medRxiv preprint doi: https://doi.org/10.1101/2020.03.09.20032219.this version posted March 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Data-driven discovery of clinical routes for severity detection of COVID-19 pediatric cases

Hui Yu[1,*], M. M., Jianbo Shao[1,*], M. D., Yuqi Guo[2,*], Yun Xiang[1], M. D., Chuan Sun, Hai-Tao Zhang[2,#] , Ph.D., Ye Yuan[2,#] , Ph.D.

1. Wuhan Children's Hospital, Tongji Medical School, Huazhong University of Science and Technology, Wuhan 430010, P. R. China
2. School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, P. R. China

* Equal contribution;

Corresponding Authors: Prof. Ye Yuan and Prof. Hai-Tao Zhang, School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China. Email: {yye,zht}@hust.edu.cn

The outbreak of COVID-19 epidemic has caused worldwide health concerns since Nov., 2019 [1]. A previous study [1] described the demographic, epidemiologic, and clinical features for infected infants. However, compared with adult cases, little attention has been paid to the infected pediatric cases [2-6].

Severity detection is challenging for children since most of children patients have mild symptoms [2] no matter they are moderately or critically ill therein.

Methods

For this retrospective study, we identified 105 infected children admitted to Wuhan Children's Hospital, the sole designated hospital in Wuhan for COVID-19 children patients, from Feb. 1 to Mar. 3, 2020. The epidemiological, clinical laboratory, and outcome data were extracted from the medical records of these patients. The study was approved by Wuhan Children's Hospital Ethics Committee.

Throat-swab, anal-swab or urine specimens were taken at admission for real-time RT-PCR, which was performed by Wuhan Huada Gene Biology. Symptoms on admission were collected together with laboratory results, chest radiography and CT findings, treatment received for COVID-19 and clinical outcomes.

Severe patients are generally manually detected according to the guideline from National Health Commission, China, using clinical symptoms including shortness of breath, assisted respiration, apnea, cyanosis, dehydration, and progressive increase of lactate, etc.

In this study, a supervised decision-tree classifier was developed. All clinical measurements from the last available date were used as features and set 'mild' and 'severe' as labels. "-1" was used to complement the incomplete clinical measures to avoid bias. We used standard F1-score [4] to evaluate the performance of the classifier. We started from one feature and increased the number of features in the clinical route, until the F1-score converges. We picked the classifier with less incomplete measurements for all the patients, when two classifiers with the same number of features performed the same.

Results

105 children aged 1-16, including 64 males and 41 females, were infected with COVID-19, of which 8 were critically ill. The youngest was 1 day after birth, and the oldest was 15-year-old. Male infection rate (60.95%) is higher than that of the female (39.05%). This is opposite to a previous report [2]. Children over 6-year-old have the highest infection rate (60.95%). All the children lived in Wuhan.

Based on the current available clinical data, we discovered a clinical route that can achieve 100% F1 score (shown in Figure 1a). Figure 1b depicts the mild and severe children patients over the proposed two-feature based clinical route. As a result, we have extracted merely two features, i.e., Direct Bilirubin (DBIL) and alaninetransaminase (ALT), by which 8 critically ill pediatric cases can be precisely identified from other 97 mild patients.

Discussion

Due to the scantiness of clinical data from confirmed COVID-19 children cases, especially for severe ones, it is an urgent yet challenging

medRxiv preprint doi: https://doi.org/10.1101/2020.03.09.20032219.this version posted March 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

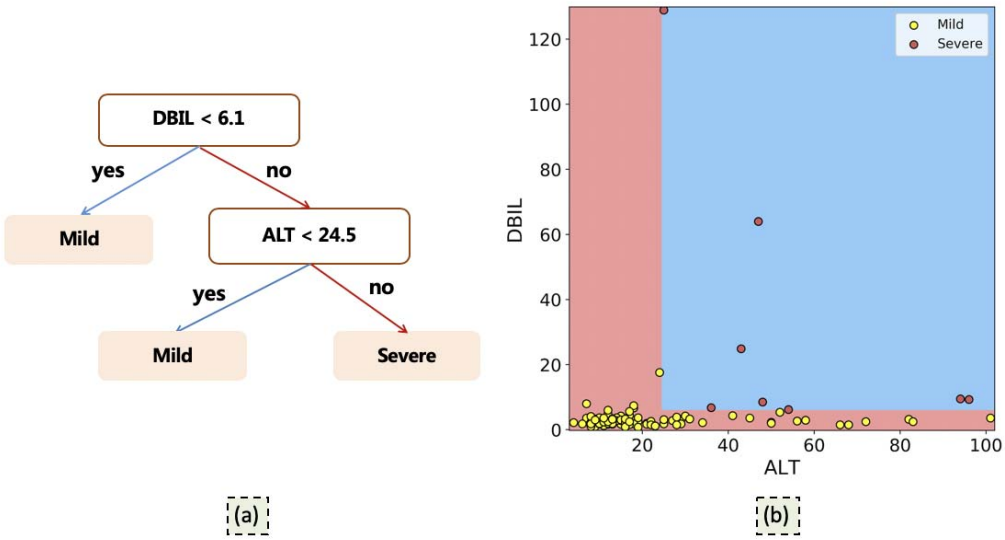

Figure 1. (a) A diagram showing identification of severe cases using the discovered clinical route. These two features are automatically selected from over 300 features from the proposed classification algorithm. (b) Visualization of mild and severe patients over the selected features.

mission to promptly distinguish the severe ones from the mild cases for early diagnosis and intervention. To this end, with the assistance of machine learning methods, we identified that DBIL and ALT, surfacing from over 300 clinical features, were able to serve as a combination index to screen out all the critically ill cases. Although the increase of DBIL and ALT has been reported to reflect tissue destruction or injury, for the first time, their combination is revealed as a precise indicator for the severity of COVID-19 pediatric cases, which is quite different from the discovered clinical route for adult [4].

The study was limited to a small number of patients from a single center in Wuhan. Further studies from multiple centers in a larger cohort would be beneficial to the validation of the proposed route as well as understanding of the disease.

Word count of the manuscript text: 598

Reference

1. Paules CI, Marston HD, Fauci AS. Coronavirus infections more than just the common cold. *JAMA*, doi:10.1001/jama. 2020.0757
2. Wei M, Yuan J, Liu Y, Fu T, Yu X, Zhang Z-J, Novel Coronavirus Infection in Hospitalized Infants Under 1 Year of Age in China. *JAMA*, doi:10.1001/jama.2020.2131.
3. Chen N, Zhou M, Dong X., *et al*., Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The Lancet*, 395: 507-513, 2020.
4. Yan L, Zhang H-T, Xiao Y, *et al*, Prediction of survival for severe Covid-19 patients with three clinical features: development of a machine learning-based prognostic model with clinical data in Wuhan, *medRxiv*, doi: 10.1101/2020.02.27.20028027.
5. Wang D, Hu B, Hu C, *et al.* Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China. *JAMA*, doi:10.1001/jama.2020.1585.
6. Guan WJ, Ni ZY, Hu Y, *et al.* Clinical characteristics of 2019 novel coronavirus infection in China. *New England Journal of Medicine*, doi: 10.1056/NEJMoa2002032.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.09.20032219.this version posted March 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Author Contributions

H. Yu designed the clinical study; Y. Yuan conceptualized the idea; H. Yu, J. Shao, H.-T. Zhang and Y. Yuan collected and statistically analyzed the data; H. Yu, H.-T. Zhang and Y. Yuan wrote the letter; H. Yu carried out the clinical laboratory inspection; Y. Yuan and Y. Guo discovered the clinical features and routes.

Conflict of Interest Disclosures:

None reported.

Funding/Support:

None