

Towards an Artificial Intelligence Framework for Data-Driven Prediction of Coronavirus Clinical Severity

Xiangao Jiang¹, Megan Coffee^{2,3,*}, Anasse Bari^{4,*}, Junzhang Wang⁴, Xinyue Jiang⁵, Jianping Huang¹, Jichan Shi¹, Jianyi Dai¹, Jing Cai¹, Tianxiao Zhang⁶, Zhengxing Wu¹, Guiqing He¹ and Yitong Huang⁷

Abstract: The virus SARS-CoV2, which causes coronavirus disease (COVID-19) has become a pandemic and has spread to every inhabited continent. Given the increasing caseload, there is an urgent need to augment clinical skills in order to identify from among the many mild cases the few that will progress to critical illness. We present a first step towards building an artificial intelligence (AI) framework, with predictive analytics (PA) capabilities applied to real patient data, to provide rapid clinical decision-making support. COVID-19 has presented a pressing need as a) clinicians are still developing clinical acumen to this novel disease and b) resource limitations in a surging pandemic require difficult resource allocation decisions. The objectives of this research are: (1) to algorithmically identify the combinations of clinical characteristics of COVID-19 that predict outcomes, and (2) to develop a tool with AI capabilities that will predict patients at risk for more severe illness on initial presentation. The predictive models learn from historical data to help predict who will develop acute respiratory distress syndrome (ARDS), a severe outcome in COVID-19. Our results, based on data from two hospitals in Wenzhou, Zhejiang, China, identified features on initial presentation with COVID-19 that were most predictive of later development of ARDS. A mildly elevated alanine aminotransferase (ALT) (a liver enzyme), the presence of myalgias (body aches), and an elevated hemoglobin (red blood cells), in this order, are the clinical features, on presentation, that are the most predictive. The predictive models that learned from historical data of patients from these two hospitals achieved 70% to 80% accuracy in predicting severe cases.

¹ Department of Infectious Diseases, Wenzhou Central Hospital, Wenzhou, 325000, China.

² Division of Infectious Diseases and Immunology, Department of Medicine, New York University, New York, USA.

³ Department of Population and Family Health, Mailman School of Public Health, Columbia University, New York, USA.

⁴ Courant Institute of Mathematical Sciences, Computer Science Department, New York University, New York, USA.

⁵ Columbia College, Columbia University, New York, USA.

⁶ Departments of Infectious Diseases, Cangnan People's Hospital, Wenzhou, 325800, China.

⁷ Department of Gynaecology, Wenzhou Central Hospital, Wenzhou, 325000, China.

* Corresponding Authors: Anasse Bari, Email: abari@nyu.edu;

Megan Coffee, Email: Megan.Coffee@nyulangone.org.

Received: 02 March 2020; Accepted: 24 March 2020.

Keywords: SARS-CoV2, COVID-19, coronavirus, infectious diseases, artificial intelligence, predictive analytics.

1 Introduction

Since December 2019, the virus SARS-CoV2, causing the Coronavirus disease (COVID-19), has spread from Wuhan, China to every inhabited continent [World Health Organization (2020)]. As the COVID-19 outbreak is now a pandemic, it will be important to have tools to rapidly identify those at most risk of morbidity and mortality. Infections often result in nosocomial spread, affecting health workers and the general provision of healthcare. Caseloads can overwhelm hospitals, with a high need for oxygen, prolonged ventilation and even extracorporeal membrane oxygenation (ECMO), particularly for patients with acute respiratory distress syndrome (ARDS). However, over 80% of cases appear to be mild [Novel Coronavirus Pneumonia Emergency Response Epidemiology Team (2020)]. Symptoms usually begin as mild in all patients, with cough, fever, and occasional dyspnea, without a sudden onset of severe disease. In a minority of patients, severe symptoms including shortness of breath, pneumonitis and ARDS, may develop 5-8 days into the illness [Xu, Wu, Jiang et al. (2020); Guan, Ni, Hu et al. (2020); Wang, Hu, Hu et al. (2020); del Rio and Malani (2020)]. Those who become more severely ill are more likely to be male and older, with progressively more risk with each decade over the age of 50 [Novel Coronavirus Pneumonia Emergency Response Epidemiology Team (2020)]. Despite these poor outcomes, most cases are mild; and there are asymptomatic infections in all age groups and both genders, as well as among some young adults [Bai, Yao, Wei et al. (2020); Kam, Yung, Cui et al. (2020); National Institute of Infectious Diseases (2020); Cai, Xu, Lin et al. (2020)].

Acute respiratory distress syndrome has been a key feature of the pathophysiology and clinical course of declining outcomes in COVID-19 patients [Liu, Sun, Li et al. (2020); Xu, Shi, Wang et al. (2020)]. These clinical declines appear to be due to the virus alone; superinfections have been rare in reported studies [Yang, Yu, Xu et al. (2020)].

Clinical outcomes have varied greatly within China, where mortality has been substantially lower outside of Hubei province [World Health organization (2020)]. In Wenzhou, Zhejiang, China, which has faced one of the larger outbreaks outside of Hubei, there has only been one death, which occurred outside of this study [Health Commission of Wenzhou (2020)].

Faced with the initial mild presentation of COVID-19 in patients, it may be difficult to identify who will more likely develop severe illness by using established risk factors alone such as age, gender, and comorbidities. The ability to predict which patients on presentation are more likely to develop ARDS, and as such to face a greater risk of complications including death, would assist in triage. This is particularly important in a novel and accelerating outbreak when critical care resources and hospital beds are limited, and clinicians are forced to make difficult decisions without past specific experience to guide clinical acumen.

Artificial intelligence (AI) has begun to tackle these difficult challenges in healthcare, and can provide clinical decision support if used carefully [Shortliffe and Sepulveda (2018); Gianfrancesco, Tamang, Yazdany et al. (2018)]. Deep learning algorithms can use large data sets to identify risk, oftentimes based on unexpected characteristics [Sun and McIntosh

(2018)]. Such techniques can predict risk of myocardial infarction from retinal images, risk stratify Ebola patients, or screen chest imaging for tuberculosis and mammograms for cancer [Geras, Wolfson, Shen et al. (2017); Colubri, Silver, Fradet et al. (2016); Qin, Sander, Rai et al. (2019); Bai, Yao, Wei et al. (2020)]. Decision trees, a predictive analytics technique used in this study, have been previously used for pneumonia risk prediction [Caruana, Lou, Gehrke et al. (2015)].

There are as of yet no prediction models for this novel infection, which has a different clinical trajectory than many other pneumonias; there are also no public patient datasets for further validation. This represents a first step which will require further validation using different models to identify similar results, albeit the current study provides insight to help doctors in real time. AI applications present opportunities for the future of healthcare and can be harnessed at this time, as clinicians take on the complexities of responding to COVID-19 [Gianfrancesco, Tamang, Yazdany et al. (2018); Yang, Yu, Xu et al. (2020); Richardson, Griffin, Tucker et al. (2020)]. Here, we explore the use of data available on initial presentation with a novel illness to better predict who will develop more serious disease.

2 Experimental design and participants

This case series was approved by the institutional ethics board of Wenzhou Central Hospital and Cangnan People's Hospital in Wenzhou, China.

All consecutive patients with confirmed COVID-19 admitted to Wenzhou Central Hospital and Cangnan People's Hospital in Wenzhou, China, during the time period of the study, were included. Wenzhou Central Hospital is one of the major tertiary teaching hospitals and is responsible for the treatments for COVID-19 assigned by the government. Another hospital later began admitting patients after this study began. There has only been one reported death in Wenzhou as of yet, but not in this case series [Health Commission of Wenzhou (2020)]. A line listing was developed for expected clinical and epidemiologic characteristics of patients. Patient medical records were collected and analyzed by the clinical team from Wenzhou Central Hospital and Cangnan's People Hospital. Epidemiological past medical history, clinical, laboratory, and radiological characteristics on admission presentation were collected and analyzed by physicians. Treatments (including antivirals, corticosteroids, antibiotics, and IVIG) and clinical outcomes data were obtained as they were collected over the following weeks until all patients were discharged. ARDS was the endpoint of interest and was identified by clinicians during the course of hospitalization using the Berlin definition [The ARDS Definition Task Force (2012)]. The durations from any known exposure and onset of disease to hospital admission and length of stay were recorded.

3 Data description

Of the 53 hospitalized patients with COVID-19 in the dataset, all tested positive by throat swab with real-time reverse transcription polymerase chain reaction (RT-PCR) assay.

3.1 Presenting characteristics

The median age was 43 years (interquartile range 32-48.5, 13-67 years) and 33 (62.2%)

were men (Tab. 1).

Table 1: Baseline characteristics of patients diagnosed with COVID-19

| | No. (%) |
|---------------------------|---------------------|
| | Total (N=53) |
| Age, median (IQR), y | 43 (32-48.5) 13-67 |
| Sex | |
| Female | 20 (37.8) |
| Male | 33 (62.2) |
| Wuhan Exposure | 36 (67.9) |
| Infected | |
| Hospitalized patients | 53 (100) |
| Medical staff | 0 (0) |
| | Total (N=53) |
| Signs and Symptoms | |
| Fever | 47 (88.7) |
| Cough | 32 (60.4) |
| | Total (N=40) |
| Wet Cough | 13 (32.5) |
| Diarrhea | 5 (12.5) |
| Dyspnea | 9 (22.5) |
| Myalgias | 4 (10.0) |
| Wheezing | 9 (22.5) |
| | Total (N=29) |
| Nasal Congestion | 2 (6.8) |
| Sore Throat | 4 (13.9) |
| Hemoptysis | 0 (0.0) |
| | Total (N=33) |
| Comorbidities | |
| Hypertension | 7 (21.2) |
| Hepatitis B | 3 (9.0) |
| Unspecified Liver Disease | 1 (3.0) |
| Diabetes | 1 (3.0) |
| Gout | 1 (3.0) |
| Reported tobacco use | 2 (6.0) |
| | Total (N=53) |
| | Median (IQR) |
| Duration of symptoms | 3 (1-5) |

Common symptoms included fever (in 47 patients, 88.7%) and cough (in 32, 60.4%). Of the 40 patients from Wenzhou Central Hospital, a wet cough (in 13, 32.5%), diarrhea (in 5, 12.5%), dyspnea (in 9, 22.5%), myalgias (in 4, 10.0%), and wheezing (in 9, 22.5%) were also reported. Nasal congestion was noted in 2 (6.8%) and sore throat in 4 (13.9%) of the first 29 patients at

Wenzhou Central Hospital; none of these first 29 patients had hemoptysis (Tab. 1).

On admission, the median white blood cell count was 4.8×10^9 cells/L (interquartile range 3.6-6.4, 2.3-13.6) and the median lymphocyte count was 1.2×10^9 cells/L (interquartile range 0.9-1.63, 4-2.8) among all 53 patients (Tab. 2).

Table 2: Findings on presentation to the hospital of patients diagnosed with COVID-19

| | Normal Range | Median (IQR) |
|--|--------------|-----------------------------|
| | | Total (N=53) |
| White blood cell, $\times 10^9$ /L | 3.5-9.5 | 4.8 (3.6-6.4) |
| Lymphocyte count, $\times 10^9$ /L | 1.1-3.2 | 1.2 (0.9-1.63) |
| | | Total (N=40) |
| Hemoglobin, g/dl | 12.8-16.5 | 13.7 (12.9-14.4) |
| Platelets, $\times 10^9$ /L | 125-350 | 170.5 (130-221) |
| ESR, mm/hour | <17.5 | 31.5 (19.3-42) |
| CRP, mg/L | <3 | 20.8 (7-25.7) |
| AST (aspartate aminotransferase), U/L | 15-40 | 25.5(21-37) |
| ALT (alanine aminotransferase), U/L | 9-50 | 24 (15-40.5) |
| Sodium, mEq/L | 135-145 | 137.9 (136.1-139) |
| Potassium, mEq/L | 3.5-5 | 3.55 (3.4-3.7) |
| Creatinine, micromoles/L | 64-104 | 64 (55-75) |
| Creatinine kinase, U/L | <25 | 80.5 (52-112) |
| Lactate dehydrogenase, U/L | 125-243 | 205.5 (178-259) |
| Glucose, mmol/L | 3.9-7.1 | 5.6 (5.2-6.84) |
| Fio2% | 0.21 | 0.21 (0.21-0.29; 0.21-0.37) |
| | | Total (N=29) |
| Neutrophil, $\times 10^9$ /L | 1.8-6.3 | 3 (1.9-4) |
| | | Total (N=26) |
| Troponin-I, ug/L | <0.04 | 0.02 (0.02-0.05) |
| | | Total (N=7) |
| Cycle Threshold values | >40 | 28.2 (26.5-31.6) |
| | | Number (%) |
| | | Total (N=35) |
| BNP, ng/L normal | <125 | 35 (94.3) |
| | | Total (N=16) |
| Procalcitonin, ng/ml normal | <0.25 | 15 (93.8%) |
| | | Total (N=49) |
| Radiologic Findings | | |
| Ground Glass Opacities or similar findings | normal | 43 (87.7) |
| | | Total (N=40) |
| Supplemental Oxygen received | 0.21 | 11 (27.5%) |

Further clinical information was available only for Wenzhou Central Hospital's patients. Other labs were available on admission for the 40 Wenzhou Central Hospital's patients: median hemoglobin was 13.7 g/dl (interquartile range 12.9-14.4, 11.2-16.2), median platelets was 170.5×10^9 cells/L (interquartile range 130-221, 94-430), median erythrocyte

sedimentation rate (ESR) was 31.5 mm/hour (interquartile range 19.3-41, 3-83), median c-reactive protein (CRP) was 20.8 mg/L (interquartile range 7-25.7, 0.6-101.9), median aspartate aminotransferase (AST) was 25.5 U/L (interquartile range 21-37, 14-89), median alanine aminotransferase (ALT) was 24 U/L (interquartile range 15-40.5, 8-206), median sodium was 137.9 mEq/L (interquartile range 136.1-139, 131.9-143), potassium was 3.55 mEq/L (interquartile range 3.4-3.7, 2.9-4.4), median creatinine was 64 micromoles/L (interquartile range 55-75, 20-90), median creatine kinase was 80.5 U/L (interquartile range 52-112, 33-1725), median lactate dehydrogenase was 205.5 U/L (interquartile range 178-259, 128-402), median glucose was 5.6 mmol/L (interquartile range 5.2-6.84, 4.3-9.1). Procalcitonin was available for 16 patients; the value was less than 0.25 ng/ml in 15 (93.8%) and 0.05 ng/ml or lower in 13 (81.3%) (Tab. 2).

Some labs were not available on presentation for all patients. Neutrophil counts were available for 29 patients at Wenzhou Hospital; median neutrophil count was 3×10^9 cells/L (1.9-4, 1.4-7.6). D-dimer was available for 27 patients at Wenzhou Central Hospital; median was 155.5 mg/L (interquartile range 110.3-207.5; 43-747). Troponin-I was available for 26 patients at Wenzhou Central Hospital and the median was 0.02 ug/L (interquartile range 0.02-0.05, 0-0.09). BNP was over 125 ng/l in 2 (6%) of 35 patients at Wenzhou Central Hospital.

Although all tested positive for SARS-CoV2, specific cycle threshold values were charted and available for 7 patients and ranged from 25-38 with a median of 28.2.

Of the 49 patients for whom initial radiology was available, 6 (12.2%) had a normal chest computed tomography; the rest (87.8%) had ground glass opacities or equivalent findings. No patient with a normal CT scan developed ARDS.

Median days from symptom onset to hospitalization was 3 days (IQR range 1-5, 0-30). 36 infections were described as due to exposures in Wuhan; the rest locally in Wenzhou.

None of the patients in the dataset were healthcare workers or the result of known nosocomial spread. One patient was identified on a chest computed tomography and developed symptoms the following day. No patients presented more than once.

None were pregnant. No patient reported receiving the flu vaccine. Of the 33 patients at Wenzhou Central Hospital interviewed on smoking status, only two admitted to any smoking and reported they had quit. 7 were reported to have known hypertension, 3 were reported to have hepatitis B and another had unspecified liver disease. 1 had known diabetes, 1 had known gout and 1 had post brain aneurysm surgery. Liver enzymes for 3 of the 4 patients with liver disease, showed no elevation. Aspartate aminotransferase (AST) ranged from 13-50 U/L, alanine aminotransferase, 17-58 U/L in those with liver disease or hepatitis; one had a mild elevation with AST 50 and ALT 58.

On presentation, 11 of 40 patients (27.5%) at Wuhan Central Hospital were initiated on oxygen therapy. In total, median fraction of inspired oxygen (FiO₂) was 0.21 (interquartile range 0.21-0.29; 0.21-0.37). Of those on oxygen, 10 out of 11 were started on FiO₂ 0.29 (approximately 2 l/min), except for one on FiO₂ 0.37 (approximately 4 l/min). Of those 5 who would develop ARDS, 3 (60%) were initiated on oxygen at presentation; FiO₂ was 0.29 (approximately 2 l/min) for 2 and 0.37 (approximately 4 l/min) for 1.

Of those patients who developed ARDS, all were male, median age 46 (interquartile range

41-62; 23-67). None smoked. Four of the five had either high blood pressure or a potentially related condition (past brain aneurysm). All had fever, cough, wheezing, and dyspnea on arrival. 2 had myalgias. None had chills. Median CRP was 31.1 (interquartile range 22.8-37.8; 5.2-101.9), ALT 44 (interquartile range 24-66; 24-70) and hemoglobin 13.7 g/dl (interquartile range 12.8-14.4; 11.6-14.7). Highest CRP was 68.6. Procalcitonin was low in all tested (3 had values <0.25). None had a troponin over 0.05 microg/L. Two were on room air on arrival; two were on 29% FiO₂ and one was on 37% FiO₂.

3.2 Characteristics of hospital courses

During hospitalization, a total of 19 patients (47.5%) at Wuhan Central Hospital were treated with oxygen. Of those 19 on oxygen, median FiO₂ was 0.29 (interquartile range 0.29-0.35; 0.29-0.50) and 13 did not received more than FiO₂ 0.29 (2 l/min). No patients were intubated, required ECMO or dialysis. 5 developed ARDS, as charted by a clinician with information on PaO₂, FiO₂, chest imaging, and clinical course. Only 1 was cared for in the ICU and none died (Tab. 3).

During hospitalization, the peak procalcitonin was available for 16 cases in Wenzhou Central Hospital; the median was 0.045 ng/ml (interquartile range 0.02-0.095; <0.02-2.49); the value was greater than or equal to 0.05 ng/ml in 25%, and only 2 saw values over 0.25 ng/ml (12.5%). Peak CRP was available for 27 cases in Wenzhou Central Hospital; the median was 25.9 U/L (interquartile range 17.55-39.6, 1.5-129) (Tab. 3).

Table 3: Characteristics of hospitalization course for patients with COVID-19

| | No. % |
|--|---------------------|
| Laboratory values | |
| | Total (N=16) |
| Peak procalcitonin | 0.045 (0.02-0.095) |
| | Total (N=27) |
| Peak CRP | 25.9, (17.55-39.6) |
| Complications | Total (N=53) |
| ARDS | 5 (9.4) |
| ICU | 1 (1.9) |
| Supplemental Oxygen | 19 (47.5%) |
| Length of Stay | 27 (23-31.5) |
| Treatment | |
| Ritonavir/Lopinavir | 53 (100) |
| Umifenovir | 29 (55.7) |
| | Total (N=29) |
| Moxifloxacin | 4 (13.9) |
| Levofloxacin | 3 (10.3) |
| Amoxicillin/Clavulanic | 7 (24.1) |
| Moxifloxacin/Biopenem | 1 (3.4) |
| Moxifloxacin/Amoxicillin/Clavulanic | 1(3.4) |
| IVIg | 6 (20.9) |
| Methylprednisolone | 2 (6.9) |
| Rectal suppositories of recombinant human Interferon | 43 (81.1) |

All 53 patients took lopinavir and litonavir tablets; dose was 200 mg twice a day of lopinavir. 29 patients took umifenovir. 43 patients took rectal suppositories of recombinant human interferon-2a (Tab. 3).

Of the first 29 patients at Wenzhou Central Hospital, there was clinical concern for secondary infection in 13 (44.8%) based on clinical exam or chest computed tomography; no bacterial or fungal cultures are available. Of these 29 patients, 28 were determined to have pneumonia affecting both lungs. 13 patients were reported to have received antibiotics with courses largely for 3 days. 6 patients received amoxicillin/clavulanic injections 2.4 gm twice a day; 1 other received reduced amoxicillin/clavulanic dosing 600 mg twice a day intravenously (IV). 3 received 500 mg of levofloxacin, orally daily. 4 received 400 mg of moxifloxacin, orally daily. Courses involved 2 antibiotics in 2 additional patients: one changed from moxifloxacin to biapenem 300 mg three times a day, another changed to amoxicillin/clavulanic 2.4 gm twice a day, after 2 days of moxifloxacin. In addition, 2 patients were known to have received steroids (IV methylprednisolone) and 6 were known to have received IVIG.

All 53 patients have now been discharged. The median length of stay was 27 days (interquartile range 23-31.5, 9-45). Discharge required normal temperature for over three days, no respiratory and gastrointestinal symptoms, PCR swab negative twice over at least 2 days, and PCR stool sample negative as well.

4 Methods

Predictive analytics (a form of artificial intelligence) learns from historical data to help predict future outcomes. The technology uses machine learning algorithms that can extract insights and rules from experience (historical examples) in order to determine data attributes (features) with the most predictive power for making accurate predictions.

In predictive analytics, a feature (also known as variable, or observation), is an individual measurable attribute. The features for every patient that we considered in this analysis are outlined in Tabs. 1 and 2, and represent baseline characteristics.

The predictive analytics problem addressed in this report can be formulated as follows: Given historical data of patients that tested positive for coronavirus, identify features from Tabs. 1 and 2 that are predictive to ARDS; hence the severity of the patients with COVID-19. In the following section, we explain the data pre-processing methods, feature engineering algorithms and predictive models used in this study.

In predictive analytics, feature engineering (also known as feature selection) is the process of algorithmically reducing the dimensionality of the feature space to a smaller set of features with higher predictive power vis-a-vis the predictive label, which is ARDS in our case. The goal is to identify the best subset that contains the least number of dimensions that most contribute to the accuracy of predicting ARDS.

There are two major types of feature selection methods: filter methods and wrapper methods. Filter methods used in this study are based on entropy that is widely applied in information theory. In the case of this experiment, entropy is a metric that measures how much information a feature encapsulates to help predict the final class label ARDS of the

sample. The higher the entropy of a feature, the more variance that feature exhibits, and thus the more likely that feature contains valuable information for predicting the final label. In our case, the entropy of any discrete variable X is given by Bellaachia et al. [Bellaachia and Bari (2012)] as:

$$H(X) = -\sum_{\text{all possible } x} P(X = x) \cdot \ln(P(X = x)) \quad (1)$$

4.1 Information gain

We also adopted information gain as a measure to rank features. Each feature is assigned a value corresponding to its information gain, the amount of information acquired after knowing the value of the feature, with respect to the class label ARDS. The formula for information gain is represented in Eq. (2):

$$IG(V) = -\sum_{c \in \{\text{positive, negative}\}} P(c) \log_2 P(c) + \sum_{j=1}^m P(v_j) \sum_{c \in \{\text{positive, negative}\}} P(c|v_j) \log_2 P(c|v_j) \quad (2)$$

In Eq. (2), V represents the feature variable, c represents the class label, m represents the total number of subcategories of the dataset if the dataset is categorized by feature V .

4.2 Gini index

Each feature is assigned a value based on the Gini index, a measure of the impurity of the dataset. Assuming we split the dataset based on each feature, features that resulted in less impure class distributions are assigned higher values. The formula for the Gini index is shown below:

$$GI(V) = \sum_{j=1}^m P(v_j) [1 - P(\text{positive}|v_j)^2 - P(\text{negative}|v_j)^2] \quad (3)$$

In the above formula, V represents the feature variable, m represents the total number of subcategories of the dataset if split by feature V .

4.3 Chi-Squared statistics

Each feature is assigned a value based on the Chi-Square value, a statistical measure that indicates how dependent two variables are, for the class label variable and that feature variable. The higher the Chi-Square value, the more the class label is dependent on the given feature. The formula for the Chi-Squared Statistics is shown below:

$$\chi_v^2(V) = \sum_{i=1}^{2j} \frac{(O_i - E_i)^2}{E_i} \quad (4)$$

In Eq. (4), V represents the feature variable, v represents the degrees of freedom that will be used in the Chi-Squared test without affecting the value calculated, O represents the observed frequency, E represents the expected frequency, j represents the total number of possible values of the feature V , 2 represents the number of different class labels in this case.

We also adopted the feature engineering methods that are based on wrappers. A wrapper-based method uses greedy algorithms to select the best features, as opposed to ranking the feature. We adopted the forward selection algorithm that starts with an empty set of features, and iteratively adds features to the set until the inclusion of additional features stops improving the framework performance.

5 Experiment results

In the experiment, the predictive power for developing ARDS among patients testing positive for SARS-CoV2 was evaluated with the following features: lymphocyte count, white blood count, temperature, cycle threshold, creatinine, hemoglobin, gender, CRP, age, fever, CK, LDH, Glu, ALT, AST, K+ and N+; as listed in Tab. 4.

Table 4: Feature ranking

| Predictive features of ARDS in this order using feature selection algorithms described in Section 4. | |
|--|-------------------|
| 1. | ALT |
| 2. | Myalgias |
| 3. | Hemoglobin |
| 4. | Gender |
| 5. | Temperature |
| 6. | Na+ |
| 7. | K+ |
| 8. | Lymphocyte Count |
| 9. | Creatinine |
| 10. | Age |
| 11. | White Blood Count |

The accuracy of the algorithms is based on 10-fold cross validation. The algorithms were applied to all patients who tested with COVID-19 in order to predict ARDS using high predictive features following the previous experiment with ALT, myalgias, hemoglobin, gender, temp, Na+, K+, lymphocyte count, creatinine, age and white blood count. A decision tree based on the one feature ALT reached a 70% accuracy. Overall accuracy collectively reached 70-80% as shown in Tab. 5. The most predictive features were alanine aminotransferase (ALT), myalgias, and hemoglobin, in this order.

Table 5: Predictive algorithms accuracy

| Predictive Algorithm | Accuracy |
|-------------------------------------|----------|
| Logistic Regression | 50% |
| KNN (k=5) | 80% |
| Decision Tree (based on Gain Ratio) | 70% |
| Decision Tree (based on Gini Index) | 70% |
| Random Forests | 70% |
| Support Vector Machine | 80% |

There were no funds or time allocated for patient and public involvement (PPI), so we were unable to involve patients. However, we have invited patients to help us develop our dissemination strategy.

6 Discussion

Artificial intelligence (AI) can be used to recognize unexpected patterns in novel clinical presentations. Such tools can fine-tune a clinician's ability to detect a "sick" versus "not sick" diagnosis in relation to a previously unencountered infection like COVID-19.

A framework was developed here to identify the few who would develop more severe illness, specifically ARDS, from among the many patients with mild initial presentations of COVID-19. Decision trees, random forests and support vector machines are the types of machine learning models used in this study. These tools sometimes have advantages over more traditional methods such as logistic regression, which was found here to be much less predictive. This approach can more adeptly work with datasets that are small or imbalanced as seen in early epidemics, or containing data that is not linearly separable as expected with disease data analytics. In turn, this approach better allows for actionable responses early in an epidemic rather than waiting for large, complete data sets. It also avoids reliance on a single biomarker or a difficult to obtain lab, and instead pulls together an ensemble of common predictors for robust determinations. Furthermore, in the predictive analytics models adopted in this study, the decision thresholds are determined automatically during the model training process, rather than being set manually and potentially more arbitrarily as in logistic regression. We intend to expand the algorithms used here to deep learning algorithms and swarm intelligence as soon as we can obtain larger clinical data [Bellaachia and Bari (2012)].

Characteristics that are hallmarks of COVID-19 diagnosis, such as ground glass opacities on chest computed tomography, as well as fever, cough, and lymphopenia, assist in initial clinical diagnosis. However, given uniformity, these did not distinguish risk of disease progression and were not highly predictive. Trending some clinical values over the course of the illness has also been shown to be predictive, though this will not help with initial triage and resource allocation [Wang, Hu, Hu et al. (2020)]. Other characteristics, including older age, male gender, and comorbidities, have been associated with worsened outcomes in separate studies [Xu, Wu, Jiang et al. (2020); Guan, Ni, Hu et al. (2020); Wang, Hu, Hu et al. (2020)]. Age and gender were not strong predictors of outcome, as severe or mild cases have been found in all age groups. In our study, patients were in their thirties or forties and none were over age 67. Identifying risks beyond gender, age, and comorbidities will be particularly important in identifying those young adults who will go on to develop ARDS [Wang, Hu, Hu et al. (2020)].

Cycle threshold (Ct) might have been expected to predict severity of this coronavirus, as it does in other infections including in another emerging disease Ebola [Crowe, Maenner, Kuah et al. (2016)]. However, in COVID-19, Ct has so far not been shown to be predictive; high viral loads (low Ct) have been seen in asymptomatic individuals and peak soon after the onset of symptoms which may be well before some patients present to medical care [National Institute of Infectious Diseases (2020); Huang, Wang, Li et al. (2020); Zou, Ruan, Huang et al. (2020)].

Other well validated testing tools for pneumonia severity did not perform well. None of the patients who developed ARDS would have met the criteria for requiring hospitalization as set by the pneumonia patient outcomes research team (PORT) score.

Instead, a combination of factors commonly collected at first presentation was found to predict disease progression to ARDS. As there were no deaths in this study, ARDS represents the clinical syndrome of significance and has been associated with death in other studies [Liu, Sun, Li et al. (2020); Xu, Shi, Wang et al. (2020)].

The features that machine learning showed best predicted ARDS were not the indicators a

clinician would standardly select, nor were these values grossly abnormal clinically. Multiple iterations showed that the most predictive features included an increase in alanine aminotransferase (ALT) and hemoglobin, and the presence of myalgias.

These features do not need to be causal to be predictive, but correlations do raise clinical questions for physicians to consider. Liver function tests were not substantially elevated in our study; none of the patients with liver disease developed ARDS, yet small elevations in ALT featured prominently in the predictions. Myalgias are not normally featured in classification of illness severity, but could represent generalized inflammatory and cytokine response not captured well by other indicators. Higher hemoglobin levels were associated with poorer outcomes; this may be due to correlation with other factors including male gender or even unreported tobacco use.

Other data more commonly expected to affect clinical risk did contribute, but to a lesser degree. Gender, temperature, and sodium further added to the predictive features of this model, as do potassium, lymphocyte, creatinine, age, and white blood cell count. Other factors like dyspnea on presentation, which would have logically been expected to correlate with ARDS, did not feature as predictive in this model, as this was common in many patients.

7 Conclusion

This study explored the clinical spectrum of illness and predictive indicators in a case series from Wenzhou, Zhejiang, China. The outbreak seen has not matched the clinical severity seen in the initial Wuhan, Hubei epicenter or in other settings where lab testing and clinical surge capacity has been stretched. Reduced caseloads from public health measures, coupled with increased surveillance, may have resulted in detection of a larger number of milder cases. Clinical management, including treatment with antivirals, and clinical insights from earlier in the epidemic, may have modified the course as well.

A clear limitation of this study is the size of the dataset; 53 patients with some incomplete data as well as a limited spectrum of severity. Overall, the models were 70%-80% predictive for this population. Further validation and refinement of this model will require data describing a wider clinical spectrum.

Nonetheless, this study shows that predictive analytics can play a role in augmenting clinical skills in distinguishing between “sick” from not “sick”. The model highlights that some pieces of clinical data may be underappreciated by clinicians, such as mild increases in ALT and hemoglobin as well as myalgias. Key characteristics predictive of diagnosis, including fever, lymphopenia, chest imaging, were not as predictive of severity. Likewise epidemiologic risks such as age and gender were not as predictive; all ARDS patients in this study were male but most males did not develop ARDS.

Just as predictive text is intended to augment, but not replace writers, the goal is not to create a black box to supersede clinical reasoning, but to create models that can provide insight. Clinical acumen is based on both personal learning and collective professional learning; machine learning can add further insight.

AI tools need to be developed iteratively and include clinicians in their development to be clinically applicable. Further refinement of these models with more data, from different settings with different spectrums of severity, would strengthen the predictive power of the

model and allow it to be a useful tool in identifying early from the many with COVID-19, who will develop more serious disease and require closer clinical attention and resources including early initiation of treatments, which will likely be in limited supply, if available in the future.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

Contributions of Authors: **XJ** is the Chief Physician in the Department of Infectious Disease in Wenzhou Central Hospital and was responsible for the care of patients and data collection.

MC designed the study and directed specific data collection and data analytics, responsible for writing with **AB**.

AB designed the data analytic experiments, responsible for writing with **MC**.

XJ compiled data, provided translations, involved in article editing.

References

Bai, Y.; Yao, L.; Wei, T.; Tian, F.; Jin, D. et al. (2020): Presumed asymptomatic carrier transmission of COVID-19. *Journal of the American Medical Association*.

Bellaachia, A.; Bari, A. (2012): A flocking based data mining algorithm for detecting outliers in cancer gene expression microarray data. *International Conference on Information Retrieval and Knowledge Management, IEEE*, pp. 305-311.

Bellaachia, A.; Bari, A. (2012): Flock by leader: a novel machine learning biologically inspired clustering algorithm. *International Conference in Swarm Intelligence*, pp. 117-126, Springer, Berlin, Heidelberg.

Cai, J.; Xu, J.; Lin, D.; Yang, Z; Xu, L; et al. (2020): A case series of children with 2019 novel coronavirus infection: clinical and epidemiological features. *Clinical Infectious Diseases*.

Caruana, R.; Lou, Y.; Gehrke, J.; Koch, P.; Sturm, M. et al. (2015): Intelligible models for healthcare: predicting pneumonia risk and hospital 30-day readmission. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1721-1730.

Colubri, A.; Silver, T.; Fradet, T.; Retzepi, K.; Fry, B. et al. (2016): Transforming clinical data into actionable prognosis models: machine-learning framework and field-deployable app to predict outcome of Ebola patients. *PLoS Neglected Tropical Diseases*, vol. 10, no. 3.

Crowe, S. J.; Maenner, M. J.; Kuah, S.; Erickson, B. R.; Coffee, M. et al. (2016): Prognostic indicators for Ebola patient survival. *Emerging Infectious Diseases*, vol. 22, no. 2, pp. 217-223.

del Rio, C.; Malani, P. N. (2020): 2019 Novel coronavirus, important information for clinicians. *Journal of the American Medical Association*.

- Geras, K. J.; Wolfson, S.; Shen, Y.; Wu, N.; Kim, S. G. et al.** (2020): High-resolution breast cancer screening with multi-view deep convolutional neural networks. arXiv preprint.
- Gianfrancesco, M. A.; Tamang, S.; Yazdany, J.; Schmajuk, G.** (2018): Potential biases in machine learning algorithms using electronic health record data. *Journal of the American Medical Association Internal Medicine*, vol. 178, no. 11, pp. 1544-1547.
- Guan, W.; Ni, Z.; Hu, Y.; Liang, W.; Ou, C. et al.** (2020): Clinical characteristics of coronavirus disease 2019 in China. *New England Journal of Medicine*.
- Health Commission of Wenzhou** (2020): March 5 report. *Health Commission of Wenzhou*.
- Huang, C.; Wang, Y.; Li, X.; Ren, L.; Zhao, J. et al.** (2020): Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*, vol. 395, no. 10223, pp. 497-506.
- Kam, K.; Yung, C. F.; Cui, L.; Lin Tzer Pin, R.; Mak, T. M. et al.** (2020): A well infant with coronavirus disease 2019 (COVID-19) with high viral load. *Clinical Infectious Diseases*.
- Liu, Y.; Sun, W.; Li, J.; Chen, L.; Wang, Y. et al.** (2020): Clinical features and progression of acute respiratory distress syndrome in coronavirus disease 2019. medRxiv.
- National Institute of Infectious Diseases** (2020): Field briefing: diamond princess COVID-19 cases. *National Institute of Infectious Diseases*.
- Novel Coronavirus Pneumonia Emergency Response Epidemiology Team** (2020): The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China. *Chinese Journal of Epidemiology*, vol. 41, no. 2, pp. 145-151.
- Qin, Z. Z.; Sander, M. S.; Rai, B.; Titahong, C. N.; Sudrungrot, S. et al.** (2019): Using artificial intelligence to read chest radiographs for tuberculosis detection: a multi-site evaluation of the diagnostic accuracy of three deep learning systems. *Scientific Reports*, vol. 9, no. 1, pp. 1-10.
- Richardson, P.; Griffin, I.; Tucker, C.; Smith, D.; Oechsle, O. et al.** (2020): Baricitinib as potential treatment for 2019-nCoV acute respiratory disease. *Lancet*, vol. 395, no. 10223, pp. e30-e31.
- Shortliffe, E. H.; Sepulveda, M. J.** (2018): Clinical decision support in the era of artificial intelligence. *Journal of American Medical Association*, vol. 320, no. 21, pp. 2199-2200.
- Sun, H.; McIntosh, S.** (2018): Analyzing cross-domain transportation big data of New York City with semi-supervised and active learning. *Computers, Materials & Continua*, vol. 57, no. 1, pp. 1-9.
- The ARDS Definition Task Force** (2012): Acute respiratory distress syndrome: the Berlin definition. *Journal of the American Medical Association*, vol. 307, no. 23, pp. 2526-2533.
- Wang, D.; Hu, B.; Hu, C.; Zhu, F.; Liu, X. et al.** (2020): Clinical characteristics of 138 hospitalized patients with 2019 novel Coronavirus-infected pneumonia in Wuhan, China. *Journal of the American Medical Association*.
- World Health Organization** (2020): Coronavirus disease 2019 (COVID-19) situation report 43.
- World Health Organization** (2020): Report of the WHO-China joint mission on coronavirus disease (COVID-19).

Xu, X.; Wu, X.; Jiang, X.; Xu, K.; Ying, L. et al. (2020): Clinical findings in a group of patients infected with the 2019 novel coronavirus (SARS-Cov-2) outside of Wuhan, China: retrospective case series. *British Medical Journal*, vol. 368.

Xu, Z.; Shi, L.; Wang, Y.; Zhang, J.; Huang, L. et al. (2020): Pathological findings of COVID-19 associated with acute respiratory distress syndrome. *Lancet Respiratory Medicine*.

Yang, X.; Yu, Y.; Xu, J.; Shu, H.; Xia, J. et al. (2020): Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *Lancet Respiratory Medicine*.

Zou, L.; Ruan, F.; Huang, M.; Liang, L.; Huang, H. et al. (2020): SARS-CoV-2 Viral load in upper respiratory specimens of infected patients. *New England Journal of Medicine*, vol. 10, no. 1056, pp. 1-3.