Severity detection for the coronavirus disease 2019 (COVID-19) patients using a machine learning model based on the blood and urine tests

Running title：Severity detection for COVID-19

Haochen Yao[2,†], Nan Zhang[1,†], Ruochi Zhang[3,†], Meiyu Duan[3], Tianqi Xie[4], Jiahui Pan[2], Ejun Peng[5], Juanjuan Huang[2], Yingli Zhang[1], Xiaoming Xu[1], Hong Xu[1,*], Fengfeng Zhou[3,*], Guoqing Wang[2,*]

[1]The First Hospital of Jilin University, Jilin University, Changchun, China

[2]Department of Pathogenobiology, The Key Laboratory of Zoonosis, Chinese Ministry of Education, College of Basic Medical Science, Jilin University, Changchun, China

[3]BioKnow Health Informatics Lab, College of Software, and Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun 130012, Jilin, China

[4]School of Computing and Information, University of Pittsburgh, 135 N Bellefield Ave, Pittsburgh, PA 15213, United States

[5]Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

*Correspondence:

Guoqing Wang

qing@jlu.edu.cn;

Fengfeng Zhou

ffzhou@jlu.edu.cn;

Hong Xu

chxuhong@163.com

† These authors have contributed equally to this work.

number of words: 3654

number of figures:5

number of tables:1

ABSTRACT

The recent outbreak of the coronavirus disease-2019 (COVID-19) caused serious challenges to the human society in China and across the world. COVID-19 induced pneumonia in human hosts and carried a highly inter-person contagiousness. The COVID-19 patients may carry severe symptoms, and some of them may even die of major organ failures. This study utilized the machine learning algorithms to build the COVID-19 severeness detection model. Support vector machine (SVM) demonstrated a promising detection accuracy after 32 features were detected to be significantly associated with the COVID-19 severeness. These 32 features were further screened for inter-feature redundancies. The final SVM model was trained using 28 features and achieved the overall accuracy 0.8148. This work may facilitate the risk estimation of whether the COVID-19 patients would develop the severe symptoms. The 28 COVID-19 severeness associated biomarkers may also be investigated for their underlining mechanisms how they were involved in the COVID-19 infections.

Keywords: Severity detection, COVID-19, model, blood and urine tests, biomarkers

INTRODUCTION

Multiple cases of pneumonia patients were linked to the coronavirus disease-2019 (COVID-19) occurred in December 2019 [1]. The virus 2019-nCoV demonstrated a substantial capability of inter-human transmissions [2] and has rapidly spread around the world, in particular South Korea and Japan [3]. Patients infected with COVID-19 had significantly varied symptoms and their outcomes ranged from mild to death, and the mortality rate was approximately 4.3% [4]. It is necessary to mention that 61.5% of the COVID-19 pneumonia patients with critical symptoms died within 28 days after admission [5]. The discrimination of severely ill patients with COVID-19 from those with mild symptoms may help understand the individualized variations of the COVID-19 prognosis. The knowledge may also facilitate the establishing of early diagnosis of the COVID-19 severeness.

The diagnosis of COVID-19 heavily relies on the epidemiological features, clinical characteristics, imaging findings, and nucleic acid screening [6], etc. The delivery of the diagnosis result by these technologies was time consuming and error prone [7]. Multiple types of clinical data were collected for a patient with COVID-19 infection and they were manually integrated by the clinicians to make the diagnosis decisions. The stochastic transmission model was also used to investigate how the COVID-19 transmitted locally and globally [8]. Machine learning algorithms were widely used to integrate the heterogeneous biomedical data sources for the diagnosis decision [9,10]. So they may also be utilized to produce more delicate prediction models for the severeness diagnosis of the COVID-19 patients. The biomarkers used for an

accurate diagnosis model of patients with COVID-19 may serve as the drug targets for this global infectious disease.

This study investigated the detection of severely ill patients with COVID-19 from those with mild symptoms using the clinical information and the blood/urine test data. The clinical information consisted of age, sex, body temperature, heart rate, respiratory rate and blood pressure. The blood/urine tests may be carried out using the technically-easy and cost-efficient procedures. An accurate severeness detection model of the patients with COVID-19 based on those features above may improve the prognosis of this disease in large scale clinical practices. The following sections will firstly describe the data collection and modeling methods, and then utilized the popular machine learning algorithms to build the best severeness detection model.

MATERIALS AND METHODS

Data collection

This study recruited 137 clinically confirmed cases of COVID-19, which were collected from the Tongji Hospital Affiliated to Huazhong University of Science and Technology. Patients were hospitalized from January 18, 2020, to February 13, 2020. The cohort consisted of 17 mild cases, 45 moderate ones and 75 severely ill patients. 21 of the severe cases eventually died. This study investigated the binary classification problem between 75 severe/deceased cases and 62 mild/moderate ones. Each participant was regarded as a sample in this study. This study was approved by the Ethics Commission of the First Hospital of Jilin University(2020-236). With informed consent was waived for this emerging infectious disease.

Patient information including age, sex, body temperature, heart rate, respiratory rate, blood pressure and the blood/urine tests data. Each clinically-obtained value was regarded as a feature in this study. In summary, each sample has 100 features, consisting of 8 clinical, 76 blood test and 16 urine test values.

Data pre-processing

The missing entries were filled in the following procedure. We assumed a missing entry to be within the normal range and filled this entry with the median of that normal range. If there is no normal range for a missing entry, we filled it with zero (0). The samples were randomly split into 80% as training and 20% as test datasets in a stratified fashion. Features in continuous values were normalized by the values in the training dataset. The categorial features were encoded by the one-hot strategy.

Feature selection

The principle of Occam's razor suggested that a model using fewer features was preferred over a complicated model with a similar prediction performance [11]. Feature selection algorithms may

be utilized to remove those unrelated features [12] and may usually increase the model prediction performances [13,14].

The student t-test (abbreviated as T-test) is a filter algorithm and it evaluates the statistical association of each feature with the disease severeness of a sample. The features with the T-test calculated Pvalues below 0.05 were usually considered to be statistically significantly associated with the disease severeness [15,16].

Prediction algorithms

This study evaluated several classification algorithms to build the prediction models of the severely ill patients with COVID-19. The predictive logistic regression (LR) model is a regression analysis for the dataset with the binary dependent variable, i.e., the class label [17]. LR has been widely used to build clinical decision models [18,19]. A probability is calculated by LR to describe whether the sample belongs to a class and a threshold for the probability is usually utilized to make the predictive decision. LR firstly calculates the log-odds $l=\log_b[p/(1-p)]=\beta_0+\beta_1x+\ldots+\beta_nx$, and the probability $p=1/[1+b^{-(\beta0+\beta1x+\ldots+\beta nx)}]$, where β_i is the model parameter.

Support vector machine (SVM) is a supervised machine learning algorithm that may accomplish both classification and regression tasks [20]. SVM tries to find a hyperplane to separate data by the highest margin. The learning strategy of SVM is spacing maximization, which can be formalized as a problem of solving convex quadratic programming [21]. This algorithm has been widely used to build the prediction models using the data of blood test [22,23] and urine test [24,25].

Random forest (RF) is an ensemble algorithm that summarizes the prediction results of multiple tree-based classifiers [26]. RF may improve the model performances and avoid over-fitting by averaging the results of models trained over various sub-samples of the dataset. Its model complexity renders itself computation-intensive and RF runs slower than many prediction algorithms. RF is another popular algorithm for building the prediction models using the clinical data [27,28].

K nearest neighbor (KNN) is an instance-based learning algorithm and summarizes the prediction based on the class labels of the query sample's k nearest neighbors [29]. KNN simply assigns the query sample with the class label of its majority nearest neighbors. And its prediction performance heavily relies on the definition of the inter-sample distances. Nicholas Schaub, et al., demonstrated that selecting the best biomarkers may be essential to improve the KNN models [30].

The boosting-based algorithm AdaBoost iteratively trained weak learners and summarized these weak learners' results into a weighted sum [31]. Multiple variants of Adaboost were proposed for recognizing human actions [32], diagnosing the dog hypoadrenocorticism [33], and predicting protein binding sites [34], etc.

The above algorithms are implemented using Python programming language (version 3.6) and Scikit-learn package (version 0.22).

Prediction performance evaluation metrics

The binary classification model was evaluated using four classification performance metrics, as defined in the followings. The severely ill patients were regarded as positive samples and the other patients constituted the negative dataset. The number of correctly predicted positive samples was defined as true positive (TP), and the number of the other positive samples was false negative (FN). The true negative (TN) and the false positive (FP) were defined as the numbers of correctly and incorrectly predicted negative samples, respectively. So the overall accuracy Acc was defined as Acc=(TP+TN)/(TP+FN+TN+FP). The model's sensitivity (Sn) and specificity (Sp) were defined as Sn=TP/(TP+FN) and Sp=TN/(TN+FP). The three metrics Acc, Sn and Sp measured the percentages of correctly predicted all, positive and negative samples, respectively. The Matthew's Correlation Coefficient (MCC) described the overall correlation of the predicted and the real class labels, and MCC was defined as MCC=(TP×TN-TP×FN)/sqrt[(TP+FP) ×(TP+FN) ×(TN+FP) ×(TN+FN)], where sqrt() was the square root function [35,36].

Each model was randomly trained for twenty runs with different random seeds and the metric averaged accuracy aAcc=[Acc(1)+Acc(2)+...+Acc(20)]/20, where Acc(i) was the accuracy of the i^{th} model. The metric aAcc was used to find the best prediction model. The metrics aSn, aSp and aMCC were the averaged Sn, averaged Sp and averaged MCC over the twenty random runs.

Ethics statement

This study was approved by the Ethics Commission of the First Hospital of Jilin University(2020-236). With informed consent was waived for this emerging infectious disease.

RESULTS

Baseline characteristics of the 2019-nCoV pneumonia participants

This study recruited 137 COVID-19 patients to build the detection model of severely ill (positive) samples against the patients with mild symptoms. All the 100 features were screened for their association with the class label, i.e., Positive or Negative. There were 8 clinical values, 76 blood test values and 16 urine test values, respectively. Thirty-two features achieved the T-test Pvalue<0.05, and were kept for further analysis in the following sections，as summarized in the Supplementary Table S1.

The feature of the patient's age at diagnosis (Age) demonstrated a significant difference (Pvalue=1.75e-6) between the two groups of samples, and the severely ill patients were on average 13.5695 years older than the patients with mild symptoms. This supported the

observation that patients aged around 65 years old tended to have more severe symptoms than those aged around 51 years old [37]. The sex also demonstrated severe-specific Pvalue=7.71e-5, suggesting that male patients were at higher risks of developing severe symptoms [38], as shown in Figure 1 (A).

We also summarized three blood test values and three urine test values with the most significant differences between the two groups of samples, as shown in Figure 1 (A). Overall, the blood test values demonstrated much more significant inter-group differences than the urine test values. The summary data suggested that the percentage of neutrophil cells was significantly enriched in the blood of the severely ill patients, with P values 4.14e-11. In addition, the serum calcium level and the monocyte percentage were also significantly lower in the severely ill patients than those mild ones.

Three urine test values demonstrated weak inter-group differential significances. The two values “Urine | Urine protein” and “Urine | Red blood cell(occult)” demonstrated the elevated levels in the severely ill patients with Pvalues 1.44e-2 and 2.83e-2, respectively. But their variations were very larger, which rendered neither of them as good disease severeness biomarkers. A minor decrease (0.1028) in the urine pH value (feature “Urine | PH(Urine)”) in the severely ill patients achieved the inter-group differential significance Pvalue 4.25e-2.

In the following sections. The detailed summary may be found in the Supplementary Table S1.

Evaluation of feature correlations with the group labels

We firstly evaluated the correlation between the 32 features and the class label using Pearson Correlation Coefficient (PCC), as shown in Figure 1 (B). The PCC value ranges between -1 and 1. This study focused on the whether a feature was correlated with the class label. So the absolute value of PCC was calculated in Figure 1 (B).

Some features showed strong correlations with the 2019-nCoV pneumonia severeness, which was the class label. The feature “Blood | Neutrophil percentage” demonstrated the largest PCC=0.53 with the disease severeness (class label). This provided another piece of evidence that the neutrophil cell percentage was positively correlated with the 2019-nCoV severeness. Another feature “Blood | Calcium” achieved the second-best PCC=0.49. The age at diagnosis (feature Age) achieved the third-best PCC=0.40 with the class label, suggesting that the elder patients were under higher risks of developing severe symptoms.

Figure 1 (B) suggested that some of the 32 features were highly correlated with the class label and they may facilitate the training of a reasonably-accurate detection model for the 2019-nCoV pneumonia severeness. The existence of high inter-feature correlations suggested that some redundant features may need to be removed to further improve the detection model.

Comparison of different prediction algorithms

Five prediction algorithms were evaluated for their detection performances using their default parameters on all the 98 features of the 2019-nCoV pneumonia patients, as shown in Figure 2. Firstly, all the five prediction algorithms achieved at least 0.7130 in Acc on all the 32 features, suggesting that the severely ill COVID-19 patients may have severeness-specific patterns. The prediction algorithm SVM achieved the best prediction accuracy Acc=0.7926 and its standard deviation in Acc was only 0.0715. SVM achieved the sensitivity Sn=0.7666 much better than the specificity Sp=0.6993. The prediction sensitivity was the detection accuracy of the positive samples, i.e., the severely ill patients. So the following sections used the prediction algorithm SVM as the default predictor and the prediction model was further refined by optimizing the SVM parameters and selecting the best features.

Choosing the best threshold

A threshold may be tuned to find the balanced model performances for both positive and negative samples, as shown in Figure 3. The metric Youden's index was introduced by W.J. Youden in 1950 to catch the best performance of a dichotomous diagnostic model [39]. Youden's index assigns equal weights for sensitivity and specificity and tries to maximize the index value J =(Sn+Sp-1) [39]. Figure 3 illustrated the changing curves of Sn and Sp with different thresholds for the prediction scores of the samples. The maximal value of J was achieved at the threshold 0.7318, and averaged accuracy of SVM was improved to 0.8148.

The Youden's index was used to find the best threshold of the SVM models with different parameters and features in the following sections.

Tuning the parameters of the SVM model

The grid search strategy was carried out to evaluate how different parameter values affected the disease severeness detection model, as shown in Figure 4. Parameter tuning was a time-consuming step. So this section randomly split the training dataset into 80% sub-training dataset and 20% validation dataset. Each model was trained using the sub-training dataset and the performance was calculated on the validation dataset. The model detection performance didn't change with the linear kernel and different choices of the parameter Gamma, as shown in Figure 4 (b). And the best accuracy=0.8636 of the linear kernel SVM was achieved when C=0.1 or 1. The best model with the RBF kernel achieved Acc=0.9091 for the validation dataset, where C=100 and Gamma=0.0010. The other three metrics Sp=1.0000, Sn=0.8333 and MCC=0.8333 were also the best values in Figure 4. The SVM model with the above-mentioned parameters achieved Acc=0.8148 on the independent test dataset. So the following sections used these two choices of the parameters C and Gamma.

Remove redundant features to improve the model

The existence of strong inter-feature correlations in Figure 1 (B) suggested that some features may be removed to further improve the model. This section carried out a conservative recursive feature elimination (cRFE) strategy to eliminate the redundant features while ensuring the model performance was not decreased. The model performance was evaluated for its threshold-independent metric AUC value [40,41]. Firstly, all the 32 features were ranked by the ascending order of their T-test Pvalues. Then, the detection model was evaluated by eliminating each feature. A feature was eliminated if the model's AUC was improved with its removal. Otherwise that feature was kept. The final feature set was returned after all the features were evaluated.

The feature selection procedure should avoid using the test samples, so this section calculated the performance metrics on the validation dataset using the model trained over the sub-training dataset, as shown in Figure 5. The heuristic cRFE strategy ensured by its nature that the model performance would not be decreased, and the rising line segment indicated the removal of the feature on the horizontal axis. Four features were removed, i.e., "Age", "Blood | Interleukin-10", "Blood | Prothrombin time,", and "Blood | Oxygen partial pressure". Figure 1 (B) illustrated that all these four features were strongly correlated with some other features, with the PCC values at least 0.51.

The remaining 28 features achieved Acc=0.9917 on the validation dataset, and Acc=0.8148 on the independent test dataset. Although the COVID-19 severeness detection performance was not improved, the model complexity was reduced and the clinical screening cost was reduced with fewer features.

A web site was established to help the clinicians to try this COVID-19 infection severity estimation model, and the users may access: http://dVirusSeverity.HealthInformaticsLab.org/ .

DISCUSSION

The emergence of SARS-CoV-2 marked the third of highly pathogenic coronavirus in humans in the twenty-first century, after severe acute respiratory syndrome (SARS) in 2003, and Middle East respiratory syndrome (MERS) in 2012 [42,43]. SARS-CoV-2 belongs to the coronavirus family, β-coronavirus genera and belongs to the cluster of betacoronaviruses [44]. Based on Sequence analysis, the amino acid sequences of SARS-CoV-2 showed 94.4% identity with SARS-CoV [45]. It is suggested that SARS-CoV-2 was more closely related to SARS-like bat CoV. In comparison, SARS-CoV-2 was more distant from the MERS-CoV [46,47]. The mortality of critically ill patients with COVID-19 is considerable. The survival time of the dead patients may be within 1-2 weeks after ICU admission [48].

The present diagnosis of COVID-19 didn't achieve a satisfying accuracy. Both false positives and false negatives need to be decreased [49-51]. The clinical decisions of COVID-19 infections are usually confirmed by epidemiological features, clinical manifestations, imaging factors, and

nucleic acid screenings, etc. Some of the COVID-19 patients may develop severe symptoms and these patients are at a much higher mortality rate than the other patients. This challenge raised the scientific question of finding the COVID-19 severeness specific biomarkers, which may help reduce the overall mortality.

This study investigated the binary classification problem between 75 severely illed COVID-19 infected patients and the other 62 patients with mild symptoms. A comprehensive optimization procedure led to the best SVM-based COVID-19 severeness detection model using only 28 features. The experimental data suggested that the severely illed patients had a higher serum level of neutrophil percentage and lower serum levels of monocyte percentage and calcium compared with those mild ones. Urine test contributed three weak group-specific biomarkers, i.e., urine pH value, urine protein and urine red blood cell. Compared with the urine pH value, the variations of urine protein and urine red blood cell were very large and these two urine features may not serve well as COVID-19 infection severeness biomarkers. The blood test features demonstrated much more significant inter-group differences than the urine test features. The summary data suggested these three blood test features as candidate severeness biomarkers, i.e., serum ferritin, hs-CRP, interleukin-2R, and tumor necrosis factor-α.

COVID-19 severeness detection model achieved the overall accuracy 0.8148 on the independent test dataset with only 28 clinical biomarkers. Twenty-one out of these 28 biomarkers were investigated in the coronavirus. Two serum values “Blood | Tumor necrosis factor-α” (56 papers) and “Blood | Sodium” (57 papers) were known to be associated with the coronavirus infections. The tumor necrosis factor-alpha (TNF-alpha) was observed to have elevated expression levels in the serum of the coronavirus-infected mice [52]. The serum sodium level was slightly increased by 2.01% in the severely ill patients in the cohort used in this study. Hoffman, et al., proposed that the pulmonary complication were more frequently observed in the hypernatremia patients [53]. So it would be interesting to investigate the underlining mechanism of how the serum sodium may induce the COVID-19 severeness. The feature “Urine | PH(Urine)” is the pH level in the urine, and quite a few investigations observed the aberrant pH levels in the body fluid or fecal matter of the coronavirus-infected animals [54,55]. Although the urine pH level was not investigated in the coronavirus-infected animals, this may be worth of an investigation. The sex bias was also observed that coronavirus tended to infect males [56,57]. Our data suggested that males were at a higher risk to be infected by COVID-19 and to develop more severe symptoms.

An accurate severeness detection model of the patients with COVID-19 based on those features may improve the prognosis of this disease in large scale clinical practices, and reduce the incidence of COVID-19 severeness and mortality. The biomarkers used for an accurate diagnosis model of patients with COVID-19 may serve as the drug targets for this global infectious disease.

There are some limitations that should be noted. First, the number of patients with COVID-19 is relatively small, which may limit the accuracy of severeness detection model. Second, since all subjects in our study were Chinese patients with COVID-19, the results may not be applied to other ethnicities. Third, the data of this study is only the preliminary establishment of COVID-19 severeness detection model. Further studies are still needed.

This study utilized the machine learning algorithms to detect the COVID-19 severely ill patients from those with only mild symptoms. Our experimental data demonstrated strong correlations with the COVID-19 severeness. And the final COVID-19 severeness detection model achieved the accuracy 0.8148 on the independent test dataset using only 28 clinical biomarkers. The detection model itself is in urgent need for the current epidemic situation that the severely ill patients are at a very high mortality rate. The 28 biomarkers may also be investigated for their underlining mechanisms of their roles in the COVID-19 severely ill patients.

AUTHOR CONTRIBUTIONS

NZ collected the data. HY, and RZ analyzed the data and wrote the paper. MD, TX, JP, EP, JH, YZ, and XX did literature search. HX, FZ and GW conceived and designed this study. The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report of this study. The corresponding author had full access to all the study data and had final responsibility for the decision to submit for publication. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We thank all patients.

FUNDING

This work was supported by grants from The epidemiology, early warning and response techniques of major infectious diseases in the Belt and Road Initiative (#2018ZX10101002), National Natural Science Foundation of China (#81871699), Jilin Provincial Key Laboratory of Big Data Intelligent Computing (20180622002JC), the Education Department of Jilin Province (JJKH20180145KJ), Foundation of Jilin Province Science and Technology Department (#172408GH010234983), and the startup grant of the Jilin University. This work was also partially supported by the Bioknow MedAI Institute (BMCPP-2018-001), the High Performance

Computing Center of Jilin University, and the Fundamental Research Funds for the Central Universities, JLU.

COMPETING INTERESTS

The authors have declared no competing interests.

REFERENCES

[1] Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med 2020;382:727-33. doi: 10.1056/NEJMoa2001017

[2] Chan JF, Yuan S, Kok KH, To KK, Chu H, Yang J, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet 2020;395:514-23. doi:10.1016/S0140-6736(20)30154-9

[3] Li Q, Guan X, Wu P, Li Q, Guan X, Wu P, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. N Engl J Med 2020. doi:10.1056/NEJMoa2001316

[4] Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. JAMA 2020. doi:10.1001/jama.2020.1585

[5] Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. Lancet Respir Med 2020. doi:10.1016/S2213-2600(20)30079-5

[6] Shi H, Han X, Jiang N, Cao Y, Alwalid O, Gu J, et al. Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. Lancet Infect Dis 2020. doi:10.1016/S1473-3099(20)30086-4

[7] Xie X, Zhong Z, Zhao W, Zheng C, Wang F, Liu J. Chest CT for typical 2019-nCoV pneumonia: relationship to negative RT-PCR testing. Radiology 2020:200343. doi:10.1148/radiol.2020200343

[8] Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. Lancet Infect Dis 2020. doi:10.1016/S1473-3099(20)30144-4

[9] Hu F, Zeng W, Liu X. A Gene Signature of Survival Prediction for Kidney Renal Cell Carcinoma by Multi-Omic Data Analysis. Int J Mol Sci 2019;20. doi:10.3390/ijms20225720

[10] Thompson JA, Christensen BC, Marsit CJ. Methylation-to-Expression Feature Models of Breast Cancer Accurately Predict Overall Survival, Distant-Recurrence Free Survival, and Pathologic Complete Response in Multiple Cohorts. Sci Rep 2018;8:5190. doi:10.1038/s41598-018-23494-0

[11] Koller D, Sahami M. Toward optimal feature selection. Stanford InfoLab;1996.

[12] Ebrahimpour MK, Zare M, Eftekhari M, Aghamolaei G. Occam's razor in dimension reduction: Using reduced row Echelon form for finding linear independent features in high dimensional microarray datasets. Engineering Applications of Artificial Intelligence. 2017;62:214-221. doi:

[13] Yang S, Li B, Zhang Y, Duan M, Liu S, Zhang Y, et al. Selection of features for patient-independent detection of seizure events using scalp EEG signals. Computers in Biology and Medicine. 2020:103671. doi:10.1016/j.compbiomed.2020.103671

[14] Zhang Y, Chen C, Duan M, Liu S, Huang L, Zhou F. BioDog, biomarker detection for improving identification power of breast cancer histologic grade in methylomics. Epigenomics. 2019;11(15):1717-1732. doi:10.2217/epi-2019-0230

[15] Govindan RB, Massaro A, Vezina G, Chang T, du Plessis A. Identifying an optimal epoch length for spectral analysis of heart rate of critically-ill infants. Comput Biol Med. 2019;113:103391. doi:10.1016/j.compbiomed.2019.103391

[16] Peng T, Trew ML, Malik A. Predictive modeling of drug effects on electrocardiograms. Comput Biol Med. 2019;108:332-344. doi:10.1016/j.compbiomed.2019.03.027

[17] Kleinbaum DG, Dietz K, Gail M, Klein M, Klein M. Logistic regression. Springer; 2002.

[18] Heijnen BJ, Bohringer S, Speyer R. Prediction of aspiration in dysphagia using logistic regression: oral intake and self-evaluation. Eur Arch Otorhinolaryngol. 2020;277(1):197-205. doi:10.1007/s00405-019-05687-z

[19] Luo CL, Rong Y, Chen H, Zhang W, Wu L, Wei D, et al. A Logistic Regression Model for Noninvasive Prediction of AFP-Negative Hepatocellular Carcinoma. Technol Cancer Res Treat. 2019;18:1533033819846632. doi:10.1177/1533033819846632

[20] Suykens JA, Vandewalle J. Least squares support vector machine classifiers. Neural processing letters. 1999;9(3):293-300. doi:10.1023/A:1018628609742

[21] Shilton A, Palaniswami M, Ralph D, Tsoi AC. Incremental training of support vector machines. IEEE transactions on neural networks. 2005;16(1):114-131. doi:10.1109/TNN.2004.836201

[22] Li C, Hou L, Sharma BY, Li H, Chen C, Li Y, et al. Developing a new intelligent system for the diagnosis of tuberculous pleural effusion. Comput Methods Programs Biomed. 2018;153:211-225. doi:10.1016/j.cmpb.2017.10.022

[23] Li D, Feng S, Huang H, Chen W, Shi H, Liu N, et al. Label-free detection of blood plasma using silver nanoparticle based surface-enhanced Raman spectroscopy for esophageal cancer screening. J Biomed Nanotechnol. 2014;10(3):478-484. doi:10.1166/jbn.2014.1750

[24] Osredkar J, Gosar D, Macek J, Kumer K, Fabjan T, Finderle P, et al. Urinary Markers of Oxidative Stress in Children with Autism Spectrum Disorder (ASD). Antioxidants (Basel). 2019;8(6). doi:10.3390/antiox8060187

[25] Zhou H, Li L, Zhao H, Wang Y, Du J, Zhang P, et al. A Large-Scale, Multi-Center Urine Biomarkers Identification of Coronary Heart Disease in TCM Syndrome Differentiation. J Proteome Res. 2019;18(5):1994-2003. doi:10.1021/acs.jproteome.8b00799

[26] Pal M. Random forest classifier for remote sensing classification. International Journal of Remote Sensing. 2005;26(1):217-222.

[27] Wu J, Bai J, Wang W, Xi L, Zhang P, Lan J, et al. ATBdiscrimination: An in Silico Tool for Identification of Active Tuberculosis Disease Based on Routine Blood Test and T-SPOT.TB Detection Results. J Chem Inf Model. 2019;59(11):4561-4568. doi:10.1021/acs.jcim.9b00678

[28] Zhang C, Leng W, Sun C, Lu T, Chen Z, Men X, et al. Urine Proteome Profiling Predicts Lung Cancer from Control Cases and Other Tumors. EBioMedicine. 2018;30:120-128. doi:10.1016/j.ebiom.2018.03.009

[29] Dudani SA. The distance-weighted k-nearest-neighbor rule. IEEE Transactions on Systems, Man, and Cybernetics. 1976(4):325-327.

[30] Schaub NP, Jones KJ, Nyalwidhe JO, Cazares LH, Karbassi ID, Semmes OJ, et al. Serum proteomic biomarker discovery reflective of stage and obesity in breast cancer patients. J Am Coll Surg. 2009;208(5):970-978; discussion 978-980. doi:10.1016/j.jamcollsurg.2008.12.024

[31] Ratsch G, Onoda T, Muller K-R. Soft margins for AdaBoost. Machine learning. 2001;42(3):287-320.

[32] Lv F, Nevatia R. Recognition and segmentation of 3-d human action using hmm and multi-class adaboost. Paper presented at: European conference on computer vision2006.

[33] Reagan KL, Reagan BA, Gilor C. Machine learning algorithm as a diagnostic tool for hypoadrenocorticism in dogs. Domest Anim Endocrinol. 2019;72:106396. doi:10.1016/j.domaniend.2019.106396

[34] Qiao L, Xie D. MIonSite: Ligand-specific prediction of metal ion-binding sites via enhanced AdaBoost algorithm with protein sequence information. Anal Biochem. 2019;566:75-88. doi:10.1016/j.ab.2018.11.009

[35] Cogan T, Cogan M, Tamil L. MAPGI: Accurate identification of anatomical landmarks and diseased tissue in gastrointestinal tract using deep learning. Comput Biol Med. 2019;111:103351. doi:10.1016/j.compbiomed.2019.103351

[36] Khurana S, Rawi R, Kunji K, Chuang GY, Bensmail H, Mall R. DeepSol: a deep learning framework for sequence-based protein solubility prediction. Bioinformatics. 2018;34(15):2605-2613. doi:10.1093/bioinformatics/bty166

[37] Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet. 2020;395(10223):507□513. doi:10.1016/S0140-6736(20)30211-7

[38] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet 2020;395:497-506. doi:10.1016/0010-7824(81)90078-0

[39] Youden WJ. Index for rating diagnostic tests. Cancer 1950;3:32-5. doi:10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3

[40] Kourou K, Rigas G, Papaloukas C, Mitsis M, Fotiadis DI. Cancer classification from time series microarray data through regulatory Dynamic Bayesian Networks. Comput Biol Med 2020;116:103577. doi: 10.1016/j.compbiomed.2019.103577

[41] Deepak S, Ameer PM. Brain tumor classification using deep CNN features via transfer learning. Comput Biol Med 2019;111:103345. doi:10.1016/j.compbiomed.2019.103345

[42] BDrosten, C. et al. Identification of a novel coronavirus in patients with severe acute

respiratory syndrome. N. Engl. J. Med. 348, 1967–1976 (2003).

[43] Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. M. E. & Fouchier, R. A. M. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N. Engl. J. Med. 367, 1814–1820 (2012)

[44] Chen Y, Liu Q, Guo D. Emerging coronaviruses: Genome structure, replication, and pathogenesis. J Med Virol. 2020;92(4):418–423. doi:10.1002/jmv.25681.

[45] A pneumonia outbreak associated with a new coronavirus of probable bat origin.

[46] Lu R, Zhao X, Li J, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet. 2020;395(10224):565–574. doi:10.1016/S0140-6736(20)30251-8.

[47] Wu et al., Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China, Cell Host & Microbe (2020), https://doi.org/10.1016/j.chom.2020.02.001.

[48] Yang X, Yu Y, Xu J, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study [published correction appears in Lancet Respir Med. 2020 Apr;8(4):e26]. Lancet Respir Med. 2020;8(5):475-481. doi:10.1016/S2213-2600(20)30079-5

[49] Yan G, Lee CK, Lam LTM, Yan B, Chua YX, Lim AYN, et al. Covert COVID-19 and false-positive dengue serology in Singapore. Lancet Infect Dis 2020. doi:10.1016/S1473-3099(20)30158-4

[50] Li D, Wang D, Dong J, Wang N, Huang H, Xu H, et al. False-Negative Results of Real-Time Reverse-Transcriptase Polymerase Chain Reaction for Severe Acute Respiratory Syndrome Coronavirus 2: Role of Deep-Learning-Based CT Diagnosis and Insights from Two Cases. Korean J Radiol 2020. doi:10.3348/kjr.2020.0146

[51] Li Z, Yi Y, Luo X, Xiong N, Liu Y, Li S, et al. Development and Clinical Application of A Rapid IgM-IgG Combined Antibody Test for SARS-CoV-2 Infection Diagnosis. J Med Virol 2020. doi:10.1002/jmv.25727

[52] Zalinger ZB, Elliott R, Rose KM, Weiss SR. MDA5 Is Critical to Host Defense during Infection with Murine Coronavirus. J Virol 2015;89:12330-40. doi:10.1128/JVI.01470-15

[53] Hoffman H, Verhave B, Chin LS. Hypernatremia is associated with poorer outcomes following aneurysmal subarachnoid hemorrhage: a nationwide inpatient sample analysis. J Neurosurg Sci 2018. doi:10.23736/S0390-5616.18.04611-8

[54] Yuan P, Yang Z, Song H, Wang K, Yang Y, Xie L, et al. Three Main Inducers of Alphacoronavirus Infection of Enterocytes: Sialic Acid, Proteases, and Low pH. Intervirology 2018;61:53-63. doi:10.1159/000492424

[55] Raabis SM, Ollivett TL, Cook ME, Sand JM, McGuirk SM. Health benefits of orally administered anti-IL-10 antibody in milk-fed dairy calves. J Dairy Sci 2018;101:7375-82. doi:10.3168/jds.2017-14270

[56] Petrarca L, Nenna R, Frassanito A, Pierangeli A, Di Mattia G, Scagnolari C, et al. Human bocavirus in children hospitalized for acute respiratory tract infection in Rome. World J Pediatr 2019. doi:10.1007/s12519-019-00324-5

[57] Habib AMG, Ali MAE, Zouaoui BR, Taha MAH, Mohammed BS, Saquib N. Clinical outcomes among hospital patients with Middle East respiratory syndrome coronavirus (MERS-CoV) infection. BMC Infect Dis 2019;19:870. doi:10.1186/s12879-019-4555-5

FIGURE 1. Baseline summary of the recruited cohort

(A) There were 75 positive and 62 negative samples, respectively. The columns "Positive Std" and "Negative Std" gave the standard deviations of the specific feature in each sample group. The last column "Pvalue" gave the T-test Pvalue of that specific feature between the two sample groups. A feature name starting with "Blood | " and "Urine | " was collected from the blood test and urine test, respectively. **(B)** The heatmap matrix of the inter-feature Pearson correlation coefficient (PCC) for all the features and the group value. The values ranged between 0.00 and 1.00, and the color was linearly rendered according the inter-feature PCC. The feature names starting with "Blood | " and "Urine | " were from the blood test and urine test, respectively.

FIGURE 2. Performance metrics of five prediction algorithms

The horizontal axis was the four performance metrics, aAcc, aSn, aSp, and aMCC, which were averaged over the twenty random runs. The vertical axis gave the values of these four metrics. The bar heights and the error bars of these histograms were the averages and standard deviations of these metrics over the twenty random runs of each algorithm.

FIGURE 3. Youden index of different SVM thresholds

The three line plots were Sn, Sp and J, respectively. The horizontal axis was the threshold values sorted in the descending order. The vertical axis was the value of these three metrics Sn, Sp and J.

FIGURE 4. Heatmaps of the SVM parameter tuning

Two kernel functions were evaluated, i.e., **(A)** RBF and **(B)** Linear. The SVM parameter C had five value choices: 0.01, 0.1, 1, 10 and 100. The other parameter Gamma had five value choices: 1, 0.1, 0.01, 0.001 and "scale", where "scale" was the default value in the Python library. The four detection performance metrics Acc/Sn/Sp/MCC were used to evaluate the models.

FIGURE 5. Recursive eliminating the features

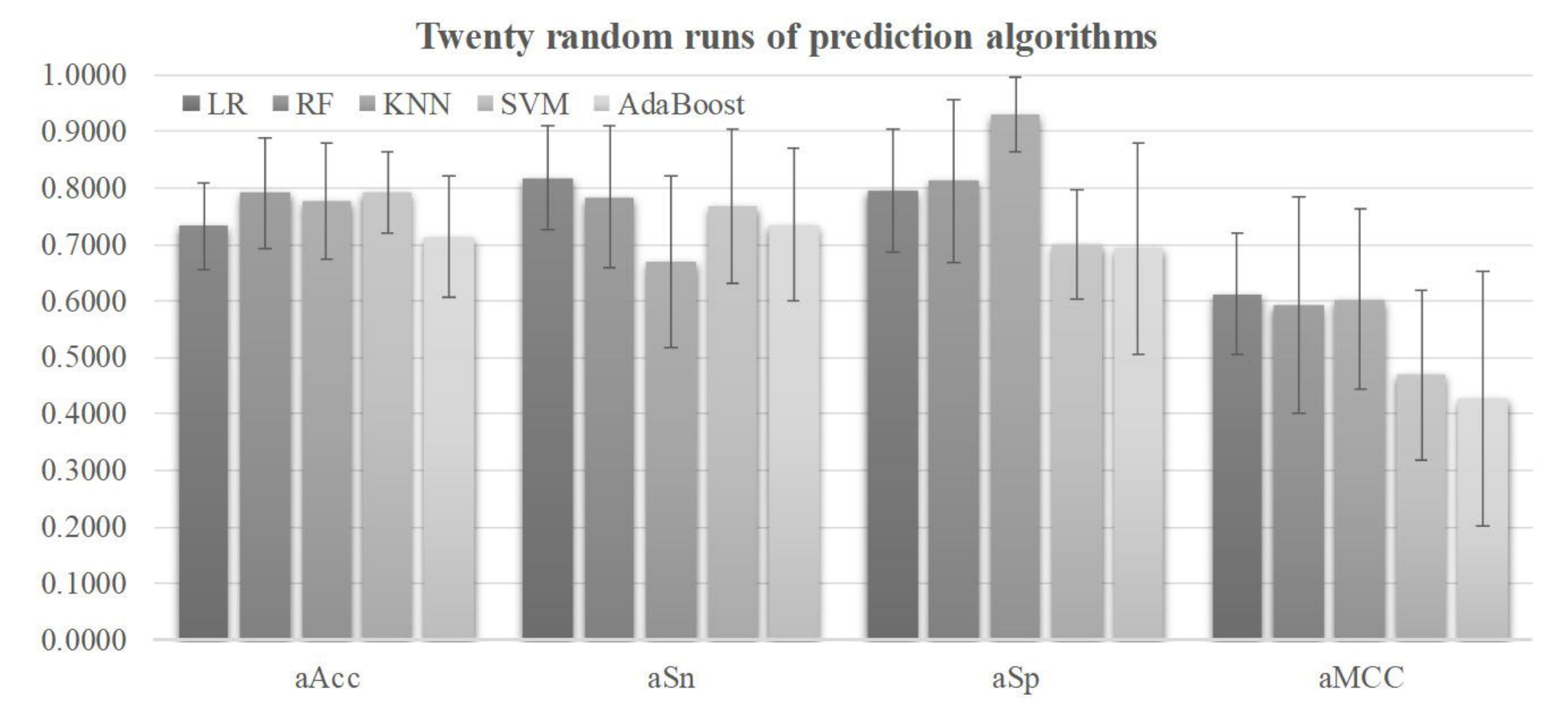
The horizontal axis listed the features ranked in the ascending order by their T-test Pvalues. The vertical axis was the threshold-independent metric Area Under the Curve (AUC) achieved by the model using the feature set specified by the horizontal axis.

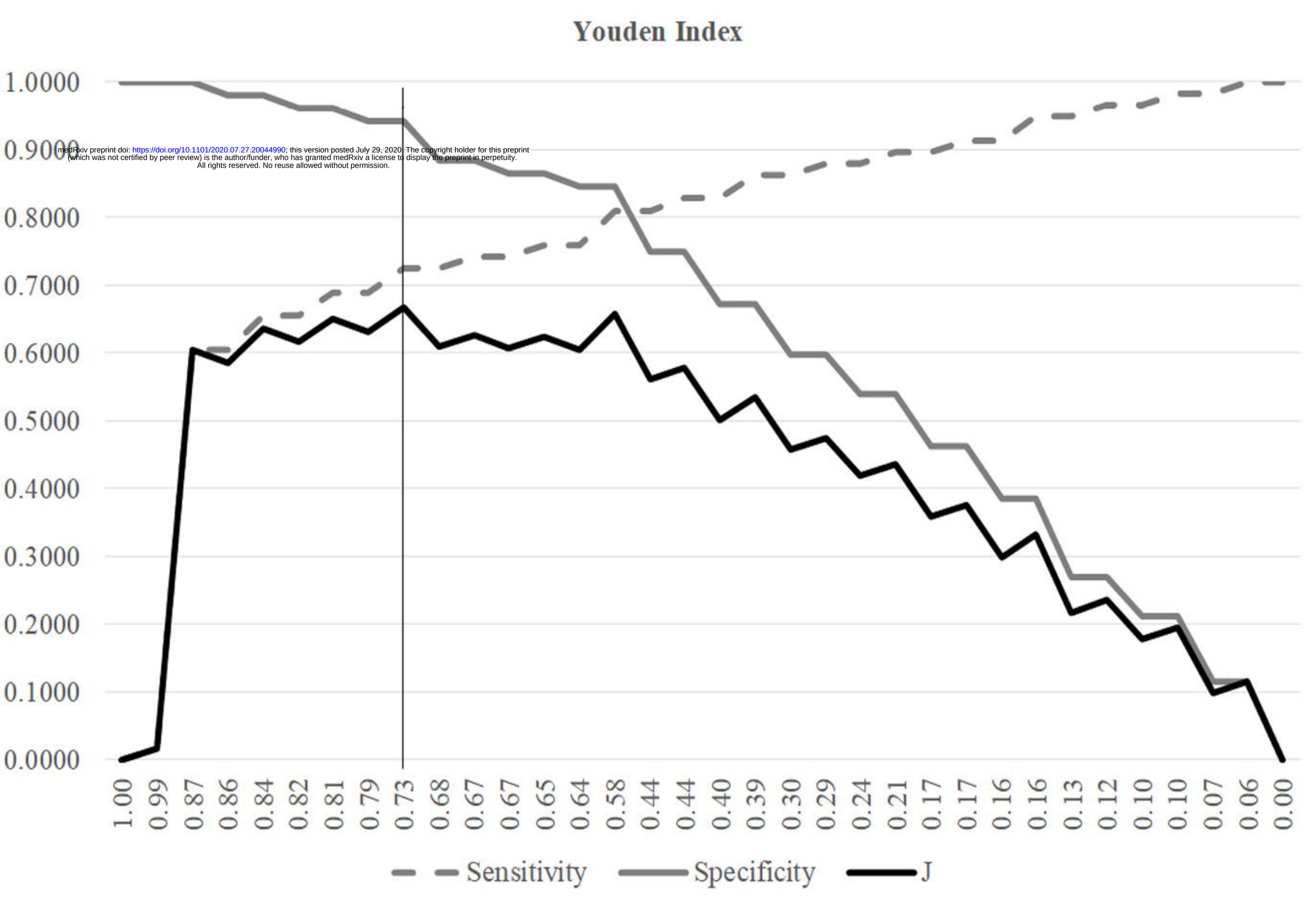
(A)

Features	Positive samples (75)	Positive Std	Negative samples (62)	Negative Std	Pvalue
Age	65.0533	14.5462	51.4839	17.2196	1.75E-06
Sex (Male/Female)	59/16	0.4124	29/33	0.5030	7.71E-05
Blood | Neutrophil percentage	82.5440	14.5598	58.7823	23.7638	4.14E-11
Blood | Calcium	1.0819	0.1884	2.2119	0.1149	9.55E-10
Blood | Monocyte percentage	8.6567	9.7006	17.3052	12.9063	1.62E-05
Urine | Urine protein	0.4467	0.7692	0.1613	0.3708	8.41E-03
Urine | Red blood cell(occult)	0.4400	0.8052	0.1774	0.5364	2.98E-02
Urine | PH(Urine)	6.1633	0.3275	6.2661	0.2432	4.25E-02

(B)

Age	1.00	0.14	0.36	0.25	0.29	0.32	0.14	0.32	0.19	0.14	0.22	0.21	0.23	0.06	0.30	0.12	0.27	0.16	0.18	0.08	0.19	0.18	0.07	0.14	0.03	0.12	0.02	0.17	0.17	0.14	0.08	0.28	0.40
Sex	0.14	1.00	0.17	0.22	0.21	0.16	0.21	0.02	0.11	0.16	0.08	0.20	0.21	0.09	0.20	0.15	0.03	0.17	0.08	0.08	0.16	0.07	0.07	0.18	0.02	0.21	0.04	0.12	0.12	0.09	0.09	0.17	0.33
Blood | Neutrophil percentage	0.36	0.17	1.00	0.48	0.43	0.41	0.33	0.34	0.37	0.06	0.15	0.28	0.22	0.58	0.21	0.19	0.30	0.25	0.23	0.11	0.14	0.20	0.14	0.14	0.08	0.15	0.12	0.14	0.04	0.22	0.25	0.11	0.53
Blood | Calcium	0.25	0.22	0.48	1.00	0.28	0.26	0.30	0.16	0.24	0.06	0.10	0.15	0.15	0.35	0.06	0.20	0.18	0.21	0.36	0.16	0.09	0.05	0.02	0.05	0.15	0.17	0.12	0.12	0.03	0.13	0.02	0.05	0.49
Blood | Monocyte percentage	0.29	0.21	0.43	0.28	1.00	0.38	0.21	0.46	0.24	0.09	0.68	0.28	0.09	0.21	0.10	0.17	0.42	0.12	0.64	0.12	0.21	0.29	0.15	0.07	0.09	0.09	0.11	0.13	0.02	0.12	0.15	0.11	0.36
Blood | Thrombin time	0.32	0.16	0.41	0.26	0.38	1.00	0.02	0.56	0.31	0.27	0.45	0.47	0.15	0.26	0.29	0.36	0.65	0.34	0.31	0.17	0.09	0.48	0.12	0.44	0.41	0.15	0.49	0.39	0.07	0.32	0.32	0.20	0.35
Blood | hs-CRP	0.14	0.21	0.33	0.30	0.21	0.02	1.00	0.08	0.04	0.10	0.05	0.17	0.19	0.26	0.05	0.10	0.05	0.27	0.11	0.05	0.15	0.06	0.14	0.01	0.01	0.12	0.02	0.01	0.09	0.11	0.07	0.13	0.33
Blood | Neutrophil Number	0.32	0.02	0.34	0.16	0.46	0.56	0.08	1.00	0.14	0.31	0.55	0.37	0.08	0.18	0.20	0.01	0.73	0.04	0.39	0.17	0.03	0.67	0.14	0.02	0.18	0.13	0.29	0.32	0.14	0.28	0.22	0.08	0.32
Blood | Glucose	0.19	0.11	0.37	0.24	0.24	0.31	0.04	0.14	1.00	0.09	0.07	0.23	0.29	0.20	0.09	0.37	0.06	0.18	0.14	0.00	0.18	0.16	0.12	0.28	0.02	0.17	0.01	0.23	0.03	0.11	0.07	0.02	0.31
Blood | Leucocytes	0.14	0.16	0.06	0.06	0.09	0.27	0.10	0.31	0.09	1.00	0.19	0.14	0.00	0.11	0.38	0.09	0.21	0.06	0.12	0.12	0.12	0.13	0.09	0.03	0.10	0.02	0.15	0.12	0.20	0.12	0.11	0.07	0.31
Blood | Basophil percentage	0.22	0.08	0.15	0.10	0.68	0.45	0.05	0.55	0.07	0.19	1.00	0.28	0.00	0.14	0.14	0.13	0.69	0.06	0.55	0.23	0.11	0.45	0.04	0.04	0.16	0.12	0.20	0.11	0.08	0.05	0.09	0.12	0.29
Blood | Total bilirubin	0.21	0.20	0.28	0.15	0.28	0.47	0.17	0.37	0.23	0.14	0.28	1.00	0.14	0.14	0.20	0.20	0.31	0.51	0.23	0.11	0.04	0.46	0.02	0.54	0.39	0.11	0.42	0.07	0.03	0.17	0.21	0.33	0.28
Blood | Interleukin-2R	0.23	0.21	0.22	0.15	0.09	0.15	0.19	0.08	0.29	0.00	0.00	0.14	1.00	0.12	0.03	0.23	0.00	0.41	0.03	0.12	0.12	0.15	0.13	0.44	0.09	0.42	0.06	0.04	0.22	0.04	0.01	0.12	0.28
Blood | Packed cell volume	0.06	0.09	0.58	0.35	0.21	0.26	0.26	0.18	0.20	0.11	0.14	0.14	0.12	1.00	0.06	0.00	0.26	0.11	0.25	0.41	0.02	0.10	0.03	0.06	0.03	0.04	0.11	0.01	0.30	0.15	0.11	0.01	0.28
Blood | RBC distribution width SD	0.30	0.20	0.21	0.06	0.10	0.29	0.05	0.20	0.09	0.38	0.14	0.20	0.03	0.06	1.00	0.20	0.18	0.07	0.07	0.20	0.04	0.21	0.09	0.22	0.30	0.03	0.24	0.09	0.54	0.13	0.21	0.16	0.27
Blood | Interleukin-10	0.12	0.15	0.19	0.20	0.17	0.36	0.10	0.01	0.37	0.09	0.13	0.20	0.23	0.00	0.20	1.00	0.00	0.39	0.08	0.08	0.09	0.05	0.07	0.54	0.21	0.22	0.20	0.08	0.14	0.19	0.18	0.18	0.26
Blood | Prothrombin time	0.27	0.03	0.30	0.18	0.42	0.65	0.05	0.73	0.06	0.21	0.69	0.31	0.00	0.26	0.18	0.00	1.00	0.10	0.36	0.17	0.02	0.60	0.10	0.05	0.13	0.21	0.23	0.27	0.09	0.12	0.09	0.04	0.26
Blood | Serum ferritin	0.16	0.17	0.25	0.21	0.12	0.34	0.27	0.04	0.18	0.06	0.06	0.51	0.41	0.11	0.07	0.39	0.10	1.00	0.02	0.17	0.04	0.02	0.01	0.51	0.26	0.29	0.30	0.08	0.12	0.09	0.13	0.31	0.25
Blood | Eosinophil percentage	0.18	0.08	0.23	0.36	0.64	0.31	0.11	0.39	0.14	0.12	0.55	0.23	0.03	0.25	0.07	0.08	0.36	0.02	1.00	0.11	0.03	0.25	0.03	0.04	0.08	0.05	0.12	0.13	0.00	0.12	0.11	0.05	0.24
Blood | Platelet distribution width	0.08	0.08	0.11	0.16	0.12	0.17	0.05	0.17	0.00	0.12	0.23	0.11	0.12	0.41	0.20	0.08	0.17	0.17	0.11	1.00	0.08	0.08	0.09	0.15	0.17	0.01	0.20	0.07	0.01	0.07	0.07	0.01	0.24
Blood | Oxygen partial pressure	0.19	0.16	0.14	0.09	0.21	0.09	0.15	0.03	0.18	0.12	0.11	0.04	0.12	0.02	0.04	0.09	0.02	0.04	0.03	0.08	1.00	0.06	0.47	0.06	0.09	0.13	0.07	0.10	0.03	0.07	0.03	0.04	0.23
Blood | Alkaline phosphatase	0.18	0.07	0.20	0.05	0.29	0.48	0.06	0.67	0.16	0.13	0.45	0.46	0.15	0.10	0.21	0.05	0.60	0.02	0.25	0.08	0.06	1.00	0.08	0.02	0.08	0.11	0.14	0.38	0.19	0.14	0.12	0.05	0.23
Blood | Total carbon dioxide	0.07	0.07	0.14	0.02	0.15	0.12	0.14	0.14	0.12	0.09	0.04	0.02	0.13	0.03	0.09	0.07	0.10	0.01	0.03	0.09	0.47	0.08	1.00	0.09	0.08	0.02	0.03	0.23	0.08	0.14	0.04	0.08	0.22
Blood | Tumor necrosis factor-α	0.14	0.18	0.14	0.05	0.07	0.44	0.01	0.02	0.28	0.03	0.04	0.54	0.44	0.06	0.22	0.54	0.05	0.51	0.04	0.15	0.06	0.02	0.09	1.00	0.41	0.15	0.39	0.02	0.09	0.19	0.17	0.36	0.20
Blood | Chlorine	0.03	0.02	0.08	0.15	0.09	0.41	0.01	0.18	0.02	0.10	0.16	0.39	0.09	0.03	0.30	0.21	0.13	0.26	0.08	0.17	0.09	0.08	0.08	0.41	1.00	0.09	0.93	0.05	0.18	0.23	0.30	0.20	0.19
Blood | Alanine aminotransferase	0.12	0.21	0.15	0.17	0.09	0.15	0.12	0.13	0.17	0.02	0.12	0.11	0.42	0.04	0.03	0.22	0.21	0.29	0.05	0.01	0.13	0.11	0.02	0.15	0.09	1.00	0.04	0.31	0.24	0.06	0.02	0.05	0.19
Blood | Sodium	0.02	0.04	0.12	0.12	0.11	0.49	0.02	0.29	0.01	0.15	0.20	0.42	0.06	0.11	0.24	0.20	0.23	0.30	0.12	0.20	0.07	0.14	0.03	0.39	0.93	0.04	1.00	0.11	0.10	0.23	0.29	0.19	0.19
Blood | Aspartase aminotransferase	0.17	0.12	0.14	0.12	0.13	0.39	0.01	0.32	0.23	0.12	0.11	0.07	0.04	0.01	0.09	0.08	0.27	0.08	0.13	0.07	0.10	0.38	0.23	0.02	0.05	0.31	0.11	1.00	0.02	0.04	0.02	0.03	0.18
Blood | RBC distribution width CV	0.17	0.12	0.04	0.03	0.02	0.07	0.09	0.14	0.03	0.20	0.08	0.03	0.22	0.30	0.54	0.14	0.09	0.12	0.00	0.01	0.03	0.19	0.08	0.09	0.18	0.24	0.10	0.02	1.00	0.03	0.06	0.08	0.17
Urine | Urine protein	0.14	0.09	0.22	0.13	0.12	0.32	0.11	0.28	0.11	0.12	0.05	0.17	0.04	0.15	0.13	0.19	0.12	0.09	0.12	0.07	0.07	0.14	0.14	0.19	0.23	0.06	0.23	0.04	0.03	1.00	0.62	0.31	0.22
Urine | Red blood cell(occult)	0.08	0.09	0.25	0.02	0.15	0.32	0.07	0.22	0.07	0.11	0.09	0.21	0.01	0.11	0.21	0.18	0.09	0.13	0.11	0.07	0.03	0.12	0.04	0.17	0.30	0.02	0.29	0.02	0.06	0.62	1.00	0.16	0.19
Urine | PH(Urine)	0.28	0.17	0.11	0.05	0.11	0.20	0.13	0.08	0.02	0.07	0.12	0.33	0.12	0.01	0.16	0.18	0.04	0.31	0.05	0.01	0.04	0.05	0.08	0.36	0.20	0.05	0.19	0.03	0.08	0.31	0.16	1.00	0.17
Class	0.40	0.33	0.53	0.49	0.36	0.35	0.33	0.32	0.31	0.31	0.29	0.28	0.28	0.28	0.27	0.26	0.26	0.25	0.24	0.24	0.23	0.23	0.22	0.20	0.19	0.19	0.19	0.18	0.17	0.22	0.19	0.17	1.00


Twenty random runs of prediction algorithms
1.0000
0.9000
0.8000
0.7000
0.6000
0.5000
0.4000
0.3000
0.2000
0.1000
0.0000
LR
RF
KNN
SVM
AdaBoost
aAcc
aSn
aSp
aMCC



(A)

Acc		Gamma				
		1.0000	0.1000	0.0100	0.0010	scale
C	0.01	0.4545	0.4545	0.4545	0.4545	0.4545
	0.1	0.4545	0.4545	0.4545	0.4545	0.4545
	1	0.4545	0.5909	0.8636	0.4545	0.8636
	10	0.4545	0.5909	0.8182	0.8636	0.7273
	100	0.4545	0.5909	0.8182	0.9091	0.7273

Sp		Gamma				
		1	0.1	0.01	0.001	scale
C	0.01	1.0000	1.0000	1.0000	1.0000	1.0000
	0.1	1.0000	1.0000	1.0000	1.0000	1.0000
	1	1.0000	1.0000	1.0000	1.0000	1.0000
	10	1.0000	1.0000	0.9000	1.0000	1.0000
	100	1.0000	1.0000	1.0000	1.0000	1.0000

Sn		Gamma				
		1	0.1	0.01	0.001	scale
C	0.01	0.0000	0.0000	0.0000	0.0000	0.0000
	0.1	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.0000	0.2500	0.7500	0.0000	0.7500
	10	0.0000	0.2500	0.7500	0.7500	0.5000
	100	0.0000	0.2500	0.6667	0.8333	0.5000

MCC		Gamma				
		1	0.1	0.01	0.001	scale
C	0.01	0.0000	0.0000	0.0000	0.0000	0.0000
	0.1	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.0000	0.3627	0.7596	0.0000	0.7596
	10	0.0000	0.3627	0.6500	0.7596	0.5590
	100	0.0000	0.3627	0.6901	0.8333	0.5590

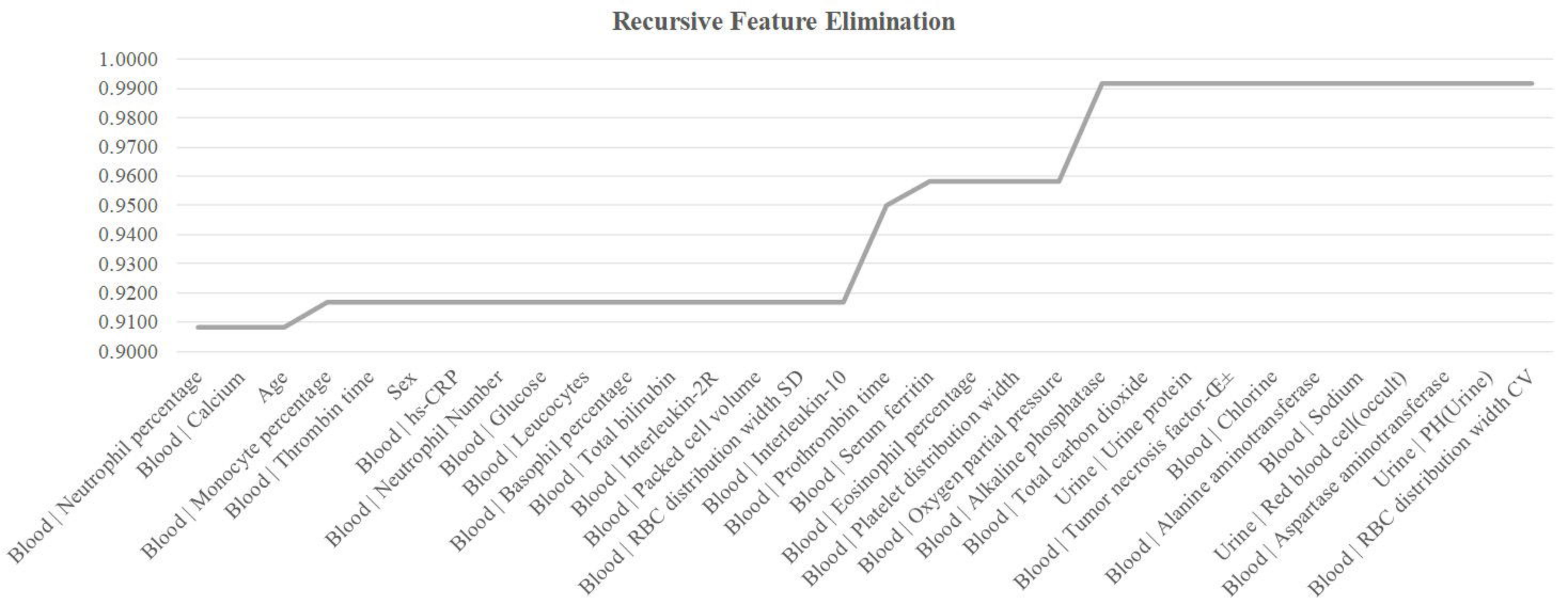
(B)

Acc		Gamma				
		1	0.1	0.01	0.001	scale
C	0.01	0.8182	0.8182	0.8182	0.8182	0.8182
	0.1	0.8636	0.8636	0.8636	0.8636	0.8636
	1	0.8636	0.8636	0.8636	0.8636	0.8636
	10	0.7273	0.7273	0.7273	0.7273	0.7273
	100	0.7727	0.7727	0.7727	0.7727	0.7727

Sp		Gamma				
		1	0.1	0.01	0.001	scale
C	0.01	1.0000	1.0000	1.0000	1.0000	1.0000
	0.1	1.0000	1.0000	1.0000	1.0000	1.0000
	1	1.0000	1.0000	1.0000	1.0000	1.0000
	10	0.8000	0.8000	0.8000	0.8000	0.8000
	100	0.8000	0.8000	0.8000	0.8000	0.8000

Sn		Gamma				
		1	0.1	0.01	0.001	scale
C	0.01	0.6667	0.6667	0.6667	0.6667	0.6667
	0.1	0.7500	0.7500	0.7500	0.7500	0.7500
	1	0.7500	0.7500	0.7500	0.7500	0.7500
	10	0.6667	0.6667	0.6667	0.6667	0.6667
	100	0.7500	0.7500	0.7500	0.7500	0.7500

MCC		Gamma				
		1	0.1	0.01	0.001	scale
C	0.01	0.6901	0.6901	0.6901	0.6901	0.6901
	0.1	0.7596	0.7596	0.7596	0.7596	0.7596
	1	0.7596	0.7596	0.7596	0.7596	0.7596
	10	0.4667	0.4667	0.4667	0.4667	0.4667
	100	0.5477	0.5477	0.5477	0.5477	0.5477


Recursive Feature Elimination
1.0000
0.9900
0.9800
0.9700
0.9600
0.9500
0.9400
0.9300
0.9200
0.9100
0.9000
Blood | Neutrophil percentage
Blood | Calcium
Age
Blood | Monocyte percentage
Blood | Thrombin time
Sex
Blood | hs-CRP
Blood | Neutrophil Number
Blood | Glucose
Blood | Leucocytes
Blood | Basophil percentage
Blood | Total bilirubin
Blood | Interleukin-2R
Blood | Packed cell volume
Blood | RBC distribution width SD
Blood | Interleukin-10
Blood | Prothrombin time
Blood | Serum ferritin
Blood | Eosinophil percentage
Blood | Platelet distribution width
Blood | Oxygen partial pressure
Blood | Alkaline phosphatase
Blood | Total carbon dioxide
Urine | Urine protein
Blood | Tumor necrosis factor-Œ±
Blood | Chlorine
Blood | Alanine aminotransferase
Blood | Sodium
Urine | Red blood cell(occult)
Blood | Aspartase aminotransferase
Urine | PH(Urine)
Blood | RBC distribution width CV