Integrated medical resource consumption stratification in hospitalized patients: an Auto Triage Management model based on accurate risk, cost and length of stay prediction

Qin Zhong††, Zongren Li‡, Wenjun Wang§, Lei Zhang∥ & Kunlun He1,2,3,4,5*

1Medical Big Data Research Center, Chinese PLA General Hospital, Beijing 100853, China; 2Medical Artificial Intelligence Research Center, Chinese PLA General Hospital, Beijing 100853, China; 3Bio-engineering Research Center, Chinese PLA General Hospital, Beijing 100039, China; 4Department of Cardiology, Fourth Medical Center of Chinese PLA General Hospital, Beijing 100048, China; 5Key Laboratory of Ministry of Industry and Information Technology of Biomedical Engineering and Translational Medicine, Chinese PLA General Hospital, Beijing 100048, China; 6China-Australia Joint Research Center for Infectious Diseases, School of Public Health, Xi’an Jiaotong University Health Science Center, Xi’an 710061, China

Received May 19, 2021; accepted August 13, 2021; published online October 8, 2021

INTRODUCTION

Healthcare systems around the world are dealing with depleting resources at a time when the demand for healthcare is rapidly rising (Hyde, 2020). Even the United States is facing unprecedented regional resource shortages (Gatesman and Smith, 2011), already limited medical resources will likely become even scarcer in low-income countries (McMahon et al., 2020). Therefore, accurate prediction of integrated medical resource consumption by hospitalized patients and reasonable triage will help rescue efficiency and alleviate the shortage of regional medical resources.

Triage based on disease progression distinguishes patients at various disease risk levels, which helps hospitals make differentiated treatment plans for patients. Besides, long-
term hospitalization and high economic costs bring substantial pressure to medical systems and insurance. Cost is not just a reflection of hospitalisation time, since variation in patients’ treatment plans and drug administration indeed cause different money expenses, although they may be under the same length of stay (LoS). Therefore, besides the risks, predicting patients’ LoS and money expense can aid resource preparation in time.

An accurate prognostic model for severe disease manifestations is long-term needed due to the limited health care resources available. Various machine learning methods have been tried to establish a reliable prediction model. For example, Emergency Heart Failure Mortality Risk Grade helps identify a heart failure patient’s risk level (Lee and Ezekowitz, 2014). Curiati et al. (2020) developed The PRO-AGE Scoring System to predicting patients’ prolonged length of stay in emergency department. However, rarely comprehensive and multi-dimension model seems to be available on longitudinal data throughout the disease course. A patient stratification model to predict the deterioration rate, the LoS, and the hospitalization costs of the remaining treatment at any point in the course of the disease is required, especially for new emerging disease.

In this study, we developed a deep learning-based framework called Auto Triage Management (ATM). It introduces AutoML technologies and residual fully connected neural network (RFCN) (Liu et al., 2019) as a neural network backbone. By training on electric health record (EMR) data, ATM could be used to predict the risk of developing a critical illness, the hospitalization LoS, the economic costs for patients to recovery and the treatment suggestion. Furthermore, we applied the ATM framework on data collected from 3,989 coronavirus disease 2019 (COVID-19) patients into mild (27 cases in H.S.S., 59 cases in G.G.), moderate (1,973 cases in H.S.S., 1,155 cases in G.G.), severe (682 cases in H.S.S., 40 cases in G.G.), critical (41 cases in H.S.S., 9 cases in G.G.) levels when at-admission (Figure 2A). All patients completed their hospital course at study end (discharged alive or dead). The mortality outcomes of those with confirmed COVID-19 were also counted, including 2,662 cured cases, 60 deceased cases in H.S.S. and 1,261 cured cases, four deceased cases in G.G. (Figure 2B). Patients’ average LoS and total cost were also analyzed for different severity levels and age groups (Figure 2C and D). The progression event was defined as a composite event of requiring invasive mechanical ventilation, or admission to the intensive care unit, or death in the hospital. In the following analysis, we used H.S.S. for training and processed external validation in G.G.

Disease progression prediction

In order to build a disease progression prediction model, we firstly searched for the best features for analysis. We selected 19 clinical features that can best predict COVID-19 progression risk from total 75 features (concordance index (CI)=0.92 and area under the receiver operating characteristics (AUC) =0.94 in H.S.S.), outperforming the seventh version of the Chinese guideline (CI=0.82, DeLong test P<5×10−5, Figure 3C–E, Methods, Figure S1A in Supporting Information).

RESULTS

Auto Triage Management framework for risk, cost and LoS prediction

We built a deep learning based framework for biomedical researchers to easily obtain the optimal models for progression risk prediction, treatment search, total cost prediction, and LoS prediction using given EMR data, called ATM, which can analyze data of patients’ disease condition and health economics index prediction during hospitalization (Figure 1). We invented two novel deep learning networks—RFCN-Cox and RFCN-Reg for progression, total LoS and cost prediction, and used random forest for prediction of patients’ today-and-before-today hospitalized condition, and integrated them into the ATM framework. To synthetically evaluate ATM, we totally included five existing relevant algorithms for comparison in the following analysis.

Data sources and characteristics

We collected two independent datasets from Huoshenshan (H.S.S.) Hospital and Guanggu (G.G.) Hospital in Wuhan, China. They cover 2,724 and 1,265 COVID-19 patients, containing 75 and 68 patients’ at-admission features of initial laboratory tests, signs and vital indexes, baseline co-morbidities, living habits and demographic information, respectively. The average ages of H.S.S. and G.G. patients are 58.5 and 57.8 years old, and the sex ratio (male vs. female) are 0.49 and 0.42, respectively. The patients were admitted to hospitals between February 4, 2020, and April 15, 2020. The seventh version of the Chinese guideline grouped these patients into mild (27 cases in H.S.S., 59 cases in G.G.), moderate (1,973 cases in H.S.S., 1,155 cases in G.G.), severe (682 cases in H.S.S., 40 cases in G.G.), critical (41 cases in H.S.S., 9 cases in G.G.) levels when at-admission (Figure 2A). All patients completed their hospital course at study end (discharged alive or dead). The mortality outcomes of those with confirmed COVID-19 were also counted, including 2,662 cured cases, 60 deceased cases in H.S.S. and 1,261 cured cases, four deceased cases in G.G. (Figure 2B). Patients’ average LoS and total cost were also analyzed for different severity levels and age groups (Figure 2C and D). The progression event was defined as a composite event of requiring invasive mechanical ventilation, or admission to the intensive care unit, or death in the hospital. In the following analysis, we used H.S.S. for training and processed external validation in G.G.
When validated in G.G., it still showed a high performance (CI=0.85, AUC=0.87), although G.G. lacks a feature (α-hydroxybutyrate dehydrogenase (α-HBDH)) required by the model. Immune cell percentages and counts, C reactive protein (CRP), lactate dehydrogenase (LDH), D-dimer and age were good prediction markers with CI>0.7 in both hospitals (Figure 1A). Interestingly, neutrophils and lymphocyte percentages ranked top 1 and 2 in H.S.S. contributing to the risk prediction, while ranked 14 and 11 in G.G.; high-sensitivity troponin I, high sensitivity CRP and D-dimer ranked top 1–3, but ranked 51, 8 and 10 in H.S.S. (Figure 1A).

Using the features above, we further trained a RFCN-Cox model outperforming existing risk prediction methods—Coxnet and DeepSurv (Katzman et al., 2018) in both internal and external validation (Figure 3 and D, Figure S1A and B in Supporting Information). Patients’ progression risk groups could be correctly separated by the predictions (Figure 3G and Figure S1B in Supporting Information). When using three features only (neutrophil percentage, lymphocyte percentage and CRP), it still showed high performance (Figure S1D in Supporting Information). More evaluation on patient subgroups were listed in Figure S1C in Supporting Information.

**LoS and cost prediction**

For LoS and cost prediction, we included RFCN-Reg, linear regression (LR), RF, multilayer perceptron (MLP), and lightGBM for comparison. The same 19 risk features for progression prediction were used here. RFCN-Reg (Figure 4A and D) predictions were significantly positively correlated to true values (spearman correlation coefficient (SCC) =0.56 for LoS and 0.62 for expense in H.S.S., Figure 4B and E) and showed correct changing trends between patient severity groups (Figure 4C and F). RFCN-Reg outperformed other models. For the LoS and the total cost, with MAE of 5.61 and 5.90 d in H.S.S. and G.G., and MAE of 7,918.68 RMB in H.S.S. (Figure S2A and B in Supporting Information). More evaluation on subgroups between methods were listed in Figure S2 in Supporting Information.

During hospitalization period, a large quantity of day-by-day records were accumulated for patients. Patients’ remaining LoS (RLoS) and cost (RCost) were meaningful to monitor patients’ hospitalization condition. We obtained the full-cycle hospitalized records from H.S.S. and G.G., totally 6,634 and 2,085 data points for 2,702 and 1,214 patients, respectively. We trained RF model to predict the RLoS and RCost, showing the performance compared to other models (Figure 3A and B, Figure S3A in Supporting Information; day MAE=5.1135, day SCC=0.6374; RMB MAE=12,965, RMB SCC=0.71). Feature importance analysis for RLoS and RCost were listed in Figure S3C and D in Supporting Information.

**Integrative analysis of disease progression, LoS and cost**

To obtain a global view on the relationship between disease progression and medical resource consumption, we clustered
all H.S.S. patients’ at-admission features as well as their predicted risk, true total LoS and cost together after removing 58 patients who finally died (Figure 5A). Patients were automatically clustered into three large clusters—high-risk high-LoS/cost (Cluster 1, 5 and 6); low-risk low-LoS/cost (Cluster 7, 9, 2 and 3); low-risk high-LoS/cost (Cluster 8, 4 and 10, Figure 5A). Patients’ risks were globally positively correlated to LoS and cost with a SCC of 0.378, however negatively correlated in special cases, e.g., of 1,158 lower-than-median-risk patients, 423 had higher-than-median cost. To discover possible reasons for the phenomenon, we fetched the low-risk high-cost (LR-hC) and low-risk low-cost (LR-lC) patient groups and compared the features between them.

Both groups showed no significant difference in the progression event, with 0.2% vs. 0% event ratios, indicating both two groups did not ever progress to critical state. 21.0%
Figure 3 COVID-19 progression risk prediction using patients’ at-admission electronic records. A, Concordance index (CI) using single features to predict progression risk in H.S.S. and G.G. hospitals respectively, as well as CI difference of each feature between H.S.S. and G.G. NA indicates a missing feature. B, Features’ linear coefficient of H.S.S. data using the 19 grid-searched optimal features. C, Net structure of RFCN-Cox for progression risk prediction. BN, batch normalization; Activation, activation function; FC, fully connection. D, Receiver operating characteristic (ROC) curve for progression risk prediction methods evaluated in G.G. external data set. AUROC, the area under ROC. E, CI of progression risk prediction methods evaluated in H.S.S. and G.G. 10-x c.v., 10-fold cross-validation. Italic bond indicates the best evaluation score. F, Kaplan-Meier plot of G.G. patients grouped by RFCN-Cox-predicted risks. High and low risks indicate risk higher or lower than the median risk of all.

Zhong, Q., et al. Sci China Life Sci May (2022) Vol.65 No.5
Figure 4 Patients’ total LoS and cost prediction using patients’ at-admission data. A, Net structure of RFCN-Reg for LoS prediction. BN, batch normalization; Activation, activation function; FC, fully connection. B, Correlation of true LoS and predicted LoS for H.S.S. hospital. C, Boxplot of true LoS and predicted LoS for patients of different disease severity levels in H.S.S. and G.G. hospitals. D, Net structure of RFCN-Reg for cost prediction. E, Correlation of true cost and predicted cost for H.S.S. hospital. F, Boxplot of true cost and predicted cost for patients of different disease severity levels in H.S.S. hospital.
Figure 5 Integrative analysis of COVID-19 patients’ progression risk, LoS and total cost. A, Clustering heat map of H.S.S. patients based on patients’ predicted risk, true LoS and total cost and all EMR features. Categories of indexes were indicated, e.g., total cost, LoS, predicted risk, risk feature, treatment cost and other features. B, Statistics comparison for features between low-risk high-cost and low-risk low-cost patients in H.S.S. Features with t-test $P$-value<0.05 were listed.
of IR-hC once reached a severe state and 11.7% for IR-lC. The expense of antivirus, Chinese medicine, immune drugs and oxygen inhalation for IR-hC was significantly higher than that for IR-lC with >105%, 96%, 95%, 94% (Figure 5B). Besides, a significantly higher proportion of IR-hC patients showed worse health condition indicators, e.g., with diabetes, coronary artery disease (CAD), shortness of breath or cough etc. In addition, alkaline phosphatase (ALP) and blood sugar were significantly higher in IR-hC than IR-lC; blood sodium, chlorine and calcium were significantly lower in IR-hC than IR-lC, although all belonged to normal ranges in both groups (Figure 5B).

A joint model to predict risk, LoS and cost and query treatment plans

Considering patients with similar health indexes are of similar illness severity, treatment plans are probably effective to patients with similar conditions. Thus it is probably practical to search existing cohort records for treatment strategies that have cured patients. Powered by ATM, we constructed a joint model that can comprehensively predict the progression risk, curve, LoS, total expense, and query treatment plans from existing records. Here we only used patients discharged that have been cured in H.S.S. as a reference database. Based on it, we constructed a web server that allows uploading queried patients’ physical testing data, searching for the most similar patients in the database based on RFCN-Cox latent space similarity, and returning the treatment records (Figure 6). The records were arranged into days of administrating each of the eight categories, and usage ratio against the total received treatments. Besides, the web server also returned predicted COVID-19 progression risk, total LoS and cost as well as their corresponding ranks against the population in the database. We took one mild and one severe patient as an example. Queried treatment plans showed that patients similar to the severe example patient administrated more types of treatment categories than patients similar to the mild one (Figure 6).

DISCUSSION

Principal findings

We developed an automatic machine learning (AutoML) framework—ATM, specially designed for hospital EMR, which can help both predict the disease progression risk on time-to-event data and for comprehensive prediction on health economics index. Significantly, this is the first research to reveal medical consumption of cured patients with COVID-19 in detail, ranging from a maximum larger than 1.5 million RMB to a minimum of less than 1,000 RMB. Depending on ATM, we trained and validated a set of deep learning models to predict COVID-19 disease progression risk, LoS and total cost, based on EMRs for 3,989 COVID-19 patients from two Chinese hospitals. The models can help patient triage at admission and dynamically monitor patients during the hospitalization, and were integrated into a user-friendly web server.

The scarcity of resources in the emergency state does not reflect the actual utilization of medical resources by each patient so that the time and economic cost for a patient’s recovery might be underestimated. In China, by February 12, 2020, more than 18,000 medical staffs had been sent to Wuhan from other parts of China to help with the coronavirus control (Li et al., 2020). Thus H.S.S. and G.G. offered treatment care under fully satisfactory medical conditions, where the patients’ information can reflect the real LoS and cost for patients to recover fully.

A comprehensive prediction model can support both disease progress and health economics research and is capable to transfer complex and multidimensions data into tangible, easy-to-read visualizations results. Existing disease predictive models showing promising AUC performance on disease risk, however have not been integrated with LoS and cost prediction (Wynants et al., 2020), which is essential to medical policy decision-making. Here, unlike other models, ATM framework provides new clinical and economic information. For healthcare management, identifying key factors to hospitalized economic burden and categorizing patients into different classes will enable efficient and appropriate financial management of medical.

Through integrative analysis, patients can be categorized more meticulously into four groups (high risk/low cost, high risk/high cost, low risk/high cost, and low risk/low cost). Like the risk progression, healthcare consumption can also be predicted. Analysis on the H.S.S. and G.G. data showed that patients with more underlying comorbidities, e.g., diabetes, heart diseases, while at the same level of risk as other patients had higher costs; some patients were at lower risk of developing the severe disease, with relatively higher costs. It indicates that the condition of these patients may have many twists and turns, while they had a better outcome, e.g., with adequate medical treatment.

We believe that more specific classification of ATM is expected to more actively explain the patients’ outcome with various dimensions, which can provide more helpful information for regional authorities and agencies to develop guidelines, protocols based on local, regional conditions. It further helps resource allocation to meet the following four fundamental values: maximizing the benefits produced by scarce resources, treating people equally, promoting and rewarding instrumental value, and giving priority to the worst off (Biddison et al., 2014; Emanuel and Wertheimer, 2006; Persad et al., 2009).

When use ATM for modelling, several aspects should be
noted. The first is that data origins should be equal to or more than two hospitals. We found that features’ prediction capability ranked differently between H.S.S. and G.G., which might come from batch effect and systematical difference in the data of the two different origins. The second is that external validation is necessary. Quite a quantity of existing studies used data from one hospital or institute to train a model, aiming to achieve good performance using as few features as possible, however possibly over-fitted the models. For example, a recent study claimed a decision tree using only three key features: LDH, lymphocyte proportion, high sensitivity C-reactive protein (hs-CRP), which could

Figure 6  A webservice to predict COVID-19 progression and query treatment plans. A, Homepage of the webserver. B, Result page of a query using EMRs. Part 1: view the uploaded EMRs; Part 2: predict patients’ risk, LoS and total cost; Part 3: query treatment suggestion.
achieve >0.9 accuracies within one hospital, but the performance dramatically worsened in data of other origins (Barish et al., 2021). In our results, although G.G. lacks features that required by H.S.S.-trained models, the model still showed consistently good performance. The third is that modelling methods should be tried as many as possible, and model parameters should be tuned well. For example, besides classic statistics, linear models, deep learning and other non-linear methods can also be tried to use the data optimally. Fourthly, since the range and distribution of values of cost and LoS are probably different among countries or regions, it is necessary to perform transfer learning for the ATM models on the new data to be analyzed that come from other countries. In addition, the ranks of cost and LoS should be still reasonable and can be referenced used if there is no enough new training data.

ATM framework can be used for patients’ hospitalized data regardless of the patients’ disease types, only requiring the data input of certain condition. ATM takes advantage of the following data for modelling: (i) at-admission stage: patients’ baseline electronic health records as features, time-to-event survival data as the learning target of risk (e.g., death or survival, progression or remission of diseases, administration of a treatment or not) and total duration and cost in hospital or other health economics data, to predict the patients’ risk levels, LoS and total cost; (ii) hospitalized stage: patients’ day-by-day electronic health records, current-day cost and LoS as features, to predict the remaining LoS and remaining cost in the future.

In this work, we totally compared ATM with five existing algorithms—Coxnet, DeepSurv, linear regression, MLP and lightGBM. Coxnet and DeepSurv were selected for performance comparison to RFCN-Cox for disease progression prediction: the former is one of the most classical statistics method for time-to-event data and Cox analysis, the latter is the new deep learning based algorithm for Cox analysis, both of which are most representative methods in the field of risk prediction. Linear regression, MLP and lightGBM were selected for comparison to RFCN-Reg and RF for total LoS, total cost, RLoS and RCost prediction. Linear regression is the most simple and machine learning way, MLP predictor is a type of deep neural network and commonly used in supervised tasks, lightGBM is a gradient boosting framework that uses tree based learning algorithm and efficient in many machine learning fields. All of the selected algorithms cover the most popular and famous methods.

CONCLUSION

Overall, an AutoML patient triage framework—ATM was developed, which comprehensively analyze not only disease progression but also LoS and cost, thus can be used to improve the reliability of medical resource distribution for hospitalized patients. Here we took COVID-19 as an example, we first reveal the medical expenses for the patients, and found that more comorbidities lead to a complex treatment plan with more money cost and longer LoS in low risk patients at admission. In addition to COVID-19, ATM also can be applied to medical resource consumption stratification of hospitalized patients for various diseases. A detailed and more comprehensive category considering both disease risk and health economics index can offer a new methodology for hospitalized patient supervision.

METHODS

Data sources
The study was conducted at Huoshenshan hospital and Guanggu hospital in Wuhan, China. The Chinese PLA General Hospital review board approved this case series as
minimal-risk research using data collected for routine clinical practice and waived the requirement for informed consent. All patients who were sufficiently medically ill require hospital admission with confirmed severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection by a positive result on high-throughput sequencing or reverse transcriptional-quantitative PCR assay with samples from the nasal or pharyngeal swab were included. Patients were admitted to hospitals between February 4, 2020, and April 15, 2020. All patients completed their hospital course at study end (discharged alive or dead). For patients with readmission during the study period, data from the first admission is presented. The progression event was defined as a composite event requiring invasive mechanical ventilation, ICU care, or death in hospital.

Data collected included patient demographic information, baseline comorbidities, signs and triage vitals (e.g., body temperature, pulse, respiration rate, blood pressure, etc.), baseline laboratory tests, diagnoses during the hospital course, inpatient treatments (including antiviruses and antibiotics), kidney replacement therapy, etc.), Charlson Comorbidity Index (CCI), and outcomes (including the length of stay, discharge, and mortality, total cost).

All laboratory tests were done in the laboratory of H.S.S. and G.G. with standard procedures. The baseline laboratory testing results were defined as the first test results within 48 h of admission. The comorbidities listed here are based on the International Classification of Diseases, 10th Revision, Clinical Modification (ICD-10-CM) codes. CCI score predicts the 10-year mortality for a patient based on age and several comorbid severe conditions. The total costs of treatment plans were collected based on the physician order presented in the inpatient-accepting treatment plans obtained from electronic health records. According to the Epidemic Prevention and Control Medical Insurance policy, the government is responsible for all treatment costs. The seventh version of the Chinese guideline grouped COVID-19 patients into mild, severe, moderate and critical levels. Cases with fever, respiratory symptoms, and radiographic evidence of pneumonia were classified as having moderate symptoms. The case were classified as having severe symptoms if they had any of the following: breathing rate 30 breaths per min or higher; oxygen saturation level 93% or lower at rest; oxygen concentration level PaO₂/FiO₂ (ratio of arterial oxygen partial pressure to fractional inspired oxygen) 300 mmHg (1 mmHg=0.133 kPa) or lower; lung infiltrates higher than 50% within 24–48 h; respiratory failure requiring mechanical ventilation; sepsis shock; or multiple organ dysfunction or failures. All other symptomatic cases were classified as mild.

**ATM framework construction**

We divided the modelling works into two aspects: (i) disease progression, total LoS and total cost prediction using patients’ baseline data when at-admission; (ii) patients’ RLoS and RCost prediction using patients’ day-by-day hospitalized data. We invented two novel deep learning networks—RFCN-Cox and RFCN-Reg for progression, total LoS and cost prediction. Besides, we selected random forest for patients’ today-and-before-today hospitalized condition, and integrated them into the ATM framework. Then we compared ATM with other popular methods, e.g., Coxnet, DeepSurv, linear regression, MLP and lightGBM for the two aspects’ tasks.

The core network structure of RFCN-Cox and RFCN-Reg is RFCN (Liu et al., 2019), which constructs skip connections between net layers to fit features so that deeper networks can be trained while avoiding gradient vanishing (He et al., 2015).

To consider patients’ risk, cost and LoS together, ATM can be firstly used to search for the optimal predictors for each interest of targets and then finally combine all the results for integrative analysis.

**Data analysis and modelling**

For the training data—H.S.S. data, features with more than 30% of missing values were discarded first. Then missing values in features were imputed by the k-nearest neighbour (KNN) algorithm. Feature values were scaled to 0–1 range feature-wise to make it easy to converge the model. For the independent test data—G.G. data, it completely missed several features that were needed for the model. To test the robustness of the model, we firstly scaled the data into 0–1 range then directly filled the completely-missing features with 0.5. For learning targets—LoS and cost were log_{10} transformed to approximate to normal distribution.

We used H.S.S. data to train the model, feature selection for COVID-19 progression prediction was carried out first by calculating the CI of a single feature for the prediction using Coxnet. Coxnet is a Cox regression regularized by the net elastic penalty for modelling. Then grid search was used to find the optimal feature combinations for COVID-19 progression prediction. DeepSurv utilizes MLP to fit features into the final target, and RFCN-Cox takes advantage of the RFCN network for fitting. Parameters of both methods are updated by loss calculated through the negative log-likelihood (NLL) function. We used 10-fold cross-validation to test the model performance in H.S.S. data. Using the 10-fold data sets, ten models were trained for Coxnet, DeepSurv and RFCN-Cox, respectively, then we obtained a unified model for Coxnet, DeepSurv and RFCN-Cox by averaging the ten models’ predictions respectively.

For net structure optimization, we scanned a different number of layers and neuron numbers for DeepSurv, and a different number of residual blocks and the latent number of each layer for RFCN-Cox.
After the model training was finished, we tested the model performance in the G.G. data of an independent hospital. Feature importance analysis for deep learning models and RF were performed through leave-one-out (LOO) strategy, filling all values of the feature being analyzed with 0.5 and calculating the difference in performance evaluation scores (e.g., SCC and MAE) before and after the operation. The higher difference indicates a stronger influence on the prediction power.

Python package `sklearn` was used to preprocess data sources, calculate AUROC, and train LR, RF and MLP models. Python package `sksurv` was used to calculate CI, perform the log-rank test and grid search of features, and generate Kaplan-Meier plots. Python package `torch` was used to train DeepSurv, RFCN-Cox and RFCN-Reg models.

### Clustering analysis

We obtained the ranks of all at-admission features, treatment costs, predicted risks, LoS and cost for clustering to avoid cluster distance bias caused by the magnitude difference. R package `pheatmap` was used to generate heat maps with pearson correlation coefficient (PCC) as distance measurement for hierarchical clustering. The t-test was performed using raw values of each feature between groups and implemented in R.

### Webserver

The web server was constructed by python library `Flask`. User uploaded patients’ features were firstly normalized to 0–1 range using H.S.S. scaler, and then input to the COVID-19 progression risk, LoS and cost prediction models. Cosine similarity was used to measure the distance between user-uploaded patients to H.S.S. patients.

### Compliance and ethics

The author(s) declare that they have no conflict of interest.

### Acknowledgements

This work was supported by the Special Zone for National Defense Innovation of CMC Science and Technology Project (19-163-15-LZ-001-001-01).

### References

Barish, M., Bolourani, S., Lau, L.F., Shah, S., and Zanos, T.P. (2021). External validation demonstrates limited clinical utility of the interpretable mortality prediction model for patients with COVID-19. *Nat Mach Intell* 3, 25–27.

Biddison, L.D., Berkowitz, K.A., Courtney, B., De Jong, C.M.J., Devereaux, A.V., Kissoon, N., Roxland, B.E., Sprung, C.L., Dichter, J.R., Christian, M.D., et al. (2014). Ethical considerations: care of the critically ill and injured during pandemics and disasters: CHEST consensus statement. *Chest* 146, e145S–e155S.

Curiati, P.K., Gil-Junior, L.A., Morinaga, C.V., Ganem, F., Curati, J.A.E., and Avelino-Silva, T.J. (2020). Predicting hospital admission and prolonged length of stay in older adults in the emergency department: The PRO-AGE Scoring System. *Ann Emerg Med* 76, 255–265.

Emanuel, E.J., and Wertheimer, A. (2006). Who should get influenza vaccine when not all can? *Science* 312, 854–855.

Gatesman, M.I., and Smith, T.J. (2011). The shortage of essential chemotherapy drugs in the United States. *N Engl J Med* 365, 1653–1655.

He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep residual learning for image recognition. ArXiv 151203385.

Hyde, R. (2020). Europe faces worsening medicine shortages. *Lancet* 395, 481–482.

Katzman, J.L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., and Kluger, Y. (2018). DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Med Res Methodol* 18, 24.

Lee, D.S., and Ezekowitz, J.A. (2014). Risk stratification in acute heart failure. *Can J Cardiol* 30, 312–319.

Li, R., Rivers, C., Tan, Q., Murray, M.B., Toner, E., and Lipsitch, M. (2020). Estimated demand for US hospital inpatient and intensive care unit beds for patients with COVID-19 based on comparisons with Wuhan and Guangzhou, China. *JAMA Netw Open* 3, e208297.

Liu, D., Xu, C., He, W., Xu, Z., Fu, W., Zhang, L., Yang, J., Peng, G., Han, D., Bai, X., et al. (2019). AutoGenome: an AutoML tool for genomic research. BioRxiv 842526.

McMahon, D.E., Peters, G.A., Ivers, L.C., and Freeman, E.E. (2020). Global resource shortages during COVID-19: bad news for low-income countries. *PLoS Negl Trop Dis* 14, e0008412.

Persad, G., Wertheimer, A., and Emanuel, E.J. (2009). Principles for allocation of scarce medical interventions. *Lancet* 373, 423–431.

Wynants, L., Van Calster, B., Collins, G.S., Riley, R.D., Heinze, G., Schuit, E., Bonten, M.J., Dahlby, D.L., Damen, J.A.A., Debray, T.P.A., et al. (2020). Prediction models for diagnosis and prognosis of COVID-19: systematic review and critical appraisal. *BMJ* 369, m1328.

**SUPPORTING INFORMATION**

The supporting information is available online at https://doi.org/10.1007/s11427-021-1987-5. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.