

Received January 30, 2021, accepted March 15, 2021, date of publication March 19, 2021, date of current version March 29, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3067311

# An Effective Machine Learning Approach for Identifying Non-Severe and Severe Coronavirus Disease 2019 Patients in a Rural Chinese Population: The Wenzhou Retrospective Study

PEILIANG WU<sup>1</sup>, HUA YE<sup>2</sup>, XUEDING CAI<sup>1</sup>, CHENGYE LI<sup>1</sup>, SHIMIN LI<sup>3</sup>, MENGXIANG CHEN<sup>4</sup>, MINGJING WANG<sup>3</sup>, ALI ASGHAR HEIDARI<sup>5,6</sup>, MAYUN CHEN<sup>1</sup>, JIFA LI<sup>2</sup>, HUILING CHEN<sup>3</sup>, (Associate Member, IEEE), XIAOYING HUANG<sup>1</sup>, AND LIANGXING WANG<sup>1</sup>

<sup>1</sup>Department of Pulmonary and Critical Care Medicine, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou 325000, China

<sup>2</sup>Department of Pulmonary and Critical Care Medicine, Affiliated Yueqing Hospital, Wenzhou Medical University, Yueqing 325600, China

<sup>3</sup>College of Computer Science and Artificial Intelligence, Wenzhou University, Wenzhou 325035, China

<sup>4</sup>Department of Information Technology, Wenzhou Vocational College of Science and Technology, Wenzhou 325006, China

<sup>5</sup>School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran 1417466191, Iran

<sup>6</sup>Department of Computer Science, School of Computing, National University of Singapore, Singapore 117417

Corresponding authors: Huiling Chen (chenhuiling.jlu@gmail.com), Xiaoying Huang (zjwzhxy@126.com), and Liangxing Wang (38805@163.com)

This work was supported in part by the Project of Health Commission of Zhejiang Province under Grant 2020KY177 and Grant 2019RC047, in part by the Wenzhou Technology Foundation under Grant Y20180600, in part by the First Affiliated Hospital of Wenzhou Medical University Youth Excellence Project under Grant QNYC114, and in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LQ19H010003.

**ABSTRACT** This paper has proposed an effective intelligent prediction model that can well discriminate and specify the severity of Coronavirus Disease 2019 (COVID-19) infection in clinical diagnosis and provide a criterion for clinicians to weigh scientific and rational medical decision-making. With indicators as the age and gender of the patients and 26 blood routine indexes, a severity prediction framework for COVID-19 is proposed based on machine learning techniques. The framework consists mainly of a random forest and a support vector machine (SVM) model optimized by a slime mould algorithm (SMA). When the random forest was used to identify the key factors, SMA was employed to train an optimal SVM model. Based on the COVID-19 data, comparative experiments were conducted between RF-SMA-SVM and several well-known machine learning algorithms performed. The results indicate that the proposed RF-SMA-SVM not only achieves better classification performance and higher stability on four metrics, but also screens out the main factors that distinguish severe COVID-19 patients from non-severe ones. Therefore, there is a conclusion that the RF-SMA-SVM model can provide an effective auxiliary diagnosis scheme for the clinical diagnosis of COVID-19 infection.

**INDEX TERMS** COVID-19, coronavirus, support vector machine, slime mould algorithm, disease diagnosis, feature selection.

## I. INTRODUCTION

On December 31, 2019, the Health Commission of Hubei Province of China notified the World Health Organization (WHO) of a rapidly evolving outbreak of unexplained

The associate editor coordinating the review of this manuscript and approving it for publication was Jenny Mahoney.

viral pneumonia, which is now believed to have begun since December 1, 2019 [1]. On February 11, 2020, this viral pneumonia has been named ‘coronavirus disease 2019’ (COVID-19) by the WHO [2]. At the same time, according to the recommendations from the International Committee on Taxonomy of Viruses (ICTV), the 2019 novel coronavirus was named as ‘severe acute respiratory syndrome

coronavirus 2' (SARS-CoV-2) [3]. By Mar 23th, 2020, a total of 332,930 laboratory-confirmed cases of SARS-CoV-2 infection have been reported, including 14,510 deaths [4]. Therefore, COVID-19 has been described as a global pandemic [5], [6].

COVID-19 is characterized by a combination of clinical features that include fever, dry cough, sputum, myalgia, tachypnea, fatigue, lymphopenia and abnormal chest computed tomography (CT) imaging [1], [7], [8]. With the improvement of accuracy on the diagnosis of patients with COVID-19, prediction of outcome in critically ill patients with COVID-19 is becoming a more and more important and urgent issue for treatment decisions and resource allocation at high-risk infection regions. A growing number of clinical studies have shown 26.1-32% of patients with COVID-19 develop severe COVID-19, and severe COVID-19 is characterized by rapid progression, severe acute respiratory disease syndrome (ARDS), multiple organs dysfunction, and high mortality [1], [7]. The previous study has described that elderly patients are more susceptible to COVID-19 infection, especially those with diabetes, hypertension, chronic pulmonary diseases, and cerebral infarction [9]. Likewise, a high fatality rate (61.5%) was observed among 52 critically ill adult patients with COVID-19 in Wuhan, China [10]. Due to COVID-19 rapid progression, a demand for intensive care unit beds is rising in a short period of time when the resource is constrained. Hence, compared to rescue and treatment for critically ill patients, early identification and intervention play a more indispensable role in preventing COVID-19 from the mild to severe. If effective novel prognostic model systems are established in time, would help to surveil the status of patients with COVID-19, with the potential for mortality reduction.

So far, it is recognized that COVID-19 is a disease involving multiple situations. Its clinical manifestation is complicated and changeable, and the critically ill cases rapidly develop, especially in patients with acute respiratory disease syndrome [11]. However, the survival rate of critically ill cases remains low, with the rising number of critically ill patients, and the focus on controlling of the infection source and cutting off the route of transmission in health and epidemic prevention work has shifted to timely detection and effective assessment of critically ill patients, and early intervention before fatal events appear. Currently, the diagnosis of critically ill and fatal cases is usually based on a large number of clinical laboratory parameters and clinical characteristics through judgment calls. Therefore, misdiagnosis and missed diagnosis easily exist because of physicians' fatigue, pressure, lack of experience, inconsistent judgments and lack of prior knowledge. Continuously, COVID-19 outbreaks are operating, worldwide, and meanwhile the rates of misdiagnosis and missed diagnosis are becoming the world's most closely watched numbers.

In the present study, the data was retrospectively collected from patients with COVID-19 pneumonia who were admitted to the Affiliated Yueqing Hospital of Wenzhou Medical

University from Jan 21 to Mar 10, 2020, in a single-center retrospective study. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was detected by real-time reverse transcriptase-polymerase chain reaction (RT-PCR). RT-PCR assay kit was provided by Huada Genomics Co., Ltd. Throat swabs from all 51 patients tested positive for SARS-CoV-2 by RT-PCR. According to the "the New Coronavirus Pneumonia Prevention and Control" promulgated by the National Health Commission of the People's Republic of China, in 2020, all patients fulfilled diagnostic criteria of COVID-19 pneumonia. To simplify the analysis, the 51 COVID-19 patients were divided the 51 COVID-19 patients into two groups: severe COVID-19 ( $n = 21$ ) and non-severe COVID-19 ( $n = 30$ ). Severe COVID-19 diagnosis should meet at least 1 of the following criteria: (1) Respiratory distress and the respiratory rate over 30 breaths/min; (2) Resting oxygen saturation lower than 93%; (3) Oxygenation index (OI) below 300 mmHg. The research plan was permitted by the Ethics Committees of the Affiliated Yueqing Hospital of Wenzhou Medical University (Yueqing, China) (Ethical approval code: 202000002). This study was in compliance with the Helsinki declaration. Blood samples were collected from all COVID-19 patients and blood routine examination was conducted, using the whole blood by a Mindray BC-5300 automatic blood cell analyzer (Mindray, Shenzhen, China). In this study, the basic information of the COVID-19 patients and 23 blood routine parameters (features) are listed in Table 1.

In recent years, the development and application of machine learning in term of infectious disease have been beneficial for the prevention and control of infectious diseases. Machine learning methods play more and more significant roles in a wide fields of medical science [12], [13], which is widely performed in disease diagnosis, development of prediction models and identification of significant risk factors [14]. Undoubtedly, machine learning methods have some advantages including improving health professionals' diagnostic capacities, reducing much of the workload of radiologists and anatomic/clinical pathologists, enhancing the diagnostic accuracy [15]. Therefore, machine learning has emerged as an indispensable tool for physicians to lead a better understanding for patient personalized treatments. In the meantime, it is hoped that machine learning methods based on the abundant clinical data will support researches into COVID-19 epidemiology, which can cope with the current disease outbreaks [16]. Therefore, machine learning based on clinical characteristics is a promising method for the prediction of survival rate and classification of disease severity. To this end, a new method with the random forest (RF) and slime mould algorithm (SMA) optimized support vector machines (SVM) model involved was established, to analyze data of 51 patients confirmed with COVID-19 pneumonia who were admitted to the Affiliated Yueqing Hospital of Wenzhou Medical University (Yueqing, China).

Since then, several machine learning methods have been taken to tackle the related problems of COVID-19. For example, Wang *et al.* [17] firstly used a Respiratory

**TABLE 1.** List of the features used in this study and their definitions.

Number	Features	Abbreviation
F1	Gender	Gender
F2	Age	Age
F3	White blood cell	WBC
F4	Neutrophil percentage	NEU%
F5	Lymphocyte percentage	LY%
F6	Monocyte percentage	MONO%
F7	Eosinophils percentage	EOS%
F8	Basophils percentage	BA%
F9	Neutrophil count	NEU
F10	Lymphocyte count	LY
F11	Monocyte count	MONO
F12	Eosinophils count	EOS
F13	Basophils count	BA
F14	Red blood cell	RBC
F15	Haemoglobin	HB
F16	Hematocrit	HCT
F17	Mean corpuscular hemoglobin	MCH
F18	Mean corpuscular volume	MCV
F19	Mean corpuscular hemoglobin concentration	MCHC
F20	Red blood cell distribution width	RDW
F21	Red blood cell distribution width-SD	RDW-SD
F22	Mean platelet volume	MPV
F23	Blood platelet	PLT
F24	Plateletcrit	PCT
F25	Platelet distribution width	PDW
F26	Platelet-large cell ratio	P-LCR%,
F27	Neutrophil to Lymphocyte ratios	N/L
F28	Platelet to Lymphocyte ratios	P/L

Simulation Model (RSM) to fill the gap between large amounts of training data and scarce real-world data. A GRU neural network with bidirectional and attentional mechanisms (BI-AT-GRU) was then operated to classify 6 key clinically respiratory patterns for screening large-scale individuals infected with COVID-19. Hu *et al.* [18] developed an improved stacked autoencoder for modeling epidemic transmission dynamics and applying this model to predicting COVID-19 confirmed cases in real-time across China. Butt *et al.* developed a three-dimensional deep learning model by using lung CT image sets to segment candidate infection areas, and then utilizing a location attention classification model to classify these separated images into COVID-19, influenza A viral pneumonia, and healthy cases. Wu *et al.* [19] performed a calibration of logistic growth model, and generalized logistic growth model, growth model, and Richards model on the number of reported infection cases in the severely affected areas. Different models provide predictions with different upper and lower limits. Shan *et al.* [20] developed a deep learning-based segmentation system to automatically quantify possible infected areas and their volume ratios in the lungs. At the same time, a human-in-the-loop (HITL) strategy was used to assist the radiologist in segmenting the

infected area, thereby shortening the total segmentation time to 4 minutes. Yang *et al.* [21] integrated population migration data and the latest COVID-19 epidemiological data into the Susceptible-Exposed-Infectious-Removed (SEIR) model to obtain the epidemic curve, and also used artificial intelligence (AI) derived from 2003 SARS data ways to predict the next outbreak. Vivanco-Lira [22] tried to summarize the biological characteristics of COVID-19 virus and the possible pathophysiology of its disease, as well as a random model that characterizes the probability distribution of cases in Mexico and the number of Mexican cases estimated by differential equation models. Gozes *et al.* [23] proposed a method that leverages powerful 2D and 3D deep learning models to modify and adjust existing AI models, and combined them with clinically understood systems, and used 3D volumetric examinations to assess each patient's disease progression over time.

In this study, it aims to propose efficient models by using the SMA, which is employed to train an effective SVM model. Then, it is the first time that the optimized SMA-SVM is employed to diagnose the severity of COVID-19. The active model is built by using the info about patients' basic information and 26 blood routine indicators. In the

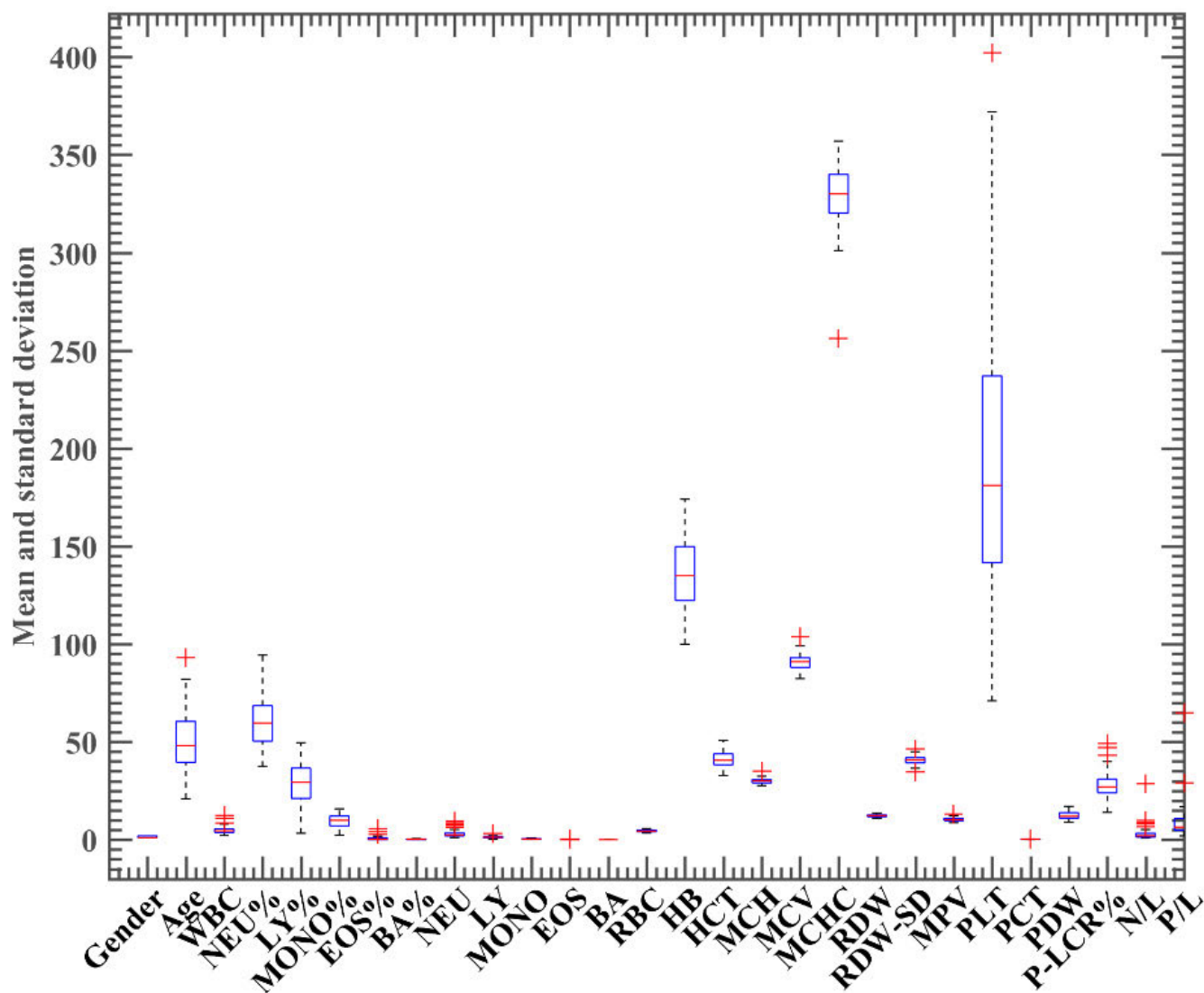


FIGURE 1. Boxplot of the 28 features.

developed model (SMA-SVM), SMA was used for training an SVM model and random forest was employed to identify the most informative risk factors of COVID-19 infected people at the same time. In the experiment, machine learning methods such as random forest (RF) combined with SMA-SVM (RF-SMA-SVM), RF combined with Harris hawks optimization-based SVM (RF-HHO-SVM), RF combined with moth-flame optimization-based SVM (RF-MFO-SVM), RF combined with particle swarm optimization-based SVM (RF-PSO-SVM), RF combined with grid search-based SVM (RF-Grid-SVM) and SMA-SVM were rigorously compared. The results showed that the established RF-SMA-SVM method performed far better than its peers on four evaluation metrics: classification accuracy (ACC), sensitivity, specificity and Matthews correlation coefficient (MCC). Therefore, the main contributions of this study can be listed below.

a) Successful applications of SMA are for parameter tuning of SVM.

b) The proposed RF-SMA-SVM was the first time to apply in the diagnosis of the severity of COVID-19.

(c) It is the first time that the age, Neutrophil-to-Lymphocyte ratio (N/L) and Platelet-to-lymphocyte ratio (P/L) combined together for predicting the prognosis of patients with COVID-19.

d) The established RF-SMA-SVM model outperforms other competitors.

The structure of the paper is as below. Section 2 describes the data and proposed the RF-SMA-SVM model in detail. The experimental setup and results are presented in Section 3. Section 4 presents the discussions. The conclusions and future works are given in Section 5.

## II. MATERIALS AND METHODS

### A. DATA COLLECTION

The boxplot of total 28 indexes are shown in Figure 1. Statistical analyses were performed by using SPSS version 21.0 (IBM, Somers, NY, US). Continuous variables (age and blood

routine parameters) were analyzed by an independent sample t-test. Results were considered statistically significant with  $p < 0.05$ . Detailed results of the statistical analysis are shown in Table 2.

**B. RF-SMA-SVM METHOD**

The main flow of the proposed method is illustrated in Figure 5. It consists of four steps: data acquisition with standardization, feature selection, model training, and classification. The first step is to normalize the acquired data to the range of  $[-1, 1]$ , and then use the random forest algorithm to select the feature set with the highest score from the data set and SMA for training the parameters  $(C, \gamma)$  in the SVM. The third step uses the optimized parameters and the feature set selected by the random forest to train the SVM again. The last step is to use the trained SVM for classification of new data. When evaluating the trained model, K-fold cross-validation (CV) is used to obtain less biased experimental results. In this experiment, K is set to 10.

1) FEATURE SELECTION VIA RANDOM FOREST

Random forest (RF) [24], is a combined classification model consisting of multiple decision tree models. It employs bootstrap resampling to take multiple samples from the original sample and models a decision tree for each bootstrap sample. Then, these decision trees are combined together, and the final classification or prediction result is obtained by voting. In addition, a large number of theoretical and empirical studies have shown that RF not only has high prediction accuracy and excellent tolerance to outliers and noise, but also has the ability to calculate the importance of a single feature and avoid easily overfit.

The steps of calculating the importance of a feature by using RF is shown in Figure 2.

2) PARAMETER OPTIMIZATION BY SMA

Swarm intelligent algorithms can always generate better solutions than the traditional gradient descent-based methods [25]–[27], SMA as a new member of the swarm intelligent algorithms, it has shown great effectiveness in many fields. SMA mainly simulates the behaviour mode of Physarum polycephalum in slime mould [28]. SMA mainly uses the weighting factor to simulate the changes of the vein structure and contraction pattern of slime mould during foraging. When the concentration of nearby food is large, the diameter of slime bacteria will increase accordingly. When the concentration of food decreases, the diameter of slime bacteria will also shrink, forming a tendency to approach food. The steps can be described by the following steps:

*Step 1:* Initialization parameters and individual positions. Such as population number  $N$ , dimension  $D$ , parameter  $z$  and randomly distribute the search agents in the solution space.

$$X(i, :) = \text{rand} \cdot (\text{UB} - \text{LB}) + \text{LB} \tag{2.1}$$

where  $X_i$  represents the  $i$ -th search agent, rand represents a random number with a value between 0 and 1, and LB and UB represent the lower and upper boundaries of the search space, respectively.

*Step 2:* Calculate fitness value and weight. First calculate the fitness value of each search agent, sort all fitness values  $F$  to get  $S$  and  $I$ , and get the best fitness values bF and wF from them. Then calculate the weight  $W$ .

$$[S, I] = \text{sort}(F) \tag{2.2}$$

$$\vec{W}(I(i), j) = \begin{cases} 1 + r \cdot \log\left(\frac{\text{bF} - S(i)}{\text{bF} - \text{wF}} + 1\right), & i \leq \frac{N}{2} \\ 1 - r \cdot \log\left(\frac{\text{bF} - S(i)}{\text{bF} - \text{wF}} + 1\right), & i > \frac{N}{2} \end{cases} \tag{2.3}$$

where  $i = 1 \dots N, j = 1 \dots D$ .

*Step 3:* Update the global optimal fitness value DF and the best search agent position BP (parameter to be optimized). If the fitness value obtained during the current iteration process is better than the global optimal fitness value, the global optimal fitness value is updated to the currently obtained fitness value and the best search agent position is updated to the currently obtained position.

*Step 4:* Update search agent location (parameter value to be optimized). First update the following parameters:

$$a = \text{atanh}\left(-\frac{T}{\text{Max}_t} + 1\right) \tag{2.4}$$

$$b = 1 - \frac{T}{\text{Max}_t} \tag{2.5}$$

where  $T$  is the current number of iterations and  $\text{Max}_t$  is the maximum number of iterations.

Then update the search agent location according to the following formula:

$$\vec{X}^* = \begin{cases} \text{rand} \cdot (\text{UB} - \text{LB}) + \text{LB}, & \text{rand} < z \\ \vec{\text{BP}}(t) + \vec{\text{vb}} \cdot (W \cdot \vec{X}_A(t) - \vec{X}_B(t)), & r < p \\ \vec{\text{vc}} \cdot \vec{X}(t), & r \geq p \end{cases} \tag{2.6}$$

where A and B are two individuals randomly selected from the group, and the formula for  $p, \vec{\text{vb}}$  and  $\vec{\text{vc}}$  are as follows:

$$p = \tanh |S(i) - \text{DF}| \tag{2.7}$$

$$\vec{\text{vb}} = [-a, a] \tag{2.8}$$

$$\vec{\text{vc}} = [-b, b] \tag{2.9}$$

Figure 3 shows the flowchart of SMA.

3) CLASSIFICATION BASED ON SVM

SVM is a popular machine learning method that has shown effectiveness in theory and applications [29]–[36]. SVM uses the principle of non-linear mapping:  $\Phi : R^n \rightarrow H$ , maps nonlinear eigendata to high dimensional space, thus using linear methods to solve nonlinear problems. For a load sample data set  $D = \{x_i, y_i\}, i = 1, 2, \dots, n, x_i \in R^n, y_i \in R$  where  $y_i$  is the load value of the output and  $x_i$  is the eigenvector corresponding to the output value, SVM maps the eigenvector

**TABLE 2. Statistical analysis of 51 COVID-19 patients (severe COVID-19 and non-severe COVID-19).**

Index	Severe (n = 21)	Non-severe (n = 30).	p-value
Age (years)	61.43±17.64	42.30±11.53	0.00
WBC (10 <sup>9</sup> /L)	5.06±2.23	4.79±1.89	0.64
NEU (%)	69.50±14.10	55.33±9.98	0.00
LY (%)	21.20±10.54	33.25±7.90	0.00
MONO (%)	8.71±4.04	10.35±3.19	0.11
EOS (%)	0.44±1.22	0.88±0.96	0.16
BA (%)	0.15±0.19	0.19±0.13	0.38
NEU (10 <sup>9</sup> /L)	3.68±2.22	2.74±1.59	0.08
LY (10 <sup>9</sup> /L)	0.95±0.46	1.50±0.45	0.00
MONO (10 <sup>9</sup> /L)	0.40±0.19	0.48±0.18	0.15
EOS (10 <sup>9</sup> /L)	0.01±0.03	0.04±0.04	0.02
BA (10 <sup>9</sup> /L)	0.01±0.01	0.01±0.01	0.22
RBC (10 <sup>12</sup> /L)	4.44±0.66	4.58±0.47	0.35
HB (g/L)	132.67±18.68	138.53± 14.71	0.22
HCT (%)	40.25±4.92	41.83±3.87	0.21
MCH (pg)	29.65±1.22	30.28±1.47	0.11
MCV (fL)	90.35±3.97	91.54±4.08	0.30
MCHC (g/L)	323.76± 21.23	330.87± 10.77	0.12
RDW (%)	12.48±0.55	12.083±0.58	0.02
RDW-SD (fL)	41.27±1.74	40.19±2.47	0.09
MPV (fL)	10.42±0.92	10.43±0.95	0.97
PLT (10 <sup>9</sup> /L)	170.14± 63.10	205.30± 69.82	0.07
PCT (%)	0.19±0.07	0.21±0.07	0.26
PDW (%)	12.62±2.29	12.77±2.01	0.81
P-LCR (%)	28.14±7.35	28.10±7.09	0.98
N/L	5.35±5.97	1.85±0.86	0.02
P/L	12.53±13.63	6.52±2.62	0.02

$x_i$  to high dimensional space by function  $\varphi(x)$ . Establish the regression function:

$$y = \omega \cdot \varphi(x) + b \tag{3.1}$$

where  $\omega$  is the weight,  $b$  is the intercept,  $x = (x_1, x_2, \dots, x_n)$  is the eigenvector,  $y = (y_1, y_2, \dots, y_n)$  is the output load value.

To find a suitable set of  $\omega$  and  $b$ , the search process was transformed into an optimization problem according to the principle of structural risk minimization. The objective function and constraints are as follows:

$$\begin{cases} \min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \\ s.t. \begin{cases} \omega \cdot \varphi(x_i) - y_i + b \leq \varepsilon + \xi_i^* \\ y_i - \omega \cdot \varphi(x_i) - b \leq \varepsilon + \xi_i \\ \xi_i^* \geq 0, \xi_i \geq 0, i = 1, 2, \dots, n \end{cases} \end{cases} \tag{3.2}$$

where  $C$  is the penalty factor,  $\xi_i$  and  $\xi_i^*$  are relaxation coefficients, allowing a certain fitting error,  $\varepsilon$  is the maximum error, and is a linear insensitive loss function parameter.

For the sake of enhancing the efficiency of optimization, Lagrange multiplier method is applied to transform it into

dual problem:

$$\begin{cases} \min \frac{1}{2} \sum_{i,j=1}^n (a_i^* - a_i) (a_j^* - a_j) K(x_i, x_j) \\ \quad + \varepsilon \sum_{i=1}^n (a_i^* + a_i) - \sum_{i=1}^n y_i (a_i^* - a_i) \\ s.t. \begin{cases} \sum_{i=1}^n (a_i - a_i^*) = 0 \\ 0 \leq a_i, a_i^* \leq C, i = 1, 2, \dots, n \end{cases} \end{cases} \tag{3.3}$$

where  $a_i$  and  $a_i^*$  are Lagrange multipliers,  $K(x_i, x_j)$  is a kernel function. Gaussian radial basis kernel function was employed:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \tag{3.4}$$

where  $\gamma$  is a kernel parameter. Eventually, the nonlinear function was obtained.

$$y = \sum_{i=1}^n (a_i - a_i^*) K(x_i, x_j) + b \tag{3.5}$$

In SVM, the penalty factor  $C$  and the kernel parameter  $\gamma$  have a crucial impact on the learning process, and the prediction effects obtained by different parameters are quite diverse.  $C$  indicates the generalization ability and error size of SVM. Larger  $C$  will make the model error larger and

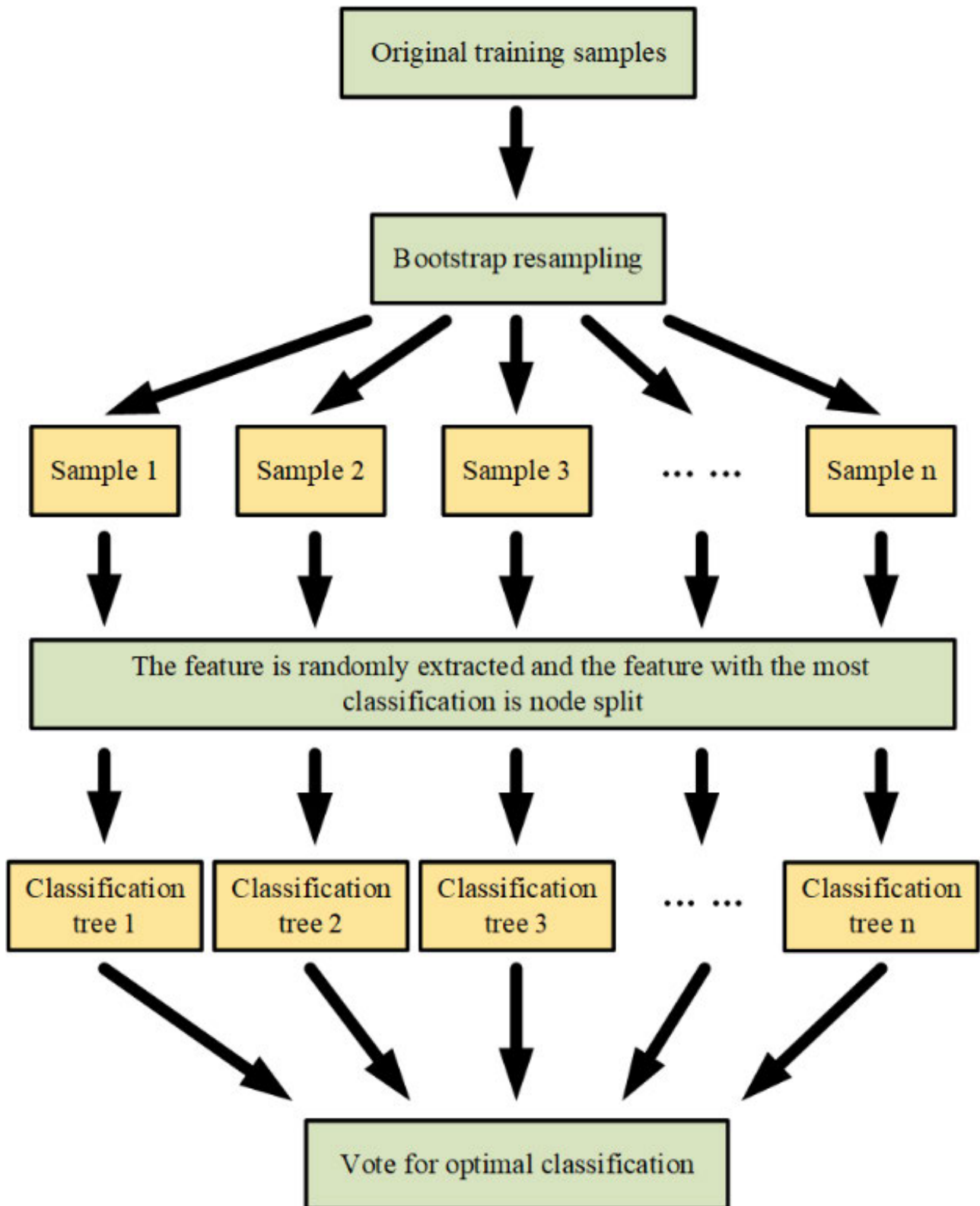


FIGURE 2. Flowchart of RF.

the precision is not enough. Smaller  $C$  will lead to weaker generalization ability.  $\gamma$  affects the fit of the model. In order to find a better set of  $C$ ,  $\gamma$  and improve the classification

accuracy of SVM, this paper combined SMA and SVM to produce a new classification model SMA-SVM and applied the model to the prediction problem in reality.

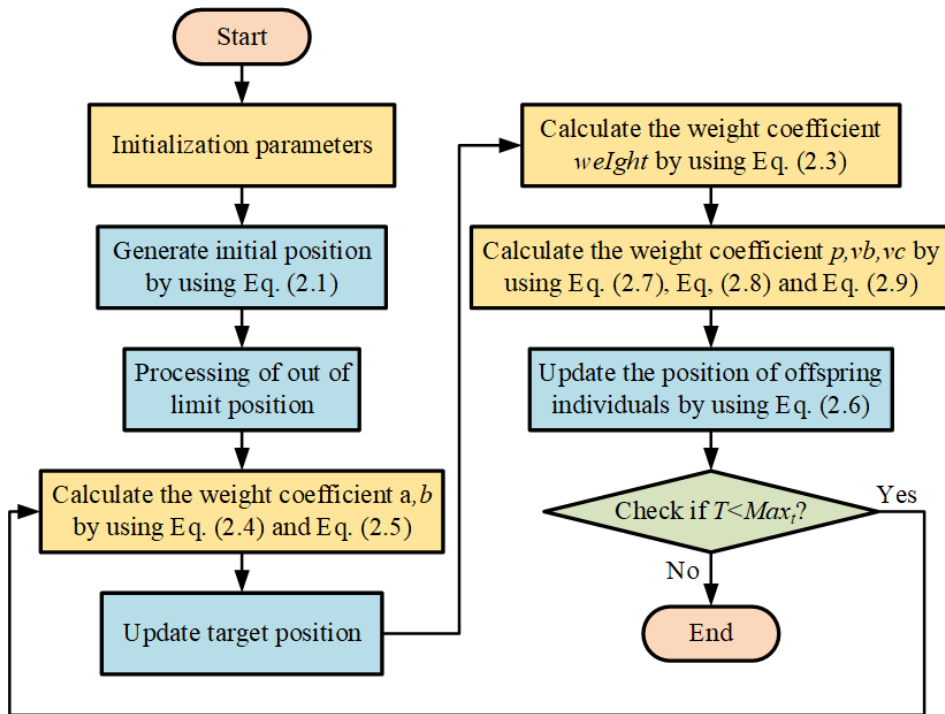


FIGURE 3. Flowchart of SMA.

4) PROPOSED SMA-SVM

To explore the potential of SVM, its parameters are tuned via the SMA. Here, we show the flowchart of parameter optimization of the proposed SMA-SVM model.

In Figure 4, adjust the position of the searched individual and calculate the fitness of all search agents using SVM classifier with the parameter pair according to the following formula:

$$fitness = \frac{\sum_{i=1}^K acc_i}{K} \tag{4.1}$$

III. EXPERIMENTAL SETUP AND RESULTS

The entire experiment was built in a MATLAB environment. The main methods involved in the comparison experiments include RF-SMA-SVM and RF-HHO-SVM, RF-MFO-SVM, RF-PSO-SVM, RF-Grid-SVM, SMA-SVM. The optimization methods including HHO, MFO, PSO, SMA, etc. were implemented using the default parameters from the original paper, while the SVM model and RF methods were implemented using the LIBSVM toolbox and RF toolbox respectively.<sup>1</sup> The input data were normalized to the range [-1, 1] before classification. The maximum number of iterations and the number of group members are set to 50 and 20, respectively. The search range for the two parameters in the SVM (penalty factor *C* and kernel parameter  $\gamma$ ) is set to  $[2^{-5}, 2^5]$ .

<sup>1</sup><https://www.csie.ntu.edu.tw/~cjlin/libsvm/>, <https://code.google.com/archive/p/randomforest-matlab>

In this experiment, the RF method was first time to use in assessment of the significant features, and the specific results are shown in Figure 6. As can be seen from the figures, the individual features exhibit different levels of importance, with the top eight most vital features being Age, LY%, LY, N/L, NEU%, P/L, RDW, EOS%, respectively. Sorting this feature sequence in descending order of importance yields with multiple feature combinations, was employed to be evaluated one by one. Considering the number of features, here we select a set of eight incremental features consisting of the first eight features for the experiment. The results of the are shown in Table 3. It can be seen from the table that the best classification accuracy is achieved when the features consisting of the first 6 features.

The detailed results of RF-SMA-SVM model are shown in Table 4. In order to ensure the accuracy of the results, we repeated 10 times, taking the average of the 10 results as the final result. The results showed that the classification accuracy (ACC) of RF-SMA-SVM was 90.10%, the Matthews correlation coefficient (MCC) was 82.18%, the sensitivity was 88.17%, the specificity was 91.41%, and the variances were 0.1245, 0.2247, 0.2210 and 0.1788, respectively. In addition, it can be observed that the SMA can automatically obtain the optimal parameters of SVM on the optimal feature space obtained from RF, which demonstrates the strong search ability of SMA. the strong search ability of SMA.

To verify the validity of this approach, a comparative study presents with five other valid machine learning models, including RF-HHO-SVM, RF-MFO-SVM, RF-PSO-SVM,



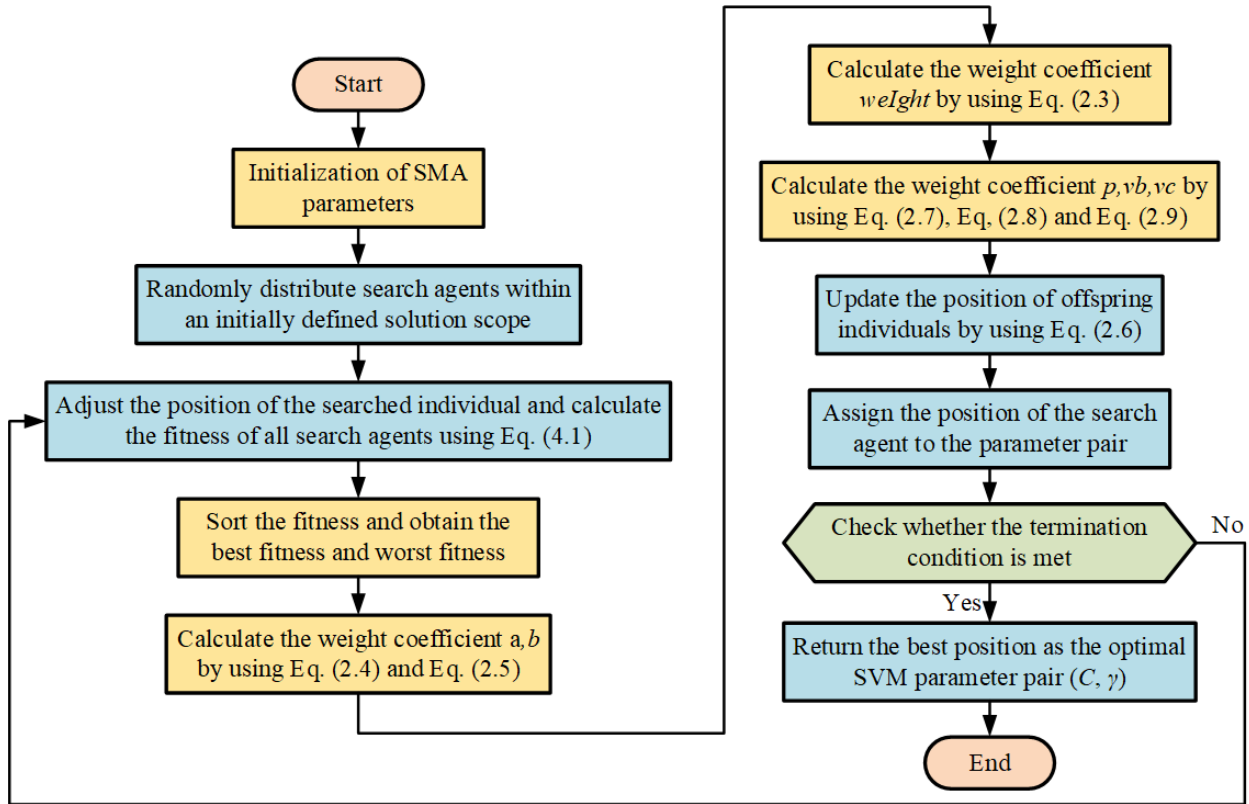


FIGURE 4. Flowchart of SMA-SVM.

TABLE 3. Classification performance of RF-SMA-SVM based on eight feature sets.

No.	Feature subset	ACC	MCC	Sensitivity	Specificity
No.1	{Age}	0.783	0.577	0.650	0.867
No.2	{Age, LY%}	0.807	0.640	0.683	0.900
No.3	{Age, LY%, LY}	0.787	0.573	0.617	0.900
No.4	{Age, LY%, LY, N/L}	0.810	0.641	0.733	0.867
No.5	{Age, LY%, LY, N/L, NEU%}	0.843	0.704	0.800	0.867
No.6	{Age, LY%, LY, N/L, NEU%, P/L}	0.900	0.812	0.850	0.942
No.7	{Age, LY%, LY, N/L, NEU%, P/L, RDW}	0.887	0.801	0.833	0.933
No.8	{Age, LY%, LY, N/L, NEU%, P/L, RDW, EOS%}	0.843	0.688	0.750	0.900

TABLE 4. Classification performance of RF-SMA-SVM in terms of ACC, MCC, sensitivity, and specificity.

Num	ACC	MCC	Sensitivity	Specificity
No.1	0.9000	0.8116	0.8500	0.9417
No.2	0.9000	0.8220	0.9000	0.9000
No.3	0.9030	0.8150	0.9000	0.9000
No.4	0.9070	0.8280	0.8500	0.9330
No.5	0.9030	0.8320	0.9170	0.9000
No.6	0.9000	0.8220	0.9000	0.9000
No.7	0.9000	0.8250	0.8500	0.9330
No.8	0.9070	0.8280	0.8500	0.9330
No.9	0.9030	0.8320	0.9170	0.9000
No.10	0.8870	0.8130	0.8830	0.9000
AVG.	0.9010	0.8218	0.8817	0.9141
STD	0.1245	0.2247	0.2210	0.1788

RF-Grid-SVM, and SMA-SVM. The comparison of six methods is shown in Figure 7. The results showed that the RF-SMA-SVM model outperformed the RF-HHO-SVM,

RF-MFO-SVM, RF-PSO-SVM, RF-Grid-SVM, and SMA-SVM models in terms of ACC, MCC, and specificity, with the least variance. Besides, RF-SMA-SVM outperformed

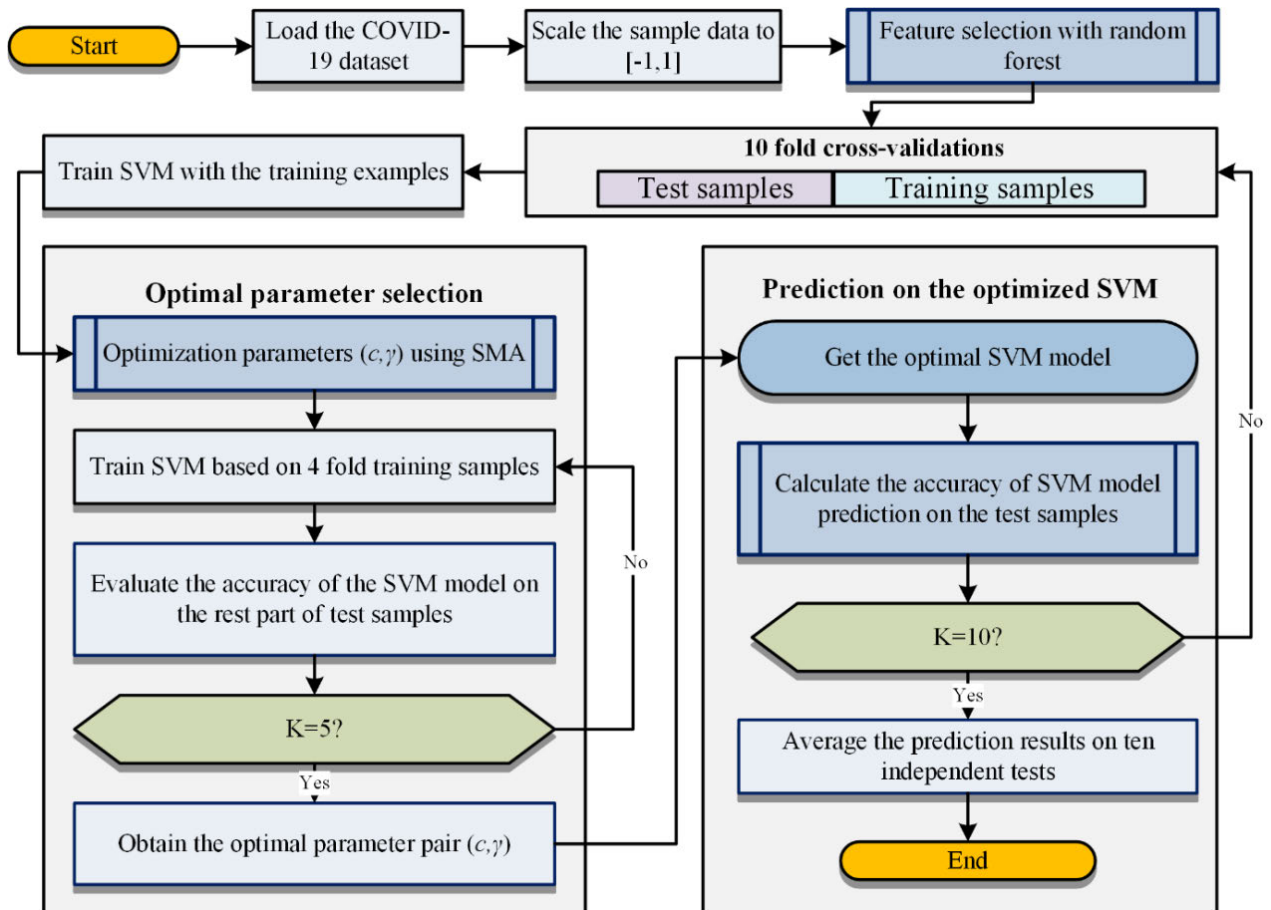


FIGURE 5. Flowchart of RF-SMA-SVM.

RF-MFO-SVM, RF-PSO-SVM and SMA-SVM in sensitivity evaluation metric had lower variance than these three models. This means that the RF-SMA-SVM model gets better classification performance and stability compared to other optimization methods. The RF-SMA-SVM model had the best results on the ACC evaluation metric, 2 percentage points higher than the RF-HHO-SVM, which ranked second. This is followed by RF-MFO-SVM, RF-PSO-SVM, RF-Grid-SVM, and SMA-SVM, with SMA-SVM being 8 percentage points lower than RF-SMA-SVM, and RF-SMA-SVM having the smallest variance of 0.1245, indicating that the RF-SMA-SVM model is more stable in resolving this problem after conduction of feature selection by RF. For the MCC evaluation indicator, the RF-SMA-SVM model still achieved the best results, followed by RF-HHO-SVM, RF-MFO-SVM, RF-PSO-SVM and RF-Grid-SVM, and otherwise SMA-SVM performed the worst, with 15.1 percentage points lower than RF-SMA-SVM. In addition, RF-SMA-SVM had the lowest variance of 0.1992. In terms of sensitivity evaluation indicators, the RF-HHO-SVM and RF-SMA-SVM models were the most effective, followed by the RF-Gird-SVM model with a difference of 3 percentage points. This is followed by RF-MFO-SVM and RF-PSO-SVM, with

RF-MFO-SVM being 3.33 percentage points higher compared to RF-PSO-SVM. Moreover, the SMA-SVM model had the worst results and RF-PSO-SVM had the largest variance of 0.2770. In terms of specificity evaluation indicator, the RF-SMA-SVM model has the best results, followed by RF-PSO-SVM, which differs from RF-SMA-SVM by 1.5 percentage points. RF-HHO-SVM, RF-MFO-SVM followed closely, with SMA-SVM being 5.4 percentage points lower than RF-SMA-SVM, RF-Grid-SVM having the largest variance of 0.212 and RF-PSO-SVM having the smallest variance of 0.1688.

The effect of original SMA in test function is far better than PSO and MFO [28], which indicates that SMA has strong exploration and development ability, and these abilities can also play an important role in medical data analysis. Figure 7 and figure 8 prove that RF-SMA-SVM is superior to RF-HHO-SVM, RF-MFO-SVM and RF-PSO-SVM. Although RF-SMA-SVM is slightly inferior to RF-HHO-SVM in sensitivity, it does not affect the superiority of RF-SMA-SVM. Furthermore, 1 is the upper limit to measure these indicators. If you want to be close to 1, the performance requirements of the algorithm are multiplied. Therefore, RF-SMA-SVM seems to have little difference from

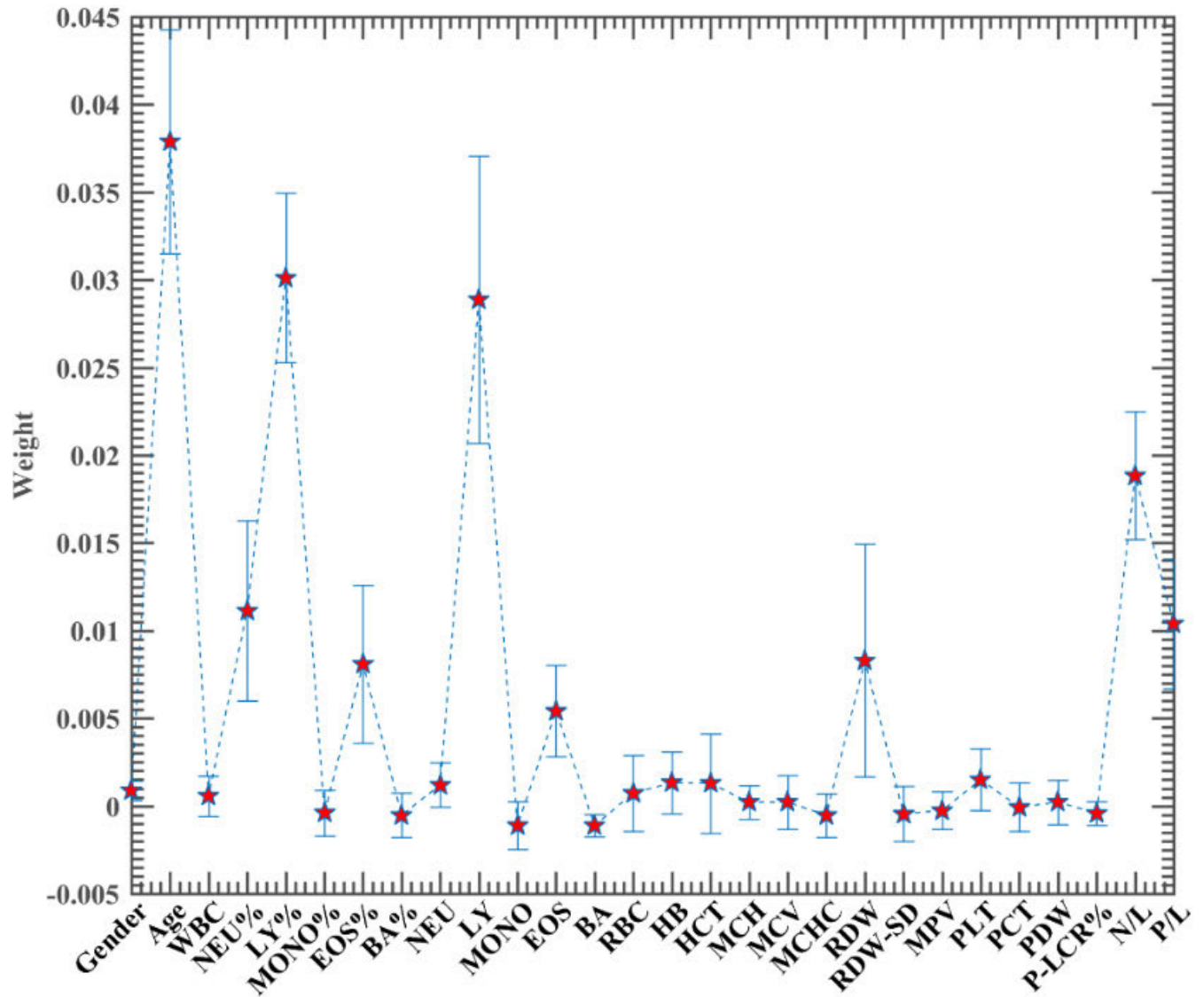


FIGURE 6. Importance of features obtained by RF.

TABLE 5. Calculation time comparison between RF-SMA-SVM and other algorithms.

Algorithm	Relative time	Rank
RF-SMA-SVM	4.597	4
RF-MFO-SVM	1.097	2
RF-PSO-SVM	1.000	1
RF-HHO-SVM	1.290	3

other comparison algorithms, but in fact there is a certain gap.

To describe the convergence of the proposed RF-SMA-SVM method, we also document the trend of the accuracy of the four SVM models with iteration. In Figure 8, it is found that after several iterations, SMA-SVM has the worst effect, and the accuracy of SMA-SVM does not improve significantly with iterations, so it is easy to fall into the

local optimum, while RF-SMA-SVM model can quickly and continuously jump out of the local optimum to reach the optimal accuracy, which shows that RF-SMA-SVM method has strong local search ability and global search ability. The accuracy of RF-MFO-SVM and RF-PSO-SVM is slightly lower than that of RF-SMA-SVM, and the accuracy increases less significantly with iteration, which makes it easier to fall into local optimization.

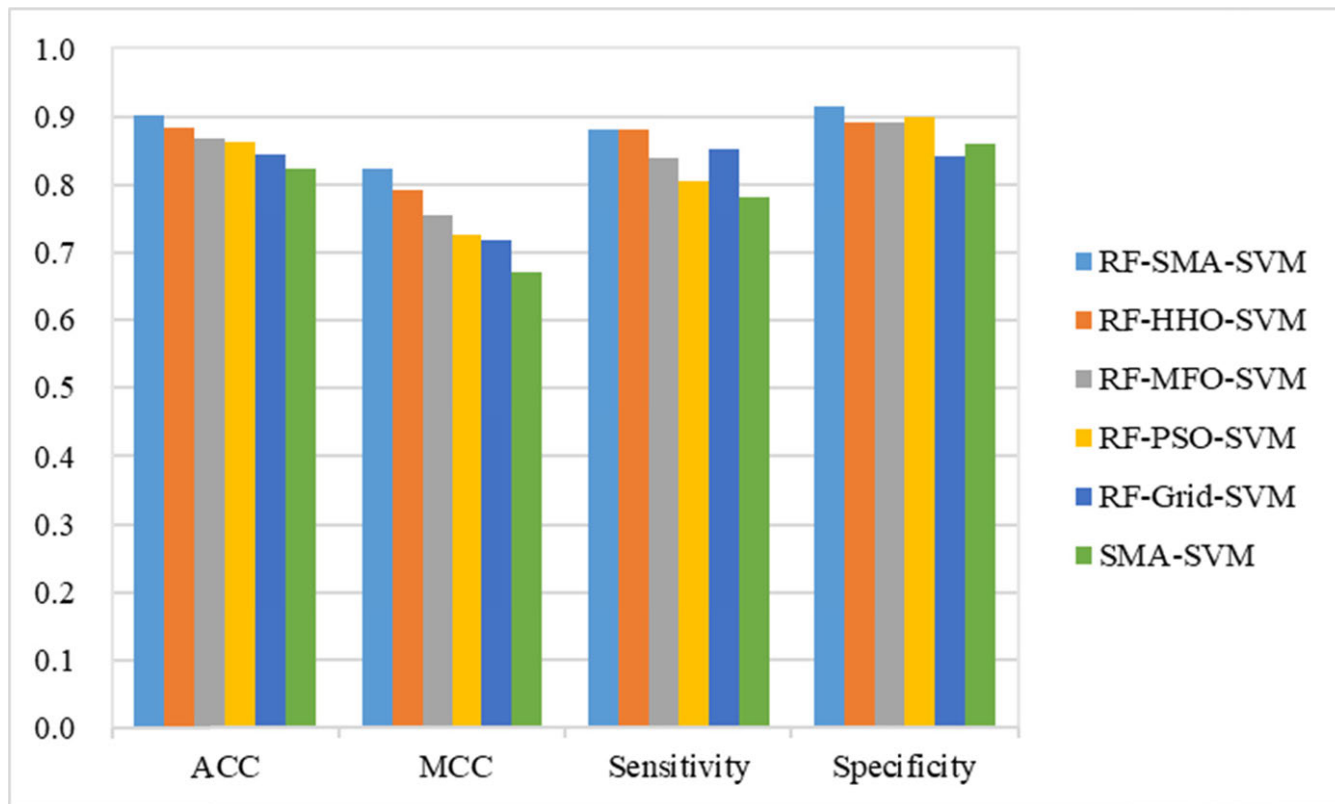


FIGURE 7. Classification efficacy of six models based on the four metrics.

At the end of the experiment, we compare the computing time of RF-SMA-SVM, RF-MFO-SVM, RF-PSO-SVM and RF-HHO-SVM. The results are shown in Table 5.

The calculation time of RF-PSO-SVM is the shortest, so the relative calculation time of RF-PSO-SVM is recorded as 1. The relative calculation time of other algorithms is calculated according to the calculation time of RF-PSO-SVM. In Table 5, although RF-SMA-SVM has the longest computation time, its accuracy is better than other algorithms.

#### IV. DISCUSSIONS

In the present study, COVID-19 patient datasets were screened by using the random-forest based feature selection method. Importantly, several key features were screened, including age, NEU%, LY%, LY, N/L and P/L. Subsequently, a SMA-SVM model is constructed to accurately assess the severity of the COVID-19 and early monitor COVID-19 progression. Therefore, we believe that RF-SMA-SVM model may help inform clinical decision-making.

Some studies demonstrate that advanced age is a strong predictor for prognosis of Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS). For instance, Choi and colleagues used multi-variable analysis to identify age older as independent predictors of mortality [37]–[39]. In line with these findings from clinical studies, studies on macaques model revealed that older macaques inoculated with SARS coronavirus

(SARS-CoV) exhibited stronger host antiviral response compared to younger macaques, which is tightly related to inflammation-related genes and type I interferon [40]. Similarly to these studies, Zhou *et al.* found that non-survivor with COVID-19 had a significantly older median age, than did the survivor with COVID-19 [41]. We also found in our study that patients with severe COVID-19 had significantly higher (1.45-fold) age than those with a non-severe COVID-19 ( $P = 0.00$ ), suggesting that disease severity may be predicted by age in COVID-19 patients. In addition, numerous clinical studies have demonstrated that advanced age is positively associated with the severity of infection and sepsis [42], [43]. The probable reason for this was that lymphocyte function decrease with increasing age may increase the risk of infection and overwhelming secretion of type 2 cytokines may produce an overstimulation of viral replication, eventually leading to body function impairment [44]–[46]. All in all, age is an important predictor of rehabilitation in patients with COVID-19.

White blood cells (WBCs) are blood cells that mediate the immune system and resist exotic microorganisms. There are five types of WBCs, including neutrophils, lymphocytes, monocytes, eosinophils, and basophils [47]. Neutrophilic granulocyte is the chief phagocytic white blood cell of the blood, comprising 50-70% of the total leukocytes in the blood [48]. Neutrophilic granulocytes play a momentous role in the unspecific immune response and

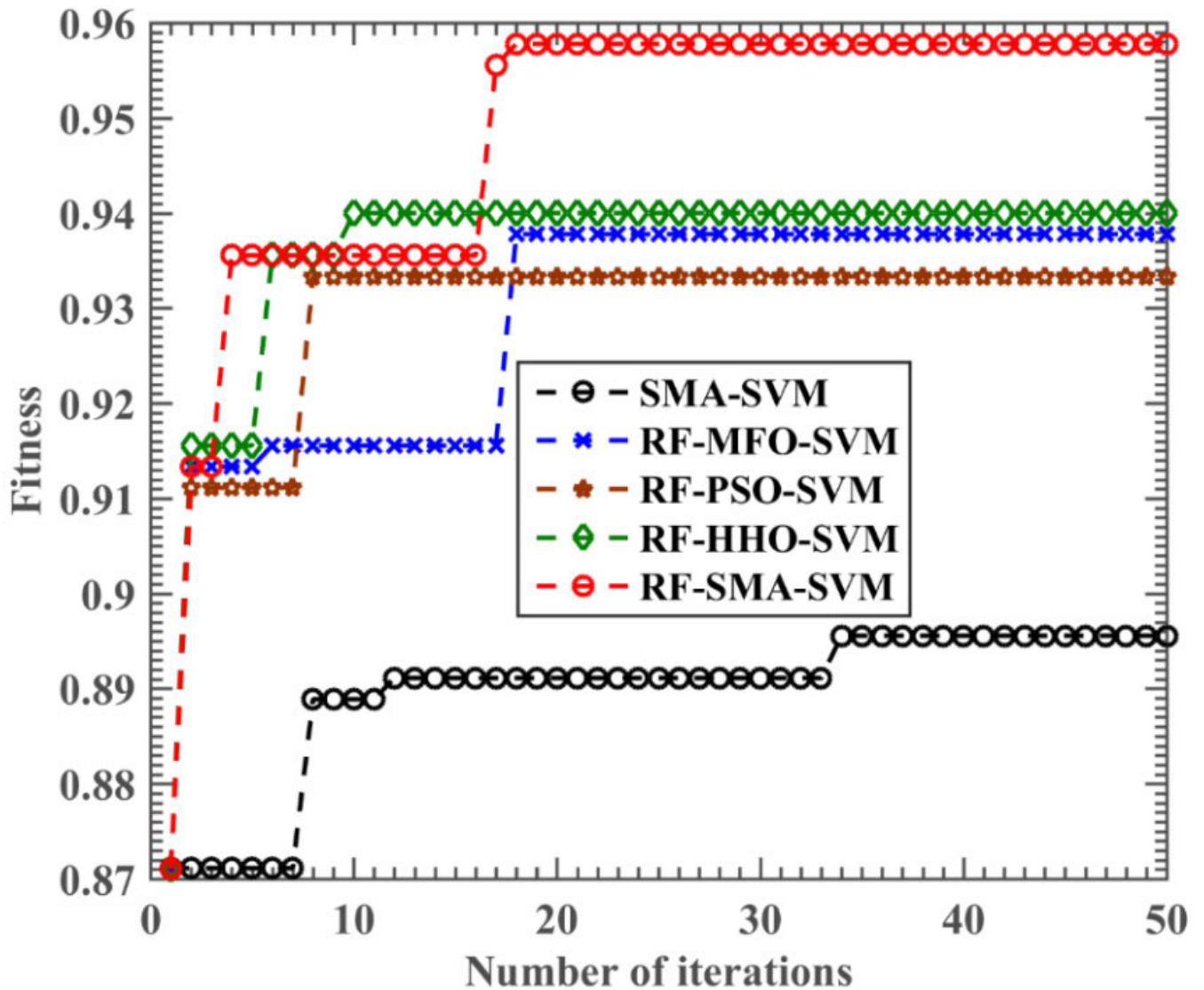


FIGURE 8. Relationship between training accuracy of several SVM variants and the number of iterations.

inflammation [49], [50]. Neutrophilic granulocytes are the first line to defense microbial pathogens, especially in pyogenic bacteria. Lymphocytes originate from the spleen, tonsils and lymph nodes are the smallest of the white blood cells, comprising 20% – 40% of the total leukocytes in blood. Lymphocytes are a necessary component of the immune response of human body, which is involved in the synthesis and distribution of antibodies in the blood [51]. The B-lymphocytes play a core role in humoral immunity or antibody immunity, and the T-lymphocytes are responsible for cellular immunity. A number of clinical observations have suggested that a reduced number of lymphocytes and an increased number of neutrophils commonly occur in COVID-19 patients [1], [52], [53]. As an example, a population-based study of more than 100 patients suggested that severe COVID-19 patients had higher neutrophil count

and lower lymphocyte count during the severe phase [54]. Liu and colleagues also found that lymphopenia and neutrophil count increased in patients with COVID-19 upon hospital admission may indicate more severe acute lung injury. Therefore, lymphopenia, neutrophil count increases may be a strong predictor of disease severity [55]. There is reason to believe that severe COVID-19 patients have immune disorders and excessive inflammation. Thus, patients with COVID-19 were rapidly screened with excessive inflammation using routine laboratory parameters in order to enhance survival rates. Hence, in recent years, N/L has been a simple parameter to reflect easily patient’s inflammatory [56]. Meanwhile, N/L was widely used as a simple and inexpensive marker for monitoring the infectious diseases, including patients with acute-on-chronic hepatitis B, pulmonary tuberculosis and bacterial community-acquired

pneumonia [57], [58]. Furthermore, N/L also has been considered to be a strong independent prognostic indicator for a variety of diseases including malignant tumors, cardiovascular diseases, sepsis, chronic obstructive pulmonary disease, and infectious disease [59]–[62]. Elevated N/L was associated with poor clinical prognosis [63]–[65]. Elevated N/L has been used as a marker in the determination of increased inflammation in acute exacerbations of chronic obstructive pulmonary disease, which is similar to C-reactive protein [66]. N/L seemed to be a safe, simple and promising marker in predicting bacteremia and in grading community-acquired pneumonia, and N/L was more accurate and objective than C-reactive protein [67]. A large population-based retrospective cohort study from London reported that more than 3,500 people living with human immunodeficiency virus cases proved elevated N/L was an independent predictor of the risk of cardiovascular disease rather than a risk traditional risk factor for cardiovascular disease [68]. In line with these findings, there are several studies describing the relationship between N/L and COVID-19 patients. Lagunas-Rangel *et al.* revealed that N/L was significantly higher in the severe COVID-19 patients compared with the non-severe group, and the elevation of N/L is closely associated with poor prognosis in COVID-19 patients [69]. Liu and colleagues also demonstrated that the incidence of severe patients aged 50 years or older and  $N/L \geq 3.13$  was up to 50%. Similar to the findings of their study, we also found that N/L was significantly higher in severe COVID-19 patients displaying about 2.9 times higher levels than in non-severe COVID-19 patients. Taken together, N/L was shown to be of significant predictive value for the discrimination of COVID-19 patients' status and might predict COVID-19 progression.

Platelets are viewed as important immune cells in the human body. Platelets have important physiological functions in the human body. Platelets have multiple physiological functions in the human body, including hemostasis, coagulation, the maintenance of vascular integrity, angiogenesis, innate immunity, inflammatory responses, viral infection, and malignant tumor. The number of platelets and platelet activity is strongly associated with multiple diseases [70]. Platelets are the smallest cell in peripheral blood cells, generating from megakaryocytes in the bone marrow [71]. In recent years, research has shown that platelets and lymphocytes play a key role in the inflammatory response process. P/L as a novel type of inflammation factor has attracted more and more attention from researchers. P/L could serve as an indicator of inflammation that may reflect the level of systemic inflammation in patients [72]. P/L serves a role in a wide spectrum of diseases, including acute ischemic stroke, urothelial carcinoma, acute pulmonary embolism and Pulmonary large cell neuroendocrine carcinoma [73]–[76]. Thus, P/L might be a reliable prognostic marker in the progression and prognosis of many diseases. Wang *et al.* carried out a retrospective study that included 134 adenocarcinoma cell lung cancer and found that high P/L was closely related

to shorter disease-free survival and lower overall survival rates [77]. Wang and colleagues revealed that elevated P/L was an independent predictor for poor prognosis of patients with hepatocellular carcinoma [78]. Wang *et al.* also reported that of the 695 lung cancer cases, lower P/L was significantly associated with a lower tumor-node-metastasis (TNM) stage and a low incidence rate of surgery [79]. Based on the results of previous research, there is a reason to believe the relationship between the P/L and COVID-19. Qu and colleagues revealed that high P/L was positively correlated with COVID-19 severity and hospital lengths of stay. They believed that P/L may provide a useful indicator for monitoring the patients with COVID-19. In this study, we found that P/L in severe COVID-19 patients was about 2.0 times higher compared to non-severe COVID-19 patients. Thus, P/L has the potential to be used as a predictive marker for COVID-19 disease severity. Until now, there are very few reports describing the cost-effective markers N/L and P/L to joint predict the severity and prognosis of the infectious disease. To the best of our knowledge, this is the first time that the combined of age, N/L and P/L for predicting the prognosis of patients with COVID-19 using a machine learning method.

However, there are some limitations in this study. First of all, our data set was based on a single-center and the sample size was not large enough. In order to improve diagnostic accuracy, we will enlarge the sample size in future research. Second, multi-center researches will be necessary to carry out external validation for the RF-SMA-SVM model. Third, we plan to include more blood indexes in our future research.

## V. CONCLUSION AND FUTURE WORKS

In this study, a validated RF-SMA-SVM model was developed to differentiate the severity of COVID-19 based on factors such as basic patient information and 26 blood routine indicators. The main innovations of this paper are: on the one hand, it is proposed for the first time that the severity of COVID-19 can be distinguished by using the combination of age, N/L and P/L indicators, and on the other hand, it is the first time that the SMA is used to train an optimal SVM classifier. Based on the experimental results, the proposed method exhibits higher prediction accuracy and more stable performance on the COVID-19 severity prediction problem than the SVM method based on other optimization algorithms, while screening out the key factors rich in discriminatory power.

In future work, we will attempt to apply the proposed RF-SMA-SVM model to the COVID-19 pre-diagnosis problem and then attempt to apply it to other disease prediction problems.

## ACKNOWLEDGMENT

(Huilin Chen, Xiaoying Huang, and Liangxing Wang contributed equally to this work.)

## REFERENCES

- [1] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, and X. Gu, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *Lancet*, vol. 395, pp. 497–506, 2020.

- [2] C.-C. Lai, T.-P. Shih, W.-C. Ko, H.-J. Tang, and P.-R. Hsueh, "Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges," *Int. J. Antimicrobial Agents*, vol. 55, no. 3, Mar. 2020, Art. no. 105924.
- [3] Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, "The species severe acute respiratory syndrome-related coronavirus: Classifying 2019-nCoV and naming it SARS-CoV-2," *Nature Microbiol.*, vol. 5, no. 4, pp. 536–544, Mar. 2020.
- [4] L. C. S. Afonso, G. H. Rosa, C. R. Pereira, S. A. T. Weber, C. Hook, V. H. C. Albuquerque, and J. P. Papa, "A recurrence plot-based approach for Parkinson's disease identification," *Future Gener. Comput. Syst.*, vol. 94, pp. 282–292, May 2019.
- [5] S. S. Tan et al., "Practical laboratory considerations amidst the COVID-19 outbreak: Early experience from Singapore," *J. Clin. Pathol.*, vol. 74, no. 4, p. 257, 2021.
- [6] H. Legido-Quigley, J. T. Mateos-García, V. R. Campos, M. Gea-Sánchez, C. Muntaner, and M. McKee, "The resilience of the Spanish health system against the COVID-19 pandemic," *Lancet Public Health*, vol. 5, no. 5, pp. e251–e252, May 2020.
- [7] N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, Y. Qiu, J. Wang, Y. Liu, Y. Wei, J. Xia, T. Yu, X. Zhang, and L. Zhang, "Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study," *Lancet*, vol. 395, no. 10223, pp. 507–513, Feb. 2020.
- [8] D. Benvenuto, M. Giovanetti, A. Ciccozzi, S. Spoto, S. Angeletti, and M. Ciccozzi, "The 2019-new coronavirus epidemic: Evidence for virus evolution," *J. Med. Virol.*, vol. 92, pp. 455–459, Apr. 2020.
- [9] S.-Q. Deng and H.-J. Peng, "Characteristics of and public health responses to the coronavirus disease 2019 outbreak in China," *J. Clin. Med.*, vol. 9, no. 2, p. 575, Feb. 2020.
- [10] X. Yang, Y. Yu, J. Xu, H. Shu, J. Xia, H. Liu, Y. Wu, L. Zhang, Z. Yu, M. Fang, T. Yu, Y. Wang, S. Pan, X. Zou, S. Yuan, and Y. Shang, "Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: A single-centered, retrospective, observational study," *Lancet Respiratory Med.*, vol. 8, no. 5, pp. 475–481, May 2020.
- [11] C.-C. Lai, T.-P. Shih, W.-C. Ko, H.-J. Tang, and P.-R. Hsueh, "Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges," *Int. J. Antimicrobial Agents*, vol. 55, no. 3, Mar. 2020, Art. no. 105924.
- [12] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, pp. 255–260, Jul. 2015.
- [13] R. C. Deo, "Machine learning in medicine," *Circulation*, vol. 132, pp. 1920–1930, Nov. 2015.
- [14] N. Kimura, Y. Aso, K. Yabuuchi, M. Ishibashi, D. Hori, Y. Sasaki, A. Nakamichi, S. Uesugi, H. Fujioka, S. Iwao, M. Jikumaru, T. Katayama, K. Sumi, A. Eguchi, S. Nonaka, M. Kakumu, and E. Maltsubara, "Modifiable lifestyle factors and cognitive function in older people: A cross-sectional observational study," *Frontiers Neurol.*, vol. 10, p. 401, Apr. 2019.
- [15] Z. Obermeyer and E. J. Emanuel, "Predicting the future—Big data, machine learning, and clinical medicine," *New England J. Med.*, vol. 375, pp. 1216–1219, Sep. 2016.
- [16] H. Ye, P. Wu, T. Zhu, Z. Xiao, X. Zhang, L. Zheng, R. Zheng, Y. Sun, W. Zhou, Q. Fu, X. Ye, A. Chen, S. Zheng, A. A. Heidari, M. Wang, J. Zhu, H. Chen, and J. Li, "Diagnosing coronavirus disease 2019 (COVID-19): Efficient Harris Hawks-inspired fuzzy K-nearest neighbor prediction methods," *IEEE Access*, vol. 9, pp. 17787–17802, 2021, doi: 10.1109/ACCESS.2021.3052835.
- [17] Y. Wang, M. Hu, Q. Li, X.-P. Zhang, G. Zhai, and N. Yao, "Abnormal respiratory patterns classifier may contribute to large-scale screening of people infected with COVID-19 in an accurate and unobtrusive manner," 2020, *arXiv:2002.05534*. [Online]. Available: <http://arxiv.org/abs/2002.05534>
- [18] Z. Hu, Q. Ge, S. Li, L. Jin, and M. Xiong, "Artificial intelligence forecasting of COVID-19 in China," 2020, *arXiv:2002.07112*. [Online]. Available: <http://arxiv.org/abs/2002.07112>
- [19] K. Wu et al., "Generalized logistic growth modeling of the COVID-19 outbreak: Comparing the dynamics in the 29 provinces in China and in the rest of the world," *Nonlinear Dyn.*, vol. 101, no. 3, pp. 1561–1581, 2020.
- [20] F. Shan, Y. Gao, J. Wang, W. Shi, N. Shi, M. Han, Z. Xue, D. Shen, and Y. Shi, "Lung infection quantification of COVID-19 in CT images with deep learning," 2020, *arXiv:2003.04655*. [Online]. Available: <http://arxiv.org/abs/2003.04655> and <https://ui.adsabs.harvard.edu/abs/2020arXiv200304655>
- [21] Z. Yang, Z. Zeng, K. Wang, S.-S. Wong, W. Liang, M. Zanin, P. Liu, X. Cao, Z. Gao, and Z. Mai, "Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions," *J. Thoracic Disease*, vol. 12, p. 165, 2020.
- [22] A. Vivanco-Lira, "Predicting COVID-19 distribution in Mexico through a discrete and time-dependent Markov chain and an SIR-like model," 2020, *arXiv:2003.06758*. [Online]. Available: <http://arxiv.org/abs/2003.06758>
- [23] O. Gozes, M. Frid-Adar, H. Greenspan, P. D. Browning, H. Zhang, W. Ji, A. Bernheim, and A. B. Siegel, "Rapid AI development cycle for the coronavirus (COVID-19) pandemic: Initial results for automated detection & patient monitoring using deep learning CT image analysis," 2020, *arXiv:2003.05037*. [Online]. Available: <https://arxiv.org/abs/2003.05037>
- [24] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [25] N. Zhang, "Semistability of steepest descent with momentum for quadratic functions," *Neural Comput.*, vol. 25, no. 5, pp. 1277–1301, May 2013.
- [26] N. Zhang, "A study on the optimal double parameters for steepest descent with momentum," *Neural Comput.*, vol. 27, no. 4, pp. 982–1004, Apr. 2015.
- [27] X. Zhang, D. Wang, Z. Zhou, and Y. Ma, "Robust low-rank tensor recovery with rectification and alignment," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 238–255, Jan. 2021.
- [28] S. Li, H. Chen, M. Wang, A. A. Heidari, and S. Mirjalili, "Slime mould algorithm: A new method for stochastic optimization," *Future Gener. Comput. Syst.*, vol. 111, pp. 300–323, Oct. 2020.
- [29] G. Liu, W. Jia, M. Wang, A. A. Heidari, H. Chen, Y. Luo, and C. Li, "Predicting cervical hyperextension injury: A covariance guided sine cosine support vector machine," *IEEE Access*, vol. 8, pp. 46895–46908, 2020.
- [30] J. Xu, X. Zhang, H. Chen, J. Li, J. Zhang, L. Shao, and G. Wang, "Automatic analysis of microaneurysms turnover to diagnose the progression of diabetic retinopathy," *IEEE Access*, vol. 6, pp. 9632–9642, 2018.
- [31] L. Hu, F. Lin, H. Li, C. Tong, Z. Pan, J. Li, and H. Chen, "An intelligent prognostic system for analyzing patients with paraquat poisoning using arterial blood gas indexes," *J. Pharmacol. Toxicol. Methods*, vol. 84, pp. 78–85, Mar. 2017.
- [32] H. Chen, L. Hu, H. Li, G. Hong, T. Zhang, J. Ma, and Z. Lu, "An effective machine learning approach for prognosis of paraquat poisoning patients using blood routine indexes," *Basic Clin. Pharmacol. Toxicol.*, vol. 120, no. 1, pp. 86–96, Jan. 2017.
- [33] Z. Cai, J. Gu, and H.-L. Chen, "A new hybrid intelligent framework for predicting Parkinson's disease," *IEEE Access*, vol. 5, pp. 17188–17200, 2017.
- [34] D. Wang, H. Qiao, B. Zhang, and M. Wang, "Online support vector machine based on convex hull vertices selection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 4, pp. 593–609, Apr. 2013.
- [35] Q. Zhang, D. Wang, and Y. Wang, "Convergence of decomposition methods for support vector machines," *Neurocomputing*, vol. 317, pp. 179–187, Nov. 2018.
- [36] N. Gu, D. Wang, M. Fan, and D. Meng, "A kernel-based sparsity preserving method for semi-supervised classification," *Neurocomputing*, vol. 139, pp. 345–356, Sep. 2014.
- [37] M. S. Majumder, S. A. Klumberg, S. R. Mekar, and J. S. Brownstein, "Mortality risk factors for middle east respiratory syndrome outbreak, South Korea, 2015," *Emerg. Infectious Diseases*, vol. 21, pp. 2088–2090, Nov. 2015.
- [38] K. Hong, J. Choi, S. Hong, J. Lee, J. Kwon, and S. Kim, "Predictors of mortality in middle east respiratory syndrome (MERS)," *Thorax*, vol. 73, pp. 286–289, Mar. 2018.
- [39] K. W. Choi, T. N. Chau, O. Tsang, E. Tso, M. C. Chiu, W. L. Tong, P. O. Lee, T. K. Ng, W. F. Ng, K. C. Lee, W. Lam, W. C. Yu, J. Y. Lai, and S. T. Lai, "Outcomes and prognostic factors in 267 patients with severe acute respiratory syndrome in Hong Kong," *Ann. Int. Med.*, vol. 139, pp. 715–723, Nov. 2003.
- [40] S. L. Smits, A. de Lang, J. M. A. van den Brand, L. M. Leijten, W. F. van IJcken, M. J. C. Eijkemans, G. van Amerongen, T. Kuiken, A. C. Andeweg, A. D. M. E. Osterhaus, and B. L. Haagmans, "Exacerbated innate host response to SARS-CoV in aged non-human primates," *PLoS Pathogens*, vol. 6, no. 2, Feb. 2010, Art. no. e1000756.

- [41] F. Zhou, T. Yu, R. Du, G. Fan, Y. Liu, Z. Liu, J. Xiang, Y. Wang, B. Song, X. Gu, L. Guan, Y. Wei, H. Li, X. Wu, J. Xu, S. Tu, Y. Zhang, H. Chen, and B. Cao, "Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: A retrospective cohort study," *Lancet*, vol. 395, no. 10229, pp. 1054–1062, Mar. 2020.
- [42] D. Pittet, B. Thievent, R. P. Wenzel, N. Li, G. Gurman, and P. M. Suter, "Importance of pre-existing co-morbidities for prognosis of septicemia in critically ill patients," *Intensive Care Med.*, vol. 19, pp. 265–272, May 1993.
- [43] G. S. Martin, D. M. Mannino, S. Eaton, and M. Moss, "The epidemiology of sepsis in the United States from 1979 through 2000," *New England J. Med.*, vol. 348, pp. 1546–1554, Apr. 2003.
- [44] M. E. Weksler, "Changes in the B-cell repertoire with age," *Vaccine*, vol. 18, no. 16, pp. 1624–1628, Feb. 2000.
- [45] D. Weiskopf, B. Weinberger, and B. Grubeck-Loebenstein, "The aging of the immune system," *Transplant Int.*, vol. 22, pp. 1041–1050, Nov. 2009.
- [46] S. M. Opal, T. D. Girard, and E. W. Ely, "The immunopathogenesis of sepsis in elderly patients," *Clin. Infectious Diseases*, vol. 41, no. 7, pp. S504–S512, Nov. 2005.
- [47] A. P. Reiner, G. Lettre, M. A. Nalls, S. K. Ganesh, R. Mathias, M. A. Austin, E. Dean, S. Arepalli, A. Britton, and Z. Chen, "Genome-wide association study of white blood cell count in 16,388 African Americans: The continental origins and genetic epidemiology network (COGENT)," *PLoS Genet.*, vol. 7, 2011, Art. no. e1002108.
- [48] R. Kyeyune, E. Saathoff, A. E. Ezeamama, T. Löscher, W. Fawzi, and D. Guwatudde, "Prevalence and correlates of cytopenias in HIV-infected adults initiating highly active antiretroviral therapy in Uganda," *BMC Infectious Diseases*, vol. 14, no. 1, p. 496, Sep. 2014.
- [49] S. Z. Wang, P. Z. Smith, M. Lovejoy, J. J. Bowden, J. H. Alpers, and K. D. Forsyth, "The apoptosis of neutrophils is accelerated in respiratory syncytial virus (RSV)-induced bronchiolitis," *Clin. Exp. Immunol.*, vol. 114, no. 1, pp. 49–54, Oct. 1998.
- [50] L. Schreiner, M. Huber-Lang, M. E. Weiss, H. Hohmann, M. Schmolz, and E. M. Schneider, "Phagocytosis and digestion of pH-sensitive fluorescent dye (Eos-FP) transfected E. Coli in whole blood assays from patients with severe sepsis and septic shock," *J. Cell Commun. Signaling*, vol. 5, pp. 135–144, Jun. 2011.
- [51] X. Wang, J. Tian, K.-T. Yong, X. Zhu, M. C.-M. Lin, W. Jiang, J. Li, Q. Huang, and G. Lin, "Immunotoxicity assessment of CdSe/ZnS quantum dots in macrophages, lymphocytes and BALB/c mice," *J. Nanobiotechnol.*, vol. 14, no. 1, p. 10, Feb. 2016.
- [52] Z. Wang, B. Yang, Q. Li, L. Wen, and R. Zhang, "Clinical features of 69 cases with coronavirus disease 2019 in Wuhan, China," *Clin. Infectious Diseases*, vol. 71, no. 15, pp. 769–777, Mar. 2020.
- [53] J. J. Zhang, X. Dong, Y.-Y. Cao, Y.-D. Yuan, Y.-B. Yang, Y.-Q. Yan, C. A. Akdis, and Y.-D. Gao, "Clinical characteristics of 140 patients infected with SARS-CoV-2 in Wuhan, China," *Allergy*, vol. 75, no. 7, pp. 1730–1741, Feb. 2020.
- [54] D. Wang, B. Hu, C. Hu, F. Zhu, X. Liu, J. Zhang, B. Wang, H. Xiang, Z. Cheng, Y. Xiong, Y. Zhao, Y. Li, X. Wang, and Z. Peng, "Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China," *J. Amer. Med. Assoc.*, vol. 323, no. 11, p. 1061, Mar. 2020.
- [55] Y. Liu, Y. Yang, C. Zhang, F. Huang, F. Wang, J. Yuan, Z. Wang, J. Li, J. Li, and C. Feng, "Clinical and biochemical indexes from 2019-nCoV infected patients linked to viral loads and lung injury," *Sci. China Life Sci.*, vol. 63, pp. 364–374, 2020.
- [56] M. Sikora, A. Stec, M. Chrabaszcz, A. Waskiel-Burnat, M. Zaremba, M. Olszewska, and L. Rudnicka, "Intestinal fatty acid binding protein, a biomarker of intestinal barrier, is associated with severity of psoriasis," *J. Clin. Med.*, vol. 8, no. 7, p. 1021, Jul. 2019.
- [57] M. Berhane *et al.*, "The role of neutrophil to lymphocyte count ratio in the differential diagnosis of pulmonary tuberculosis and bacterial community-acquired pneumonia: A cross-sectional study at Ayder and Mekelle hospitals, Ethiopia," *Clin. Lab.*, vol. 65, no. 4, 2019, doi: 10.7754/clin.lab.2018.180833.
- [58] H. Liu, H. Zhang, G. Wan, Y. Sang, Y. Chang, X. Wang, and H. Zeng, "Neutrophil-lymphocyte ratio: A novel predictor for short-term prognosis in acute-on-chronic hepatitis B liver failure," *J. Viral Hepatitis*, vol. 21, no. 7, pp. 499–507, Jul. 2014.
- [59] F. Riché, E. Gayat, R. Barthélémy, M. Le Dorze, J. Matéo, and D. Payen, "Reversal of neutrophil-to-lymphocyte count ratio in early versus late death from septic shock," *Crit. Care*, vol. 19, no. 1, p. 439, Dec. 2015.
- [60] M. E. Afari and T. Bhat, "Neutrophil to lymphocyte ratio (NLR) and cardiovascular diseases: An update," *Expert Rev. Cardiovascular Therapy*, vol. 14, no. 5, pp. 573–577, 2016.
- [61] S. S. Faria, P. C. Fernandes, Jr., M. J. B. Silva, V. C. Lima, W. Fontes, R. Freitas-Junior, A. K. Eterovic, and P. Forget, "The neutrophil-to-lymphocyte ratio: A narrative review," *Ecancermedicalscience*, vol. 10, p. 702, Dec. 2016.
- [62] D. Inoue, S. Sekiguchi, W. Yamagata, G. Maeda, D. Yamada, S. Fujiwara, S. Itou, M. Kurihara, Y. Hijioka, K. Shimoji, J. Fujiki, A. Nakazono, H. Horike, A. Yoshioka, Y. Ogura, F. Hatao, K. Imamura, and S. Namiki, "Elevation of neutrophil-to-lymphocyte ratio before first-line chemotherapy predicts a poor prognosis for second-line chemotherapy in gastric cancer," *Oncology*, vol. 96, no. 3, pp. 140–146, 2019.
- [63] X.-Q. Chen, C.-R. Xue, P. Hou, B.-Q. Lin, and J.-R. Zhang, "Lymphocyte-to-monocyte ratio effectively predicts survival outcome of patients with obstructive colorectal cancer," *World J. Gastroenterol.*, vol. 25, no. 33, pp. 4970–4984, Sep. 2019.
- [64] H. F. Zhang *et al.*, "Neutrophil-to-lymphocyte ratio improves the accuracy and sensitivity of pneumonia severity index in predicting 30-day mortality of CAP patients," *Clin. Lab.*, vol. 65, no. 10, 2019, doi: 10.7754/Clin.Lab.2019.190226.
- [65] Y. L. Ge *et al.*, "Neutrophil-to-lymphocyte ratio in adult community-acquired pneumonia patients correlates with unfavorable clinical outcomes," *Clin. Lab.*, vol. 65, no. 5, 2019, doi: 10.7754/clin.lab.2018.181042.
- [66] R. Farah, R. Ibrahim, M. Nassar, D. Najib, Y. Zivony, and E. Eshel, "The neutrophil/lymphocyte ratio is a better addition to C-reactive protein than CD64 index as a marker for infection in COPD," *Panminerva Med.*, vol. 59, pp. 203–209, Sep. 2017.
- [67] R. Furutate, T. Ishii, T. Motegi, K. Hattori, Y. Kusunoki, A. Gemma, and K. Kida, "The neutrophil to lymphocyte ratio is related to disease severity and exacerbation in patients with chronic obstructive pulmonary disease," *Internal Med.*, vol. 55, no. 3, pp. 223–229, 2016.
- [68] R. C. Merriman, O. Dissanayake, S. Alnjar, F. Burns, and R. F. Miller, "Incidence and significance of elevated platelet-to-lymphocyte and neutrophil-to-lymphocyte ratios among hospitalised HIV-positive adult patients," *Int. J. STD AIDS*, vol. 30, no. 13, pp. 1329–1332, Nov. 2019.
- [69] F. A. Lagunas-Rangel, "Neutrophil-to-lymphocyte ratio and lymphocyte-to-C-reactive protein ratio in patients with severe coronavirus disease 2019 (COVID-19): A meta-analysis," *J. Med. Virol.*, vol. 92, no. 10, pp. 1733–1734, Apr. 2020.
- [70] C. N. Jenne and P. Kubes, "Platelets in inflammation and infection," *Platelets*, vol. 26, no. 4, pp. 286–292, 2015.
- [71] W. J. Liu, J. Bai, Q. L. Guo, Z. Huang, H. Yang, and Y. Q. Bai, "Role of platelet function and platelet membrane glycoproteins in children with primary immune thrombocytopenia," *Mol. Med. Rep.*, vol. 14, pp. 2052–2060, Sep. 2016.
- [72] L. Kutlucan, A. Kutlucan, B. Basaran, M. Dagli, A. Basturk, B. Kozanhan, M. Gur, E. Senocak, and M. Kos, "The predictive effect of initial complete blood count of intensive care unit patients on mortality, length of hospitalization, and nosocomial infections," *Eur. Rev. Med. Pharmacol. Sci.*, vol. 20, pp. 1467–1473, Apr. 2016.
- [73] J.-H. Xu, X.-W. He, Q. Li, J.-R. Liu, M.-T. Zhuang, F.-F. Huang, and G.-S. Bao, "Higher platelet-to-lymphocyte ratio is associated with worse outcomes after intravenous thrombolysis in acute ischaemic stroke," *Frontiers Neurol.*, vol. 10, p. 1192, Nov. 2019.
- [74] Y. Bao, Y. Wang, X. Li, M. Pan, H. Zhang, Z. Cheng, and X. Wang, "Prognostic significance of platelet-to-lymphocyte ratio in urothelial carcinoma patients: A meta-analysis," *Cancer Cell Int.*, vol. 19, no. 1, p. 315, Dec. 2019.
- [75] T. Phan, Y. Brailovsky, J. Fareed, D. Hoppensteadt, O. Iqbal, and A. Darki, "Neutrophil-to-lymphocyte and platelet-to-lymphocyte ratios predict all-cause mortality in acute pulmonary embolism," *Clin. Appl. Thrombosis/Hemostasis*, vol. 26, pp. 1–7, Jan./Dec. 2020.
- [76] M. Shi, W. Zhao, F. Zhou, H. Chen, L. Tang, B. Su, and J. Zhang, "Neutrophil or platelet-to-lymphocyte ratios in blood are associated with poor prognosis of pulmonary large cell neuroendocrine carcinoma," *Transl. Lung Cancer Res.*, vol. 9, no. 1, pp. 45–54, Feb. 2020.
- [77] Y.-Q. Wang, Q.-J. Zhi, X.-Y. Wang, D.-S. Yue, K. Li, and R.-C. Jiang, "Prognostic value of combined platelet, fibrinogen, neutrophil to lymphocyte ratio and platelet to lymphocyte ratio in patients with lung adenocarcinoma," *Oncol. Lett.*, vol. 14, no. 4, pp. 4331–4338, Oct. 2017.



- [78] Y. Wang, B. M. Attar, H. E. Fuentes, P. Jaiswal, and A. J. Tafur, "Evaluation of the prognostic value of platelet to lymphocyte ratio in patients with hepatocellular carcinoma," *J. Gastrointestinal Oncol.*, vol. 8, no. 6, pp. 1065–1071, Dec. 2017.
- [79] L. Wang, D. Liang, X. Xu, J. Jin, S. Li, G. Tian, Z. Gao, C. Liu, and Y. He, "The prognostic value of neutrophil to lymphocyte and platelet to lymphocyte ratios for patients with lung cancer," *Oncol. Lett.*, vol. 14, pp. 6449–6456, Sep. 2017.



**PEILIANG WU** received the M.S. degree in medical science from Wenzhou Medical University, China, in 2017, where he is currently pursuing the Ph.D. degree in respiratory medicine. He is also a Resident Physician with the Pulmonary and Critical Care Medicine, The First Affiliated Hospital of Wenzhou Medical University, China. His main research interests include respiratory infectious diseases, lung cancer, and pulmonary arterial hypertension.



**HUA YE** received the M.S. degree in medical science from Wenzhou Medical University, China, in 2017. She is currently a Chief Physician of the Department of Pulmonary and Critical Care Medicine, Affiliated Yueqing Hospital, Wenzhou Medical University. Her main research interests include respiratory infectious diseases and lung tumors.



**XUEDING CAI** received the M.D. degree from the Department of Respiratory Medicine, Wenzhou Medical University, China. He is currently a Doctor with the Department of Pulmonary and Critical Care Medicine, The First Affiliated Hospital of Wenzhou Medical University. He has published more than ten articles in international journals. His research interests include respiratory medicine and artificial intelligence with application in the field of medicine.



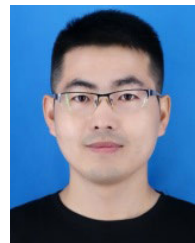
**CHENGYE LI** received the M.D. degree from the School of Medicine, Shanghai Jiaotong University, China. He is currently an Associate Chief Physician with the Department of Pulmonary and Critical Care Medicine, The First Affiliated Hospital of Wenzhou Medical University, China. His research interests include data mining and machine learning, as well as their applications to medical diagnosis.



**SHIMIN LI** is currently a Student with the College of Computer Science and Artificial Intelligence, Wenzhou University, China. He has published articles in journals, such as *Expert Systems with Applications* and *Future Generation Computer Systems*. His research interests include machine learning and data mining.



**MENGXIANG CHEN** received the master's degree from Hong Kong University of Science and Technology, China. She is currently a Teaching Assistant with the Department of Information Technology, Wenzhou Vocational College of Science and Technology. Her current research interests include higher education. She undertook several national research projects in the area of education technology and entrepreneurship education, and guided students to win first prize and second prize in fierce provincial competitions in innovation and entrepreneurship time after time, such as the Challenge Cup, the E-Commerce Ideal Competition.



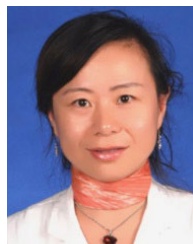
**MINGJING WANG** is currently a Special Researcher with the Institute of Research and Development, Duy Tan University, Da Nang, Vietnam. He has published several articles in the field of computer engineering in top-ranking journals, such as *Neurocomputing*, *Engineering Applications of Artificial Intelligence*, *Energy Conversion and Management*, *Applied Mathematical Modelling*, and *Applied Soft Computing*. His research interests include data mining, machine learning, evolutionary computation and their applications to medical diagnosis.



**ALI ASGHAR HEIDARI** is currently a Research Intern with the School of Computing, National University of Singapore (NUS). He is also an Exceptionally Talented Researcher with the University of Tehran. He is awarded and funded by Iran's National Elites Foundation (INEF). He has published more than 40 articles with more than 1.3K citations in prestigious international journals, such as *Information Fusion*, *Information Sciences*, *Future Generation Computer Systems*, the *Journal of Cleaner Production*, *Energy Conversion and Management*, *Energy*, *Applied Soft Computing*, *Knowledge-Based Systems*, and *Expert Systems with Applications*. His main research interests include advanced machine learning, evolutionary computation, meta-heuristics, prediction, information systems, and spatial modeling.



**MAYUN CHEN** received the M.D. and Ph.D. degrees from Wenzhou Medical University, China. He is currently a Lecturer and an Attending Physician with the Department of Pulmonary and Critical Care Medicine, The First Affiliated Hospital of Wenzhou Medical University, China. His research interests include intelligent medical image analysis, data mining and machine learning, and their applications to medical diagnosis.



**XIAOYING HUANG** received the Ph.D. degree from the Department of Clinical Discipline of Chinese and Western Integrative Medicine, Zhejiang Chinese Medical University, China. She is currently a Professor with The First Clinical Medical College of Wenzhou Medical University, China. Her current research interests include respiratory infectious diseases, lung cancer, and pulmonary arterial hypertension.

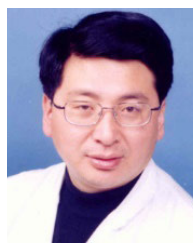


**JIFA LI** is currently the Chief Physician of the Department of Pulmonary and Critical Care Medicine, Affiliated Yueqing Hospital, Wenzhou Medical University. His research interests include respiratory chronic interstitial diseases, respiratory tract infectious diseases, and lung tumors.



**HUILIN CHEN** (Associate Member, IEEE) received the Ph.D. degree from the Department of Computer Science and Technology, Jilin University, China. He is currently an Associate Professor with the College of Computer Science and Artificial Intelligence, Wenzhou University, China. He has published more than 100 articles in international journals and conference proceedings, including the IEEE SYSTEMS JOURNAL, *Information Sciences*, *Pattern Recognition*, *Future Generation*

*Computer System*, *Expert Systems with Applications*, *Knowledge-based Systems*, *Neurocomputing*, *PAKDD*, and among others. His research interests include machine learning and data mining, as well as their applications to medical diagnosis and bankruptcy prediction. He is also serving as an Associate Editor for IEEE ACCESS. He is also a Reviewer for many journals, such as *Applied Soft Computing*, *Artificial Intelligence in Medicine*, *Knowledge-based Systems*, and *Future Generation Computer System*.



**LIANGXING WANG** received the Ph.D. degree from the Department of Respiratory Medicine, Huazhong University of Science and Technology, China. He is currently a Professor with The First Clinical Medical College of Wenzhou Medical University, China. His research interests include respiratory infectious diseases, lung cancer, and pulmonary arterial hypertension.

...