided by the authors and unedited

Table of contents

Supplementary Fig. 1 | The most significantly increased and decreased CFs2

Supplementary Fig. 2 | The resource page of iCTCF database3

Supplementary Fig. 3 | The accuracy for predicting morbidity or mortality outcomes exclusively using CT or CF data4

Supplementary Fig. 4 | The accuracy of Inception Net V3 and ChexNet for predicting morbidity or mortality outcomes using our CT data5

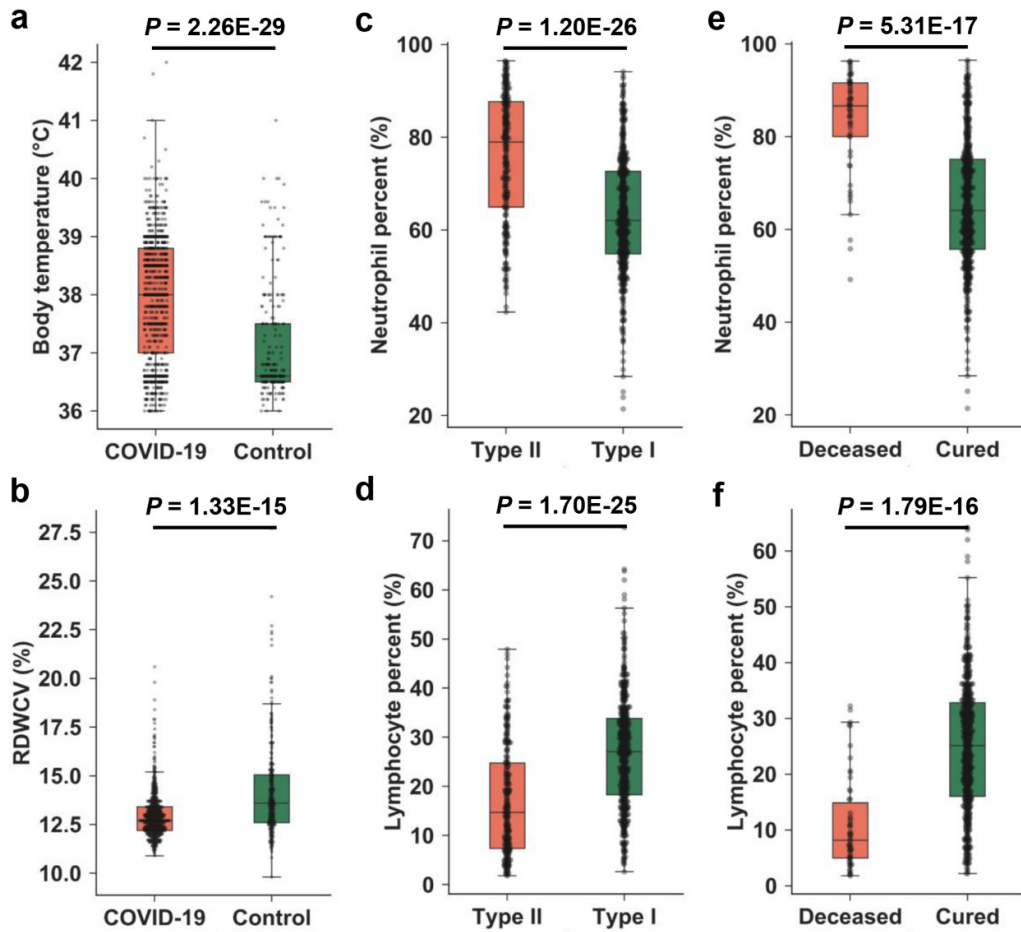
Supplementary Data 16

Supplementary Data 26

Supplementary Data 36

Supplementary Data 46

Supplementary figures



Supplementary Fig. 1 | The most significantly increased and decreased CFs. **a,b**, The most significantly increased (a) (Body temperature, $P = 2.26E-29$) and decreased (b) (RDWCV, $P = 1.33E-15$) CFs in COVID-19 patients (Type I & II, $n = 894$) against control cases ($n = 328$). **c,d**, The most significantly increased (c) (Neutrophil percent, $P = 1.20E-26$) and decreased (d) (Lymphocyte percent, $P = 1.70E-25$) CFs in Type II patients ($n = 274$) against Type I cases ($n = 620$). **e,f**, The most significantly increased (e) (Neutrophil percent, $P = 5.31E-17$) and decreased (f) (Lymphocyte percent, $P = 1.79E-16$) CFs in deceased cases ($n = 57$) against cured cases ($n = 662$). Box plots denote median and 25th to 75th percentiles (boxes) and 10th to 90th percentiles (whiskers). The data of these CFs did not follow the normal distribution (Supplementary Data 1), and p -values were calculated by the two-sided Mann-Whitney U test.

a **ICTCF Search Options**

Please search the ICTCF to find the information you need.

Search Options

Hospital: HUST-Union Hospital HUST-Liyuan Hospital

Age: Less than 40 years old 40 to 60 years old Over 60 years old

Gender: Male Female

SARS-CoV-2 nucleic acids: Positive Negative

Computed tomography (CT): Positive Negative N/A

Morbidity outcome: Critically ill Severe Regular

Mortality outcome: Mild Suspected Control

Cured Deceased Unknown

Example

b **Results**

Number of patients: 90

Hospital	Patient	Age	Gender	SARS-CoV-2 nucleic acids	CT	Morbidity	Mortality
HUST-Union Hospital	Patient 4	73	Male	Positive	Positive	Severe	Cured
HUST-Union Hospital	Patient 5	64	Female	Positive	Positive	Severe	Cured
HUST-Union Hospital	Patient 6	61	Female	Positive	Positive	Severe	Cured
HUST-Union Hospital	Patient 13	67	Female	Positive	Positive	Severe	Cured
HUST-Union Hospital	Patient 22	67	Female	Positive	Positive	Severe	Cured
HUST-Union Hospital	Patient 23	61	Female	Positive	Positive	Severe	Cured
HUST-Union Hospital	Patient 33	60	Male	Positive	Positive	Severe	Unknown
HUST-Union Hospital	Patient 44	78	Male	Positive	Positive	Severe	Deceased
HUST-Union Hospital	Patient 52	64	Female	Positive	Positive	Severe	Unknown
HUST-Union Hospital	Patient 58	66	Male	Positive	Positive	Critically ill	Deceased
HUST-Union Hospital	Patient 65	61	Male	Positive	Positive	Severe	Cured
HUST-Union Hospital	Patient 68	61	Male	Positive	Positive	Severe	Cured
HUST-Union Hospital	Patient 77	64	Female	Positive	Positive	Severe	Cured
HUST-Union Hospital	Patient 81	68	Female	Positive	Positive	Severe	Cured
HUST-Union Hospital	Patient 93	69	Male	Positive	Positive	Severe	Cured
HUST-Union Hospital	Patient 95	64	Male	Positive	Positive	Severe	Cured
HUST-Union Hospital	Patient 101	66	Female	Positive	Positive	Severe	Unknown
HUST-Union Hospital	Patient 110	65	Female	Positive	Positive	Severe	Cured
HUST-Union Hospital	Patient 134	71	Female	Positive	Positive	Severe	Cured
HUST-Union Hospital	Patient 136	65	Female	Positive	Positive	Severe	Cured

Page: 1 | First | Pre | Next | Last

c **Patient 4**

Overview

Hospital: HUST-Union Hospital
 Age: 73
 Gender: Male
 Body temperature: 38.5°C
 Underlying diseases: Aorta calcification
 SARS-CoV-2 nucleic acids: Positive
 Computed tomography (CT): Positive
 Morbidity outcome: Severe
 Mortality outcome: Cured

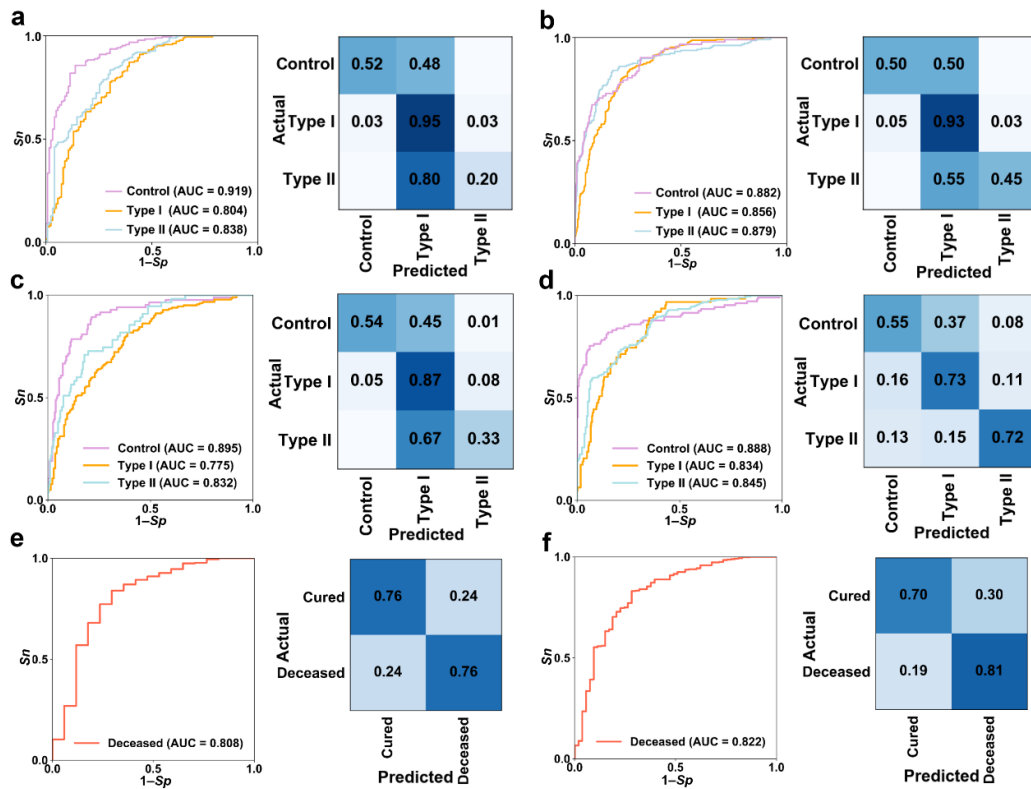
d **CT Images**

e **Clinical Features**

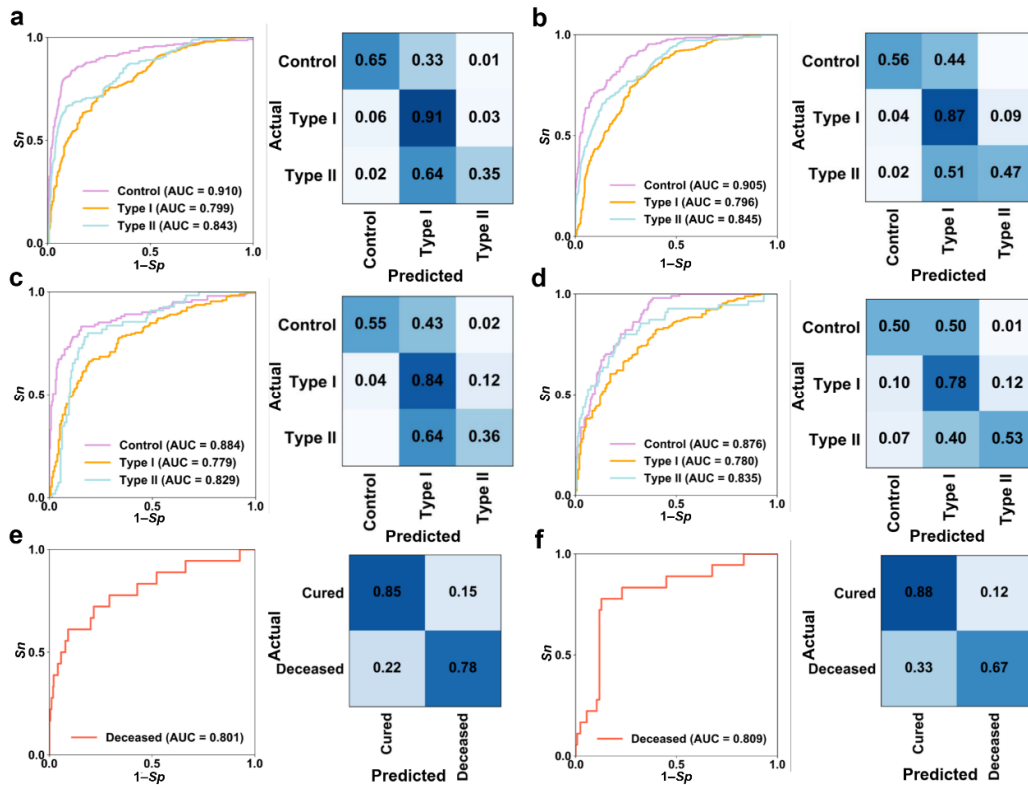
Routine Blood Test:

Clinical feature	Abbreviation	Value	Normal range
Mean corpuscular hemoglobin concentration	MCHC	337.00	316-354 g/L
Mean corpuscular hemoglobin	MCH	30.70	27-34 pg
Mean corpuscular volume	MCV	91.30	82-100 fl
Hematocrit	HCT	40.90	40-50 %
Hemoglobin	HGB	138.00	130-175 g/L
Red blood cell	RBC	4.48	4.3-5.8 T/L
Platelet distribution width	PDW	16.20	15.9-17.2 %
Plateletcrit	PCT	0.23	0.1-0.28 %
Mean platelet volume	MPV	8.30	8-12 fl
Platelet count	PLT	272.00	125-350 G/L
Basophil count	BA	0.03	<0.06 G/L
Eosinophil count	EO	0.00 ↓	0.02-0.52 G/L
Monocyte count	MO	0.30	0.1-0.6 G/L
Lymphocyte count	LY	0.73 ↓	1.1-3.2 G/L
Neutrophil count	NE	4.19	1.8-6.3 G/L
Basophil percent	Bap	0.60	0-1 %
Eosinophil percent	Eop	0.00 ↓	0.4-5.0 %
Monocyte percent	Mop	5.80	3-10 %
Lymphocyte percent	Lyp	13.90 ↓	20-50 %
Neutrophil percent	Nep	79.70 ↑	40-75 %
White blood cell	WBC	5.25	3.5-9.5 G/L
Platelet larger cell ratio	PLCR	16.30	13.0-43.0 %
Standard deviation of red cell volume distribution width	RDWSD	40.00	39.0-46.0 fl

Supplementary Fig. 2 | The resource page of iCTCF database. a, Multiple selection options were provided for searching the resource in a customizable manner. **b**, After clicking “Submit”, the results returned in a tabular list. **c**, A general summary of a given patient in the overview section. **d**, One of five representative chest CT slices in JPEG format was displayed for each patient. The other slices could be seen by clicking horizontal scrolling buttons. **e**, The numerical CFs of a given patient were shown in a tabular list.



Supplementary Fig. 3 | The accuracy for predicting morbidity or mortality outcomes exclusively using CT or CF data. Both receiver operating characteristic (ROC) curves and confusion matrices were illustrated. **a**, CT-based prediction of morbidity outcomes using the Cohort 1 (207 controls, 384 Type I patients, and 149 Type II patients having CT data). **b**, CF-based prediction of morbidity outcomes using the Cohort 1 (222 controls, 438 Type I patients, and 211 Type II patients having CF data). **c**, CT-based prediction of morbidity outcomes using the Cohort 2 (106 controls, 180 Type I patients, and 56 Type II patients having CT data). **d**, CF-based prediction of morbidity outcomes using the Cohort 2 (106 controls, 182 Type I patients, and 63 Type II patients having CF data). **e**, CT-based prediction of mortality outcomes using the merged Cohort (593 cured and 19 deceased cases having CT data). **f** CF-based prediction of mortality outcomes using the merged Cohort (662 cured and 57 deceased cases having CF data).



Supplementary Fig. 4 | The accuracy of Inception Net V3³⁶ and ChexNet³⁷ for predicting morbidity or mortality outcomes using our CT data. Both ROC curves and confusion matrices were illustrated. **a,b**, The prediction of morbidity outcomes using the Cohort 1 based on Inception Net V3 (a) and ChexNet (b) (207 controls, 384 Type I patients, and 149 Type II patients having CT data). **c,d**, The prediction of morbidity outcomes using the Cohort 2 based on Inception Net V3 (c) and ChexNet (d) (106 controls, 180 Type I patients, and 56 Type II patients having CT data). **e,f**, The prediction of mortality outcomes using the merged Cohort based on Inception Net V3 (e) and ChexNet (f) (593 cured and 19 deceased cases having CT data).

Description of the Supplementary Dataset

Supplementary Data 1

The details on 9 classes and 130 types of CFs, including CF names, abbreviations, and normal range values in HUST-UH and HUST-LH. *a.* Before statistical analysis, the normality of data distribution was evaluated by the Shapiro-Wilk test for each CF. *b.* p -value > 0.05 denoted data following normal distribution, and there were 11 CFs following the normal distribution; *c.* N/A, not available; *d.* F/M, female or male.

Supplementary Data 2

The information regarding underlying diseases and morbidity outcomes of patients from HUST-UH or HUST-LH for the two cohorts.

Supplementary Data 3

The statistical comparisons of numerical CFs in different types of patients. In the sheet "Positive vs. Control", statistically higher and lower (Marked in grey) CFs were shown in cases with COVID-19 pneumonia against control patients. In the sheet "Type II vs. Type I", statistically higher and lower (Marked in grey) CFs were shown in Type II patients against Type I patients. In the sheet "Deceased vs. Cured", statistically higher and lower (Marked in grey) CFs were shown in deceased cases against cured patients. For the above three sheets, fold changes were present, while the two-sided unpaired t-test and Mann-Whitney U test were conducted for CF data following normal distribution or not, respectively (p -value $< 10^{-4}$). Adjusted p -values were calculated using the Benjamini-Hochberg method ($< 10^{-3}$ in this study). In the sheet "Underlying diseases", the proportions of patients with and without underlying diseases (Udis) were compared (Two-sided chi-squared test) between COVID-19 patients (Type I & II) and controls, between Type II and Type I patients, and between deceased and cured cases.

Supplementary Data 4

All pre-configured parameters in CNN and DNN frameworks. *a.* Input, size of an input image (height * width * channel); *b.* Convolutional, number and size of convolutional kernels (number * height * width); *c.* Pooling, size of a pooling layer (height * width); *d.* Dense, number of nodes in the dense layer; *e.* Dropout, the dropout ratio; *f.* Output, number of nodes in the output layer; *g.* Decay, learning rate decay factor; *h.* Epochs, number of iterations.