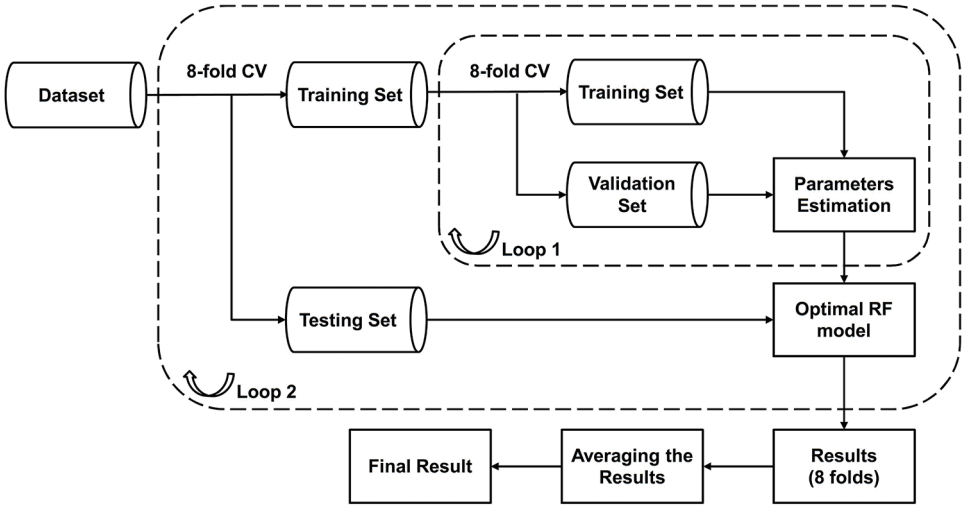

Figure S1 Scheme for establishing the RF-based mortality prediction model. Two nested loops were used. In Loop 2, the whole training and testing was based on the 8-fold cross-validation scheme. In Loop 1, 8-fold cross-validation was used to determine the optimal parameters in the RF model. RF, random forest.

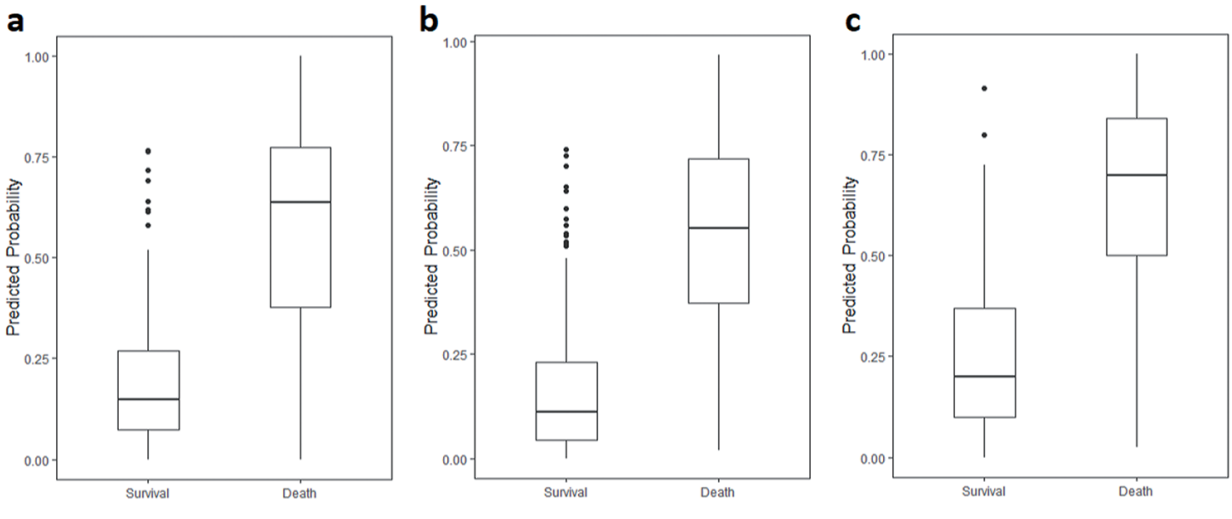

Figure S2 Discrimination plots of the RF-based mortality prediction model with Dataset 1 (the patients with complete records of the existing scoring systems from MIMIC; namely, subset of Dataset 2). (A) In-hospital mortality prediction. (B) 30-day mortality prediction. (C) 1-year mortality prediction. RF, random forest; MIMIC-III, Medical Information Mart for Intensive Care database.

http://dx.doi.org/10.21037/atm-20-6624

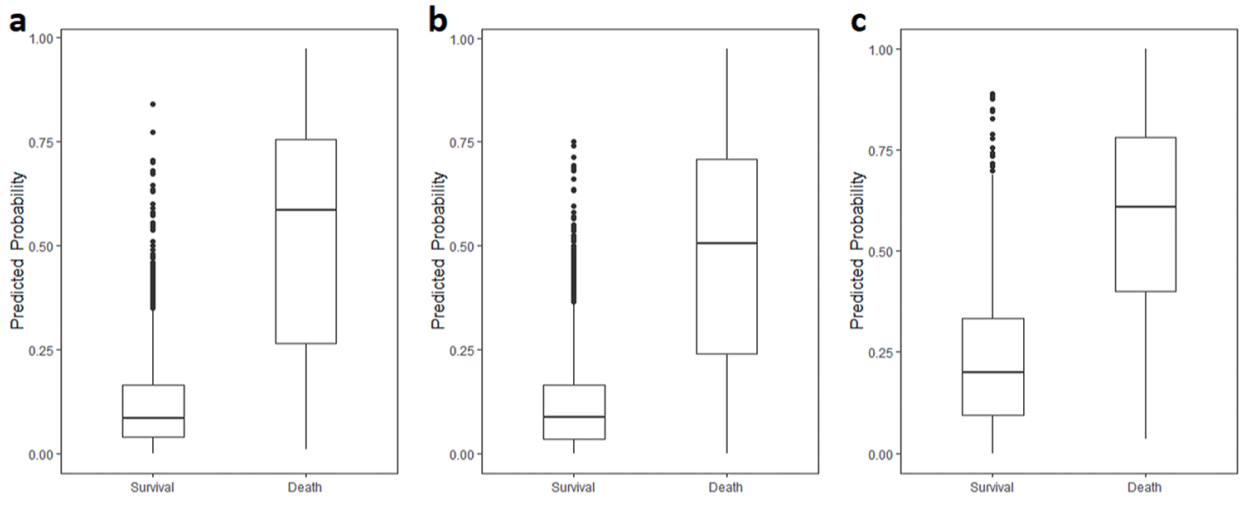

Figure S3 Discrimination plots of the RF-based mortality prediction model with Dataset 2 (2,235 patients from MIMIC-III). (A) In-hospital mortality prediction. (B) 30-day mortality prediction. (C) 1-year mortality prediction. RF, random forest; MIMIC-III, Medical Information Mart for Intensive Care database.

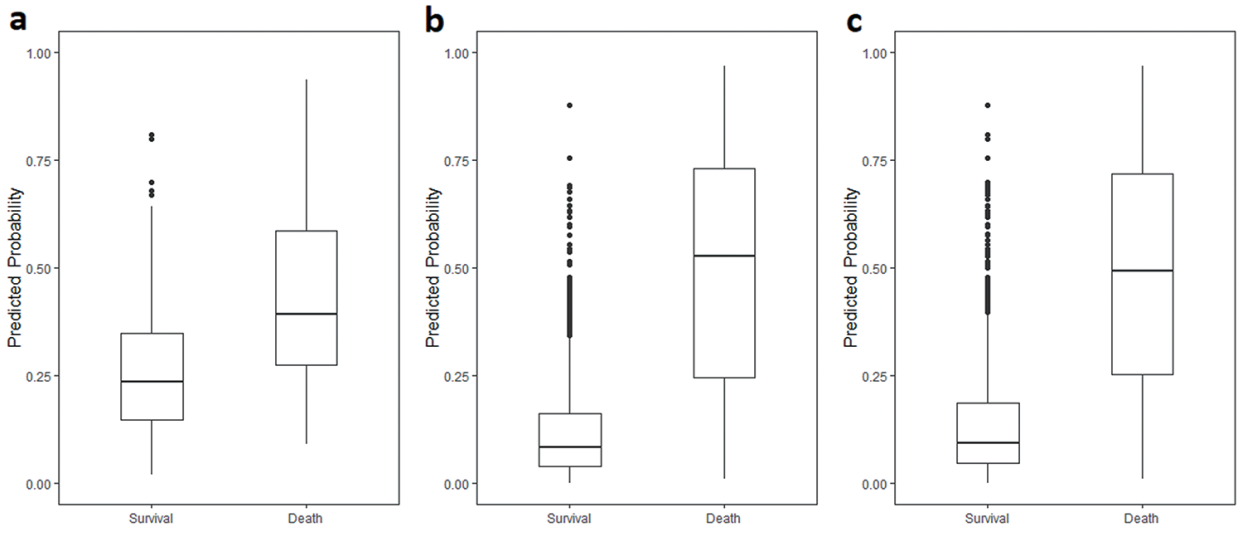

Figure S4 Discrimination plots of the RF-based mortality prediction model with multi-source data including Dataset 2 (2,235 patients from MIMIC-III) and Dataset 3 (331 patients from eICU-CRD). (A) Box plot for Dataset 3 only. (B) Box plot for Dataset 2 alone. (C) Box plot for the combination of Dataset 2 and Dataset 3. RF, random forest; MIMIC-III, Medical Information Mart for Intensive Care database; eICU-CRD, the Telehealth Intensive Care Unit Collaborative Research Database.

http://dx.doi.org/10.21037/atm-20-6624

Table S1 Calibration comparisons between the existing scoring systems and the proposed RF model in predicting ARDS mortality with Dataset 1 (subset of Dataset 2)

Methods	In-hospital mortality (Hosmer-Lemeshow goodness-of-fit Chi²)	30-day mortality (Hosmer-Lemeshow goodness-of-fit Chi²)	1-year mortality (Hosmer-Lemeshow goodness-of-fit Chi²)
OI	9.766 (P=0.282)	14.501 (P=0.070)	9.466 (P=0.305)
OSI	11.637 (P=0.168)	11.097 (P=0.196)	12.348 (P=0.136)
APPS	4.489 (P=0.213)	3.347 (P=0.341)	3.174 (P=0.366)
SOFA	10.263 (P=0.174)	5.395 (P=0.612)	6.050 (P=0.534)
SAPS-II	4.862 (P=0.677)	4.344 (P=0.739)	9.754 (P=0.203)
RF model	10.419 (P=0.237)	3.295 (P=0.915)	4.366 (P=0.823)

Null hypothesis in Hosmer-Lemeshow test: the true probabilities are those specified by the model. A two-tailed P value less than 0.05 was considered statistically significant. ARDS, acute respiratory distress syndrome; RF, random forest; OI, oxygenation index; OSI, oxygen saturation index; SOFA, Sepsis-related Organ Failure Assessment; SAPS-II, Simplified Acute Physiology Score II.

Table S2 Calibration comparisons between the SOFA, SAPS-II scoring system and proposed RF model in predicting ARDS mortality with Dataset 2 (2,235 patients from MIMIC-III)

Methods	In-hospital mortality (Hosmer-Lemeshow goodness-of-fit Chi²)	30-day mortality (Hosmer-Lemeshow goodness-of-fit Chi²)	1-year mortality (Hosmer-Lemeshow goodness-of-fit Chi²)
SOFA	5.917 (P=0.657)	6.588 (P=0.582)	5.358 (P=0.719)
SAPS-II	7.389 (P=0.495)	6.686 (P=0.571)	10.424 (P=0.237)
RF model	12.488 (P=0.131)	6.855 (P=0.552)	9.535 (P=0.299)

Null hypothesis in Hosmer-Lemeshow test: the true probabilities are those specified by the model. A two-tailed P value less than 0.05 was considered statistically significant. ARDS, acute respiratory distress syndrome; RF, random forest; SOFA, Sepsis-related Organ Failure Assessment; SAPS-II, Simplified Acute Physiology Score II; MIMIC-III, Medical Information Mart for Intensive Care.

Table S3 Calibration and discrimination comparisons between the SAPS-II, APACHE scoring systems and proposed RF model in predicting ARDS in-hospital mortality with multi-source data (the combination of Dataset 2 and Dataset 3) but performances were estimated for Dataset 2 and Dataset 3 respectively

Methods	Calibration, Hosmer-Lemeshow test Chi² (P value)	Discrimination, AUROC (95% CI)
eICU-CRD		
SAPS-II	6.341 (P=0.609)	0.511 (0.433–0.588)
APACHE	5.518 (P=0.597)	0.528 (0.452–0.605)
RF model	8.694 (P=0.369)	0.736 (0.664–0.807)
MIMIC-III		
RF model	18.076 (P=0.021)	0.905 (0.887–0.922)

Null hypothesis in Hosmer-Lemeshow test, the true probabilities are those specified by the model. A two-tailed P value less than 0.05 was considered statistically significant. SAPS-II, Simplified Acute Physiology Score II; APACHE, Acute Physiology and Chronic Health Evaluation; RF, random forest; ARDS, acute respiratory distress syndrome; AUROC, area under the receiver operating characteristic curve; CI, confidence interval; eICU-CRD, Telehealth Intensive Care Unit Collaborative Research Database; MIMIC-III, Medical Information Mart for Intensive Care.

http://dx.doi.org/10.21037/atm-20-6624

Appendix Section 1: The frequency of the present/missing data for each feature used in random forest (RF) method for acute respiratory distress syndrome (ARDS) mortality prediction

Variables	Dataset 1 (N=308)			Dataset 2 (N=2,235)			Dataset 3 (N=331)		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
Age	308	58.29	18.68	2,235	63.97	17.43	331	63.86	16.35
platelet_mean	308	274.87	136.82	2,223	268.30	136.52	331	211.13	107.44
lactate_mean	295	2.85	2.37	2,036	2.41	1.86	261	2.43	2.04
ph_mean	308	7.38	0.05	2,161	7.39	0.06	331	7.33	0.09
pao2_mean	308	115.82	24.32	2,145	124.53	49.17	331	123.96	63.83
wbc_mean	308	13.88	4.86	2,223	12.78	8.11	331	12.69	7.69
fio2_mean	308	58.84	12.80	1,878	58.96	17.70	331	57.00	23.51
creatinine_mean	308	1.24	1.04	2,223	1.30	1.19	331	1.54	1.25
temperature_max	306	38.54	0.96	1,950	38.03	1.04	330	37.71	0.79
heartrate_mean	308	91.67	12.43	1,863	90.64	13.04	331	91.55	17.71
albumin_mean	284	2.55	0.52	1,760	2.81	0.51	292	2.67	0.52
BMI	205	28.65	8.30	1,766	27.78	7.07	326	31.48	9.67

wbc_mean, mean white blood cell count; BMI, body mass index.

http://dx.doi.org/10.21037/atm-20-6624

Appendix Section 2: Hyperparameters and tuning strategies for RF model in the study

The RF model development was performed in the Python environment using scikit-learn library (version 0.22; https://www.scikit-learn.org/). RF algorithm involved several hyperparameters controlling the structure of each individual tree and the forest, including n_estimators (the number of trees), min_sample_leaf (the minimum number of samples in a terminal node), max_depth (the maximum depth of the trees) and criterion (categorical hyperparameter, the splitting rule). A grid-search strategy was performed to determine the optimal hyperparameters, in which all possible combinations of given candidate hyperparameter values were evaluated.

Hyperparameters	Grid-search settings
n_estimators	Lower bound, 11, upper bound, 301
min_sample_leaf	Lower bound, 1, upper bound, 10
max_depth	"None", the nodes are expended until all leaves are pure.
criterion	"Gini" criteria, the split that minimizes the Gini impurity

http://dx.doi.org/10.21037/atm-20-6624

Appendix Section 3

A total of 90 variables were collected in this study, including demographic data, ventilator settings, laboratory hemodynamic variables, physiological information, resuscitation status and other clinical data that may be relevant to the mortality of ARDS patients.

Type	Variables
Demographic data (4) [1]	Age*†
	Gender
	Ethnicity
	Admission type
Ventilator settings (6) [3]	Positive end expiratory pressure (minimum, maximum)
	Plateau pressure (minimum*, maximum*)
	Mean airway pressure (minimum*, maximum)
Laboratory hemodynamic variables (27) [20]	White blood cell (minimum, mean*†, maximum)
	Bilirubin (minimum*, mean*, maximum*)
	Creatinine (minimum*, mean*†, maximum*)
	Platelet (minimum*, mean*†, maximum*)
	Albumin (minimum, mean*†, maximum)
	Ph value (minimum, mean*†, maximum)
	Hemoglobin (mean*)
	Hematocrit (mean*)
	Lactate (mean*†)
	Partial pressure of oxygen, PaO_2 (mean*†)
	Fraction of inspiration O_2, FiO_2 (mean*†)
	Oxygen saturation, SpO_2 (mean*)
	Partial pressure of CO_2, PCO_2 (mean*)
	Plasma*
	Cryoprecipitate
Physiological information (21) [17]	Heartrate (minimum*, mean*†, maximum*)
	Respiratory rate (minimum, mean*, maximum*)
	Mean arterial pressure (minimum*, mean*, maximum)
	Temperature (minimum*, mean*, maximum*†)
	Tidal volume (minimum, mean*, maximum)
	Weight*
	Mean tidal volume/Weight*
	Body mass index*†
	Drug or substance input*
	Urine output*
	Output-Input*

(*continued*)

http://dx.doi.org/10.21037/atm-20-6624

(*continued*)

Type	Variables
Other clinical data (28) [1]	Transfusion*
	Pulmonary circulation
	Peripheral vascular
	Hypertension
	Paralysis
	Other neurological disorders
	Chronic pulmonary
	Uncomplicated diabetes
	Complicated diabetes
	Hypothyroidism
	Renal failure
	Liver disease
	Peptic ulcer
	AIDS
	Lymphoma
	Metastatic cancer
	Solid tumor
	Rheumatoid arthritis
	Coagulopathy
	Obesity
	Weight loss
	Fluid electrolyte
	Blood loss anemia
	Deficiency anemia
	Alcohol abuse
	Drug abuse
	Psychoses
	Depression
Resuscitation status (4) [3]	Full code (first*)
	Do not resuscitate (first*)
	Do not intubate (first)
	Comfort measure only (first*)

*, top 45 (50%) features with high Gini importance in RF model for ARDS mortality prediction based on Dataset 1. †, only 12 features that could be obtained in both MIMIC-III and eICU-CRD database were enrolled in final analysis based on Gini importance. The rest were excluded due to lack of record, equal records of the same variable for all patients or all times in eICU-CRD database. RF, random forest; ARDS, acute respiratory distress syndrome; MIMIC-III, Medical Information Mart for Intensive Care; eICU-CRD, Telehealth Intensive Care Unit Collaborative Research Database.

http://dx.doi.org/10.21037/atm-20-6624