Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Computers in Biology and Medicine 146 (2022) 105615

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

ISW-LM: An intensive symptom weight learning mechanism for early COVID-19 diagnosis

Lingling Fang *, Xiyue Liang

Department of Computing and Information Technology, Liaoning Normal University, Dalian City, Liaoning Province, China

ARTICLE INFO

Keywords:
COVID-19
Early diagnosis
Symptom weight
Learning mechanism
Machine learning models
Classification

ABSTRACT

The novel coronavirus disease 2019 (COVID-19) pandemic has severely impacted the world. The early diagnosis of COVID-19 and self-isolation can help curb the spread of the virus. Besides, a simple and accurate diagnostic method can help in making rapid decisions for the treatment and isolation of patients. The analysis of patient characteristics, case trajectory, comorbidities, symptoms, diagnosis, and outcomes will be performed in the model. In this paper, a symptom-based machine learning (ML) model with a new learning mechanism called Intensive Symptom Weight Learning Mechanism (ISW-LM) is proposed. The proposed model designs three new symptoms' weight functions to identify the most relevant symptoms used to diagnose and classify COVID-19. To verify the efficiency of the proposed model, multiple laboratory and clinical datasets containing epidemiological symptoms and blood tests are used. Experiments indicate that the importance of COVID-19 infection symptoms varies between countries and regions. In most datasets, the most frequent and significant predictive symptoms for diagnosing COVID-19 are fever, sore throat, and cough. The experiment also compares the state-of-the-art methods with the proposed method, which shows that the proposed model has a high accuracy rate of up to 97.1711%. The positive results indicate that the proposed learning mechanism can help clinicians quickly diagnose and screen patients for COVID-19 at an early stage.

1. Introduction

In December 2019, the first case of pneumonia of unknown origin was detected, which was subsequently discovered to be caused by severe acute respiratory syndrome coronavirus type 2(SARS-COV-2), named novel coronavirus disease (COVID-19) [1,2]. Although the treatment of COVID-19 patients has matured since the beginning of the outbreak, it cannot fundamentally contain the epidemic. There is an urgent need for the early prevention, screening, and diagnosis of suspected positive patients to control the spread of the disease [3]. Therefore, identifying the means to classify and diagnose suspected patients based on early examination results has become a problem worthy of investigation. Additionally, ensuring the effective control of symptom deterioration is also an urgent problem requiring a solution [4,5].

Many researchers have contributed information on how to diagnose positive cases and how to predict the course of the COVID-19 pandemic [6]. Building on the development of modern artificial intelligence (AI) and ML methods, models and technologies for coping with the COVID-19 pandemic are used to address the challenges during the outbreak. ML and AI have recently been employed to tackle the SARS-CoV-2 outbreak, SARS-CoV-2 screening and treatment, SARS-CoV-2 contact tracing, SARS-CoV-2 prediction and forecasting, SARS-CoV-2 drugs and vaccination, and other research directions [7]. The establishment of diagnostic models and techniques for COVID-19 is critical. Traditional techniques have been developed to assist doctors in making a correct diagnosis. In general, COVID-19 diagnosis can be categorized into three approaches: supervised learning approaches, unsupervised learning approaches, and hybrid approaches [8].

The common symptoms of COVID-19 patients appear approximately 1–2 weeks after exposure, including the onset of a cough, fever, general malaise, and shortness of breath [9,10]. Patients with early infection may not show significant symptoms after COVID-19 infection, and their symptoms are similar to the cold or the flu, which makes it difficult to accurately diagnose these patients [11,12]. Therefore, early detection and diagnosis using ML can help prevent and combat the COVID-19 pandemic by leveraging diverse epidemiological data [13,14]. To improve the early diagnostic capabilities of COVID-19, many methods based on symptom-based ML models have been proposed and studied

* Corresponding author.
E-mail address: fanglingling@lnnu.edu.cn (L. Fang).

https://doi.org/10.1016/j.compbiomed.2022.105615
Received 22 March 2022; Received in revised form 9 May 2022; Accepted 11 May 2022
Available online 17 May 2022

[15].

In the early stage of the pandemic, most studies were based on small datasets with fewer patients and symptoms. Besides, most of the laboratory COVID-19 data sets were used for testing. Davide Brinati et al. developed two machine learning classification models using histochemical values from routine blood and reverse transcription-polymerase chain reaction (RT–PCR) tests performed on respiratory tract specimens to discriminate between patients who are either positive or negative for SARS-CoV-2 [16]. The study included blood test results from 279 patients with symptoms of COVID-19. Of these patients, 177 had COVID-19, and 102 did not. Thomas Tschoellitsch, MD et al. used a random forest (RF) algorithm to predict a diagnosis based on laboratory blood tests with 28 unique characteristics. The reliability of the proposed method was verified by comparing it with real RT–PCR tests [17].

In general, blood and RT–PCR tests are expensive, and sometimes they may take a long time to produce results. To solve this problem, some researchers have proposed a simpler diagnostic model based on laboratory epidemiological symptoms. Rachid Zagrouba et al. presented a predictive framework incorporating support vector machine (SVM) in the forecasting of a potential outbreak of COVID-19, which can be used to predict the long-term spread of such an outbreak so that doctors can implement proactive measures in advance [18]. On this basis, Mahdi Mahdavi et al. proposed three SVM models to detect the invasive laboratory and noninvasive clinical and demographic data of COVID-19 patients at admission, which can decrease mortality by assuring efficient resource allocation and treatment planning during a pandemic [19]. In addition, Ahmed Hamed et al. proposed a novel K-nearest neighbor (K-NN) variant algorithm called K-NNV and handled incomplete heterogeneous symptom data for different diseases to achieve accurate classification of COVID-19 [20]. However, the above models are only used to classify COVID-19, without explicitly distinguishing it from other diseases. Matjaž Kukar et al. constructed a machine model for COVID-19 diagnosis using routine blood tests in 5333 patients with various bacterial and viral infections. The proposed model confirmed the five most useful routine blood parameters for COVID-19 diagnosis [21]. However, the studies obtained little symptom information from patients in the early stage of epidemic development and the reliability of the models still needs to be confirmed.

As researchers learn more about the virus and the pandemic, more patient information becomes available. Several large-scale laboratory COVID-19 datasets are also used. Buvana M and Muthumayil K explored COVID-19 datasets from the repository. Here, symptoms such as fever, body pain, runny nose, difficulty in breathing, sore throat, and nasal congestion were confirmed as the most important parameters with which to diagnose patients [22]. Warda M. Shaban et al. introduced a new detection strategy for COVID-19 infection called Distance Biased Naïve Bayes (DBNB). The researchers combined a new feature selection technique to identify the most informative and significant symptoms for diagnosing COVID-19 patients from laboratory datasets, which can quickly and accurately detect infected patients [23]. Prabh Deep Singh et al. designed and developed a novel aggregation-based classifier to predict COVID-19 cases at an early stage [24]. On this basis, Mohsin Sarker Raihan, MD et al. leveraged the concept of the COVID-19 blood test and proposed a risk-free model to identify COVID-19 patients in the blood test dataset [25].

At the beginning of the epidemic, the patients' symptoms entries in laboratory datasets were simple, such as age, gender, history of fever monitoring, and travel [26]. These were not adequate for monitoring the clinical situation. The above models have achieved accurate results when tested against laboratory COVID-19 symptom datasets. To make a more accurate diagnosis, many datasets with clinical symptoms have been studied. Nan-Nan Sun et al. proposed a prediction model based on ML for the early diagnosis of COVID-19, which aims to extract risk factors from the clinical data of patients. They also test the applicability of the model in actual clinical data and improved the accuracy and timeliness of the early diagnosis of COVID-19 infection [27]. Jiangpeng Wu et al. used the RF algorithm to extract 11 key blood indicators from the data of 49 clinically available blood tests and established the final auxiliary discriminant tool for preliminary evaluation of suspected patients, helping to obtain timely treatment and quarantine suggestions [28].

However, some studies have shown that a single diagnostic model may produce errors in the face of complex clinical situations. To collect more patient symptoms data, several studies are devoted to developing new stacked models for diagnosis to improve accuracy. Three different supervised ML techniques are used to diagnose COVID-19, such as, the bagging algorithm, K-NN, and RF to classify COVID-19 data sets datasets [29]. The symptoms are captured from COVID-19 trackers in India to evaluate model performance. However, some traditional ML models still face some limitations in determining the selection of COVID-19 symptoms. Therefore, some new classifiers are considered to assist in diagnosis. Ibrahim Arpaci et al. developed six COVID-19 diagnostic prediction models to identify positive and negative cases, including BayesNet, logistic, lazy-classifier (IBk), classification via regression (CR), rule-learner (PART), and decision-tree (J48) classifiers. The clinical dataset used was from the Taizhou Hospital of Zhejiang Province in China and contained 14 features [30]. Marcos Antonio Alves et al. presented understandable solutions based on ML techniques to deal with COVID-19 screening in routine blood tests. The sample consisted of 84 COVID-19 patients along with 608 other patients [31]. Lucas M. Thimoteo et al. proposed an interpretable artificial intelligence approach that includes two black-box models to help diagnose COVID-19 patients based on blood tests and pathogen variables [32].

However, the clinical symptoms of COVID-19 patients collected by the above models in the early stage of the pandemic are not enough to reflect the generalization of diagnostic models. To better fit the clinical setting, many studies have begun to target large-scale clinical data. Martuza Ahamad, MD et al. employed the supervised ML algorithms to identify the presentation features predicting COVID-19 disease diagnoses with high accuracy [33]. Dan Assaf et al. used three different ML models to predict patient deterioration. In this study, the selected parameters were the Acute Physiology And Chronic Health Evaluation II (APACHE II) score, white blood cell count, time from symptoms to admission, oxygen saturation, and blood lymphocyte count [34]. Maryam AlJame et al. proposed an ensemble learning model for diagnosing COVID-19 from routine blood tests, which exploits the strength of several diverse classifiers to improve the accuracy of the prediction and evaluates the importance of each feature [35]. Generally, blood tests usually take time to obtain, which slows the down subsequent analysis of the virus.

To solve this problem, L. J. Muhammad et al. developed a supervised ML model for COVID-19 positive and negative cases in Mexico using epidemiological marker datasets. The proposed method also obtains the correlation efficiency analysis between various dependent and independent features [36]. Similarly, Sakifa Aktar et al. further identified the most important symptoms and comorbidities that predict COVID-19 infection using six clinically applicable supervised ML algorithms. Pneumonia–Hypertension, Pneumonia–Diabetes, and acute respiratory distress syndrome (ARDS)–Hypertension show the most significant associations with COVID-19 mortality [37]. The above models can speed up the classification for potentially infected patients and determine the impact on the COVID-19 patients [38].

As the pandemic spreads and infection numbers soar in many countries and regions, some studies are increasingly incorporating larger and more realistic datasets to ensure accurate diagnosis and to control the spread of COVID-19. In the study by Suma L. S. et al. [39], an ML model was developed to analyze a clinical dataset containing 65,000 patient records, including 26 features, and to select the optimal subset of features needed for in COVID-19 patient screening. Krishnaraj Chadaga et al. proposed an automated framework that combines four different classifiers along with a technique called the synthetic minority oversampling technique (SMOTE) for distinguishing COVID-19 infection and

used the Shapley additive explanations (SHAP) method to calculate the gravity of each blood parameters feature [40]. In a study by Krishnaraj Chadaga et al. [41], combined multiple machine learning methods to diagnose and predict COVID-19 through routine blood tests. The experiment uses a dataset from the Israelita Albert Einstein Hospital, in Brazil. Large clinical datasets provide a large amount of patients' symptom information, but most of the classification models are tested in specific regions. The symptoms of COVID-19 infection vary by country and region. Although some symptom-based ML methods have been proposed, most of them are applied to specific datasets and cannot be applied to various situations [42].

To overcome these limitations, this paper proposes an intensive symptom weight learning mechanism, called ISW-LM, for a variety of situations using the intensive importance of symptoms to classify and diagnose early COVID-19. A new symptom weight calculation method is designed to rank the importance of symptoms. It also lists the order of intensive symptoms that can help in the early diagnosis of COVID-19. To verify the proposed model, many types of datasets are used for experiments, such as, small and large COVID-19 datasets from laboratories and clinically settings. The important symptoms in the data that help diagnose such data in patients with new coronavirus infection are listed. Several symptoms that may aggravate the infection in patients with comorbidities were also analyzed. Compared with existing techniques, the proposed model expands the application range in COVID-19 diagnosis. Furthermore, it also provides a rationale for further treatment and resource allocation.

The remainder of the paper is structured into multiple sections. The proposed method and the datasets are detailed in Section 2, which describes an intensive symptom weight learning mechanism for early COVID-19 diagnosis and the multiple datasets used in the paper. The experimental results are discussed in Section 3. Patient symptom datasets of different sizes from laboratories and clinical hospitals are used to verify the proposed model. Finally, concluding remarks and highlighting of future work are presented in Section 4.

2. Materials and methods

2.1. Datasets

2.1.1. Datasets description

The COVID-19 datasets [43–47] from open research datasets were used for research and analysis in this study. The experimental datasets are classified into small and large datasets by size. Moreover, datasets were divided into laboratory and clinical datasets based on their source. The corresponding classification chart for the datasets is shown in Fig. 1.

The datasets used in the experiment included the initial symptoms or blood index of COVID-19 patients. The laboratory datasets contain only a few patient symptoms for the study, and the clinical datasets contain information on actual COVID-19 patients at the time of admission to hospitals in some countries.

Fever, runny nose, body pain, sore throat, and difficulty breathing are the most comm symptoms in patients whose information is accessible in the datasets. The patient labels used for classification are indicated at the end of the datasets and were either COVID-19 or no COVID-19. Table 1 and Table 2 briefly describe the symptom information in the datasets.

2.1.2. Preprocessing

Due to the different sources of the datasets collected in the experiment, the format and information in the datasets are also different. It is necessary to preprocess the experimental datasets. The original data include some problems such as different data representation formats, incomplete data information, and unbalanced data distribution. Several preprocessing techniques are applied to the dataset to remedy these issues.

Most laboratory datasets have data format problems, such as the character gender information. Data transformation converts a data format from one type to another, which can standardize the datasets and smooth the experiment. Since the ML model requires that all

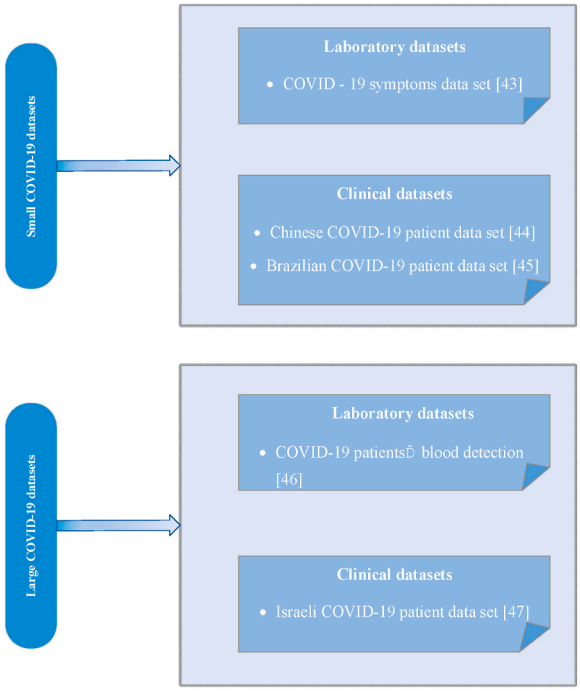


Fig. 1. The datasets used in the paper.

Table 1
The symptom descriptions of the laboratory dataset.

Symptom	Value	Description
Age	Integer	The patient's age
Fever	Integer	The patient's body temperature in Fahrenheit
Body Pain	Boolean	Develops symptoms accompanied with body pain or lower back pain; a score of 0 means no, and a score of 1 means yes
Runny Nose	Boolean	Develops a runny nose; a score of 0 means no, and a score of 1 means yes
Difficulty Breathing (Dyspnea)	0, 1 or −1	Develops symptoms of difficulty breathing or tachypnea; values of 0,1, and −1 represent the severity
Infection	Boolean	The patient had a positive contact with COVID-19

Table 2
The symptom descriptions of the clinical dataset.

Symptom	Value	Description
Gender	String	Representation of the patient's gender
Age 60 and above	Boolean	Measures of patient age with 60 set as the boundary; a score of 0 means no, and a score of 1 means yes
Cough	Boolean	Develops symptoms with a dry cough; a score of 0 means no, and a score of 1 means yes
Fever	Boolean	Develops symptoms with a high body temperature of 38 °C or more; a score of 0 means no, and a score of 1 means yes
Sore Throat	Boolean	Develops a sore, red, and swollen throat; a score of 0 means no, and a score of 1 means yes
Shortness of Breath	Boolean	Develops difficulty breathing or tachypnea; a score of 0 means no, and a score of 1 means yes
Headache	Boolean	Develops headache or nausea; a score of 0 means no, and a score of 1 means yes
Trajectory Information	String	Patient's isolation treatment status and travel history

information used be inputted be in numerical form, the character symptom information is transformed into an integer. Some datasets contain many missing values, which are not collected or are collected incorrectly. These incorrect inputs can lead to incorrect experiments and results. Deletion and completion are used to address data incompleteness, which results in a complete COVID-19 symptoms dataset. In addition, the data may be affected by uncertain and inaccurate factors. To address this problem, fuzzy logic is incorporated with data classification after inputting the data [48,49]. The proposed method can divide the data with unfixed fuzzy rules and generate fuzzy rules suitable for each data point to improve the classification performance.

Moreover, some original clinical datasets include ambiguous information about which symptoms manifest themselves clearly in the early stages of infection. In the Chinese dataset, the data contain patient symptoms in a text format that is not available in the experiment. Therefore, a string-matching algorithm is designed to search for symptom keywords and generate the regular dataset seen in Fig. 2. The selected data from six different provinces in China are processed into a proper dataset format that can be used in experiments. Labels assigned to patients at the end of the dataset indicate whether they are infected, and are then used for classification. The dataset after preprocessing provides a basis for the detection of the proposed model. The corresponding pseudocode is shown in Algorithm 1.

Algorithm 1. (Pseudocode of the string-matching algorithm).

2.2. Methodology

In medical practice, the prediction and classification of trends and severity of symptoms severity are crucial factors. ML methods can be used to analyze the importance of the different disease symptoms. Faced with the COVID-19 pandemic, there is an urgent need to identify effective predictive classification tools. Therefore, this paper establishes an intensive symptom weight learning mechanism called ISW-LM, to predict the diagnosis and risk for critical COVID-19 based on the clinical and laboratory parameters of patients. The proposed method learns the weight of patients' symptoms to diagnose and predict whether patients have COVID-19 and to classify the severity of symptoms.

In this paper, three weight functions are proposed to calculate and rank the symptoms of COVID-19 patients. According to the order of weight calculated by the functions, the intensive of symptoms is used to predict whether COVID-19 patients are infected.

2.2.1. Symptom weight measures

If the COVID-19 datasets contain $\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN}$ patients and $\text{precision} = \frac{TP}{TP+FP}$ symptoms, then the values of symptoms $\text{Recall} = \frac{TP}{TP+FN}$ across $F1\ \text{score} = 2 \times \frac{Recall \times Precision}{Recall+Precision}$ patients form an m-element vector. A comparison of the symptoms in the weight functions is produced with the value of the m-element vector, which ranges between 0 and $F1\ \text{score} = 2 \times \frac{Recall \times Precision}{Recall+Precision}$. The ranking of symptom weight represents the importance. Intense symptoms with a high weight will be used for prediction, while symptoms with a low weight can be discarded.

2.2.1.1. Support vector weight score (SV-WS). The SVM algorithm for supervised machine learning provides a theoretical foundation based on the notion of margins [50,51]. Instances on either side of a boundary hyperplane are divided into two classes, healthy or diseased. The boundary hyperplane can be obtained by calculating the symptom correlation between the two classes. According to the hyperplane definition, it can be described as follows [50]:

$$y_i(x_i \cdot w + b) - 1 \geq 0, \quad \forall i = 1, \ldots, n$$

where x_i is the i-th instance of patients and y_i is the classification label, which indicates the state of patients. w is the vector of symptom weight and b is a constant of trade-off.

By constructing a Lagrange function, the weight vector w of symptoms can be explained with the Lagrange multipliers and the training samples as follows [51]:

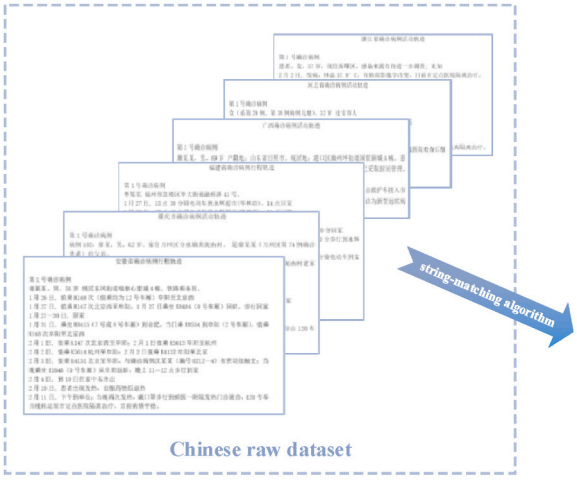
$$w = a_i y_i x_i, \quad \forall i \in [1, n]$$

Here, a_i is its corresponding class labels and Lagrange multiplier.

The correlation between each symptom and category may vary little. To distinguish the weight of symptoms clearly, the adjustment strategy of weight calculation is redefined in this paper:

$$w = a_i y_i x_i + \left(1 - \frac{1}{k}|\cos(x_i)|\right), \quad \forall i \in [1, n]$$

Here, the coefficient $1/k$ ensures that the later term of w ranges between 0 and 1. The new w ensures that the important symptoms are determined in the classification with higher weights and that the unimportant symptoms that have no effect on classification have lower weights.



After preprocessing

fever	cough	isolation	gender	age
1	0	1	1	56
0	0	1	1	46
1	0	1	0	43
0	0	0	1	56
0	0	0	1	46
1	1	1	1	26
1	0	0	1	41
0	0	1	1	38
0	0	0	0	30
1	0	1	1	54
0	0	1	0	55

Fig. 2. The Chinese COVID-19 dataset was processed by a string-matching algorithm.

```
Standardized Chinese COVID-19 dataset using feature selection
Initialize the number of patient instances and symptoms
Initialize the number of iterations i
For each patient
    While (i < the number of symptoms in a patient)
        Store text included each patient's symptoms
    end while
    Update the text included all patients' symptoms
end for
Initialize the string of patient symptoms s
Initialize a two-dimensional array for symptoms
For each patient
        Comparison of keywords with symptom string s
        if (the keyword matches the string successfully)
            generate regular data of symptoms for patients
        end if
end for
Return the two-dimensional array for symptoms
```

2.2.1.2. Information entropy weight score (IE-WS). By classifying the symptoms and finding the most representative symptoms, the state and category of patients can be accurately judged. Information entropy can be used to measure the importance of symptoms [52,53]. Since the decision tree (DT) algorithm can represent the connection between attributes and features through information entropy, the importance of COVID-19 patients' symptoms can be calculated based on the theoretical foundation of information entropy in the DT.

Information entropy is defined as the difference between patients' symptoms. A smaller entropy coefficient indicates a greater difference and more importance among the symptoms. Therefore, the symptom weight can be measured by the information entropy value as follows [52]:

$$e_i = -\sum_{i=1}^{n} p_i log_2 p_i$$

where p_i represents the probability that the patient belongs to the COVID-19 class or not.

The difference in coefficient among various symptoms is calculated by the following equation [54]:

$$d_j = 1 - e_j, \quad j \in [1, m]$$

Thus, the symptom weight can be adjusted by its importance. It can be redescribed as follows:

$$w_j = \frac{d_j}{\sum_{j=1}^{m} d_j}, \quad j \in [1, m]$$

$$w = e_i - w_j \frac{d_j}{\sum_{j=1}^{m} d_j} \cdot e_i', \quad i \in [1, n] \quad j \in [1, m]$$

where w_j is the current symptom weight and e_i' is the current symptom's information entropy.

2.2.1.3. Euclidean distance gini weight score (EDG-WS). In ML, RF is an ensemble classifier containing multiple decision trees with the same tree structure, which integrates trees through a resampling process called bagging [55,56]. The theory of ensemble learning in the RF algorithm can calculate the contribution of each symptom to each tree and calculate their average. The ratio between the symptoms can be used to determine how important the symptom is to the diagnosis of COVID-19 and its severity.

During forest growth, each tree, leaf, and root node in the forest generates a Gini value for symptom importance evaluation. The Gini value is calculated as follows [55]:

$$Gini(t) = 1 - \sum_{j=1}^{k} [p(j|t)]^2$$

where $p(j|t)$ is the probability of class j at node t and k is the number of classification results.

If there is a significant difference in the Euclidean distance of the same symptom between two different classes of patients that can distinguish whether patients are sick or serious, and the intensive symptom weight can be increased to make that symptoms more important. The weight computation formula is shown as follows:

$$w_i = w_i - \left(\frac{d(R_i, S_i)}{m} - \frac{d(R_i, D_i)}{m}\right) \cdot Gini(m), \quad (i = 1, 2, \ldots, n)$$

where w_i is the weight of the i-th symptom and m is the i-th patient. R, S and D are the samples of standard, ill or severe patients, and healthy people or mild patients, respectively.

2.2.2. The proposed method

The proposed ISW-LM is a mechanism consisting of five processing stages data preprocessing, the proposed symptom weight functions, the sort of symptoms' importance with the weight, intensive symptom weight, and the attribute prediction or diagnosis of patients. The flow chart of the proposed ISW-LM is illustrated in Fig. 3.

2.2.2.1. Data preprocessing. The purpose of data preprocessing is to eliminate outliers and balance the impact of data. In this paper, the selected COVID-19 symptom datasets are of different sizes and from different sources, and types. Therefore, it is crucially important to preprocess these datasets. This phase is the operations of handling missing values, cleaning up outliers, and balancing the data distribution.

First, datasets are divided into small and large COVID-19 datasets according to their size. Furthermore, the COVID-19 datasets can be subdivided into laboratory and clinical datasets based on their sources. Second, missing values and outliers in the data sets are processed and the COVID-19 datasets are supplemented. For the unbalanced distribution of data, datasets are balanced by randomly deleting most class patients and creating a minority of class patients. Finally, format conversion is carried out for some special COVID-19 datasets such as the original Chinese dataset. A string-matching algorithm shown in Fig. 4 is designed to transform the format, where the symptoms of COVID-19 patients in the dataset are extracted. After these steps, preprocessed and organized datasets are available for the following experiments.

2.2.2.2. Symptom weight calculation and ranking. Those infected with COVID-19 and ordinary patients appear to have many similar symptoms, so prominent symptoms of them can be given a higher weight to compare the attributes of patients.

Several designed weight functions, i.e., SV-WS, IE-WS, and EDG-WS, that are introduced in section 2 are used to calculate the contribution of each symptom to diagnose COVID-19 patients. Combined with the patient's classification labels, the corresponding weights for diagnosing COVID-19 and its severity are obtained. Then, the weight value is uniformly standardized for subsequent calculation and evaluation. The weight value with high reliability can be reserved by the designed weight functions.

COVID-19 patients always have some prominent symptoms, which are vital for diagnosing COVID-19. These can be obtained by integrating and ranking the symptoms of diagnosed patients in this work. The

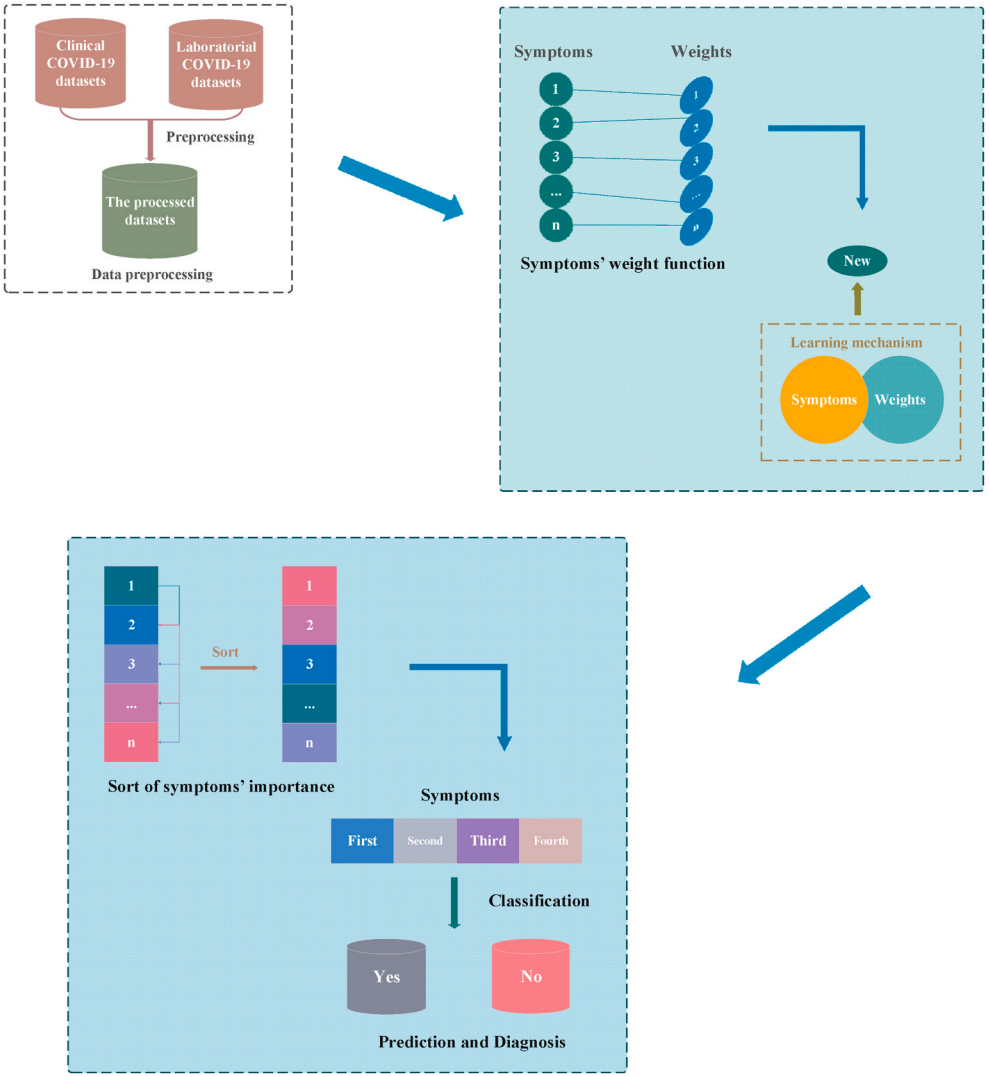
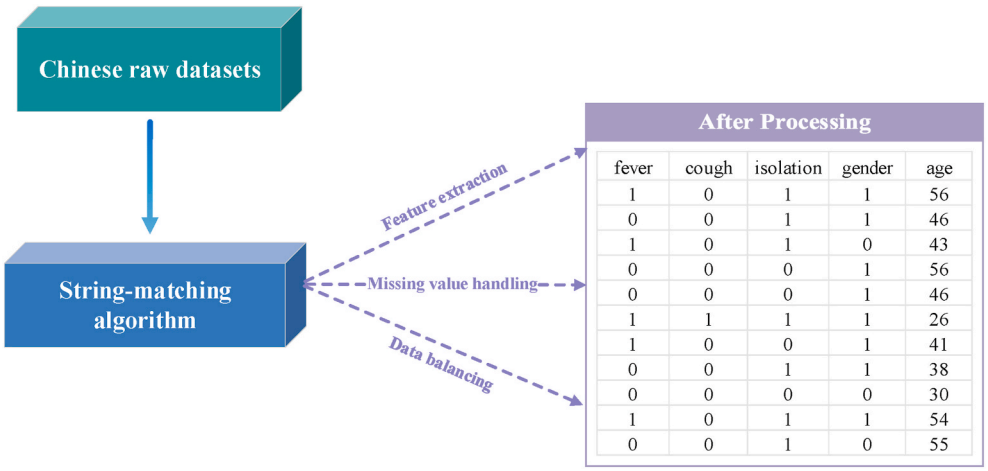


Fig. 3. The flow chart of the proposed ISW-LM.



fever	cough	isolation	gender	age
1	0	1	1	56
0	0	1	1	46
1	0	1	0	43
0	0	0	1	56
0	0	0	1	46
1	1	1	1	26
1	0	0	1	41
0	0	1	1	38
0	0	0	0	30
1	0	1	1	54
0	0	1	0	55

Fig. 4. The preprocessing dataset.

visualized steps are shown inFig. 5.

The ranking function is designed to sort out the relative importance of symptoms and can order them using the weight value. Symptoms that have higher prioritization ranks have major relativity with COVID-19 diagnosis. This can improve the ability to identify COVID-19 patients at an early stage using clinical symptoms.

2.2.2.3. ISW-LM for patient classification and diagnosis. In this paper, the ISW-LM is designed to improve the accuracy of patient diagnosis through the contribution of important symptoms. Here, each patient's symptoms are regarded as independent, and the calculated weights obtained by the above functions are incorporated into them.

The proposed method is a process for constructing the calculated weight until all symptoms are clearly represented, which is defined as intensive symptom weight. Through continuous learning and integration, the difference between the symptoms increases, and the importance of intensity becomes more prominent. Meanwhile, the binary grasshopper optimization algorithm (BGOA) [57] is integrated to process the differences that can help to classify and diagnose patients who are either infected with COVID-19 or not. The ISW-LM results provide a basis for classifying and diagnosing patients infected with early COVID-19. Fig. 6 shows the corresponding steps.

To ensure the diagnostic accuracy, the high-ranking intensive symptoms are selected as the basis for classification in BGOA. Patients with higher levels are classified as suspected or diagnosed. Besides, weights intensity can also separate already infected patients who are severe from those who are not. This can help doctors diagnose patients and perform next steps. Meanwhile, this work can satisfy the accuracy of classification results, which provides reference credibility for decision making.

3. Experimental results and discussions

3.1. Performance metrics

Performance measurement is an essential task in machine learning and can typically be measured based on a classification algorithm. Each of the following five performance metrics is used in the paper to evaluate the quality of the proposed method: accuracy, precision, recall, F1 score, and confusion matrix. They are the primary metrics for determining the class of correctly identified COVID-19 patients [58].

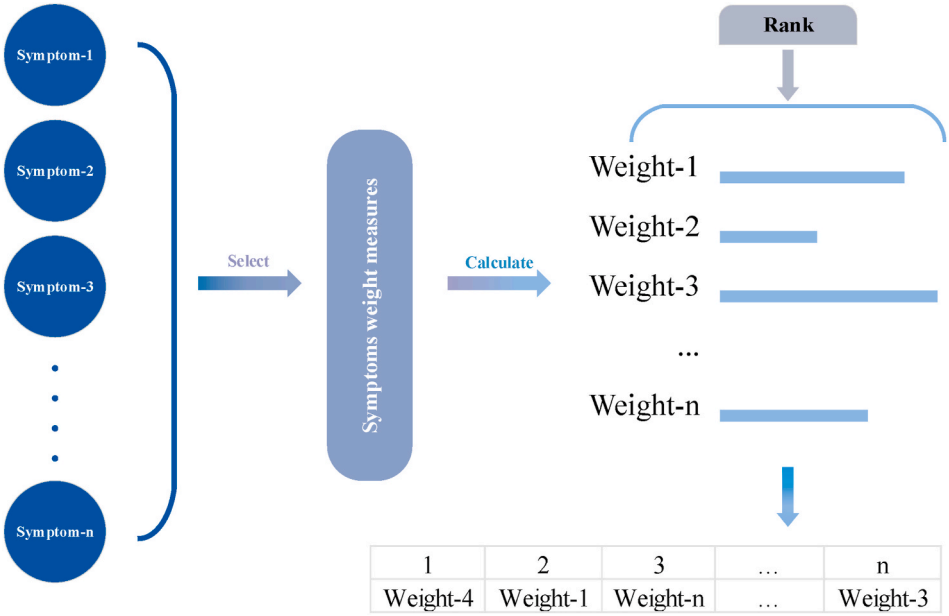


Fig. 5. Calculation and sorting of the weight functions.

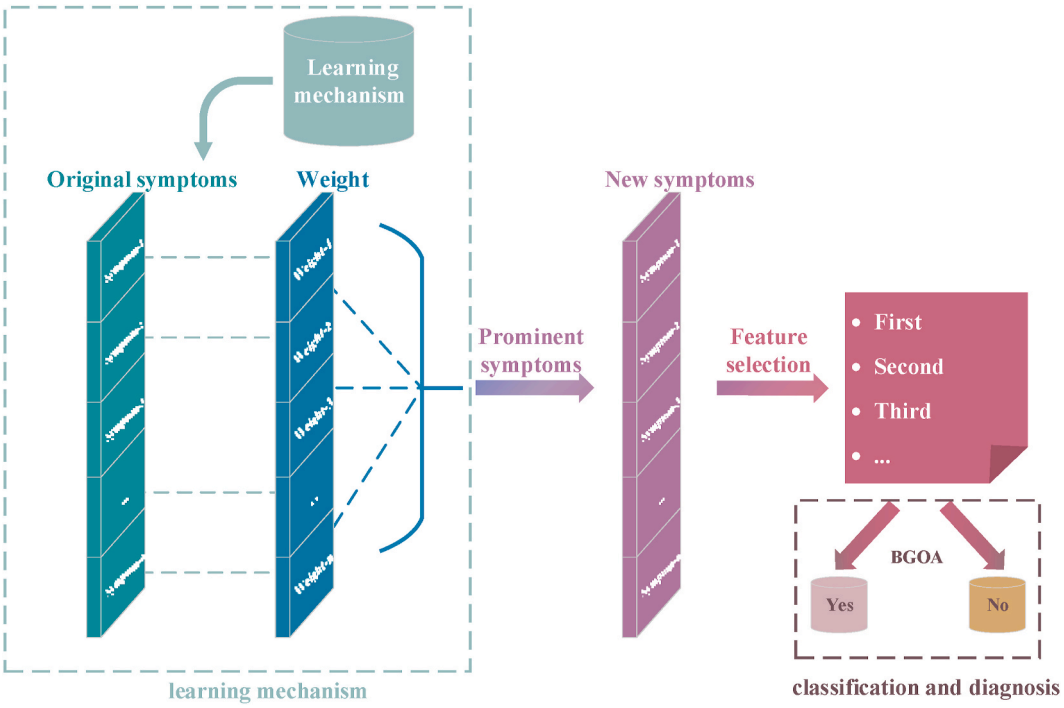


Fig. 6. Process of the ISW-LM in classification and diagnosis.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

$$precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1\ score = 2 \times \frac{Recall \times Precision}{Recall + Precision}$$

A confusion matrix is a form for evaluating the accuracy of prediction results. The columns represent the two conditions, or classes, of either having COVID-19 or not. The rows represent the actual classes and the number of patients.

Here, the true positive (*TP*) refers to the number of patients confirmed as COVID-19 positive the method correctly identifies. The true negative (*TN*) represents the number of patients without COVID-19, and the false positive (*FP*) and the false negative (*FN*) are the opposite of *TP* and *TN*, respectively [59].

3.2. Performance evaluation

In this study, multiple independent experiments are performed to ensure the reliability of the proposed method for COVID-19 prediction. The experiment is carried out in datasets, and the patient symptoms in different datasets are analyzed and ranked by the proposed ISW-LM. After the ranking of symptom weight, the diagnosis of COVID-19 depends on the intensity of some important symptoms in different datasets, so that other patients can be classified and predicted using BGOA. To evaluate the model's generalizability, the datasets are divided into 80% for training and the 20% for testing [60].

3.2.1. Evaluation of symptom importance

The proposed symptom weight functions involve selecting symptoms to obtain the best results from the ISW-LM. Combined with the weight functions, the order associated with each symptom is obtained. After numerous replications of the symptom selection experiments, the top four symptoms for diagnosing and classifying COVID-19 are identified as the optimal result. The symptoms with high-ranking values are listed in Tables 3–7.

3.2.1.1. In the small COVID-19 datasets. The orders shown in Table 3, Table 4, and Table 5 are the results of the small COVID-19 datasets calculated by different symptom weight functions in ISW-LM. Table 3 describes the four most significant symptoms that are strictly related to COVID-19 positive status. Tables 4 and 5 show the order in small clinical COVID-19 datasets from China and Brazil.

3.2.1.2. In the large COVID-19 datasets. The crucial symptoms used to

Table 3
The order of symptoms in laboratory datasets.

Dataset	Symptom weight function	The order of symptoms			
		First	Second	Third	Fourth
Symptom-1	SV-WS	Age	Fever	Body pain	Infection
	IE-WS	Age	Fever	Infection	Body pain
	EDG-WS	Fever	Infection	Age	Body pain
Symptom-2	SV-WS	Body pain	Infection	Age	Fever
	IE-WS	Fever	Age	Body pain	Infection
	EDG-WS	Fever	Age	Body pain	Runny nose
Symptom-3	SV-WS	Abroad travel	Fever	Cough	Sore throat
	IE-WS	Sore throat	Abroad travel	Dyspnea	Fever
	EDG-WS	Dyspnea	Fever	Abroad travel	Cough
Symptom-4	SV-WS	Cough	Fever	Dyspnea	Sore throat
	IE-WS	Fever	Cough	Dyspnea	Sore throat
	EDG-WS	Fever	Cough	Dyspnea	Sore throat

Table 4
The order of symptoms in the Chinese datasets.

Dataset	Symptom weight function	The order of symptoms			
		First	Second	Third	Fourth
Anhui	SV-WS	Fever	Cough	Age	Isolation
	IE-WS	Fever	Age	Cough	Isolation
	EDG-WS	Fever	Cough	Age	Isolation
Chongqing	SV-WS	Fever	Cough	Isolation	Age
	IE-WS	Age	Fever	Gender	Cough
	EDG-WS	Isolation	Age	Fever	Cough
Fujian	SV-WS	Fever	Gender	Cough	Isolation
	IE-WS	Fever	Gender	Isolation	Cough
	EDG-WS	Fever	Gender	Cough	Isolation
Guangxi	SV-WS	Fever	Cough	Gender	Isolation
	IE-WS	Age	Fever	Cough	Isolation
	EDG-WS	Gender	Fever	Age	Cough
Hebei	SV-WS	Fever	Cough	Isolation	Age
	IE-WS	Age	Fever	Gender	Isolation
	EDG-WS	Fever	Age	Cough	Isolation
Zhejiang	SV-WS	Fever	Cough	Age	Isolation
	IE-WS	Age	Fever	Cough	Isolation
	EDG-WS	Fever	Age	Gender	Isolation

Table 5
The order of symptoms in Brazilian datasets.

Dataset	Symptom weight function	The order of symptoms			
		First	Second	Third	Fourth
Brazilian dataset-1	SV-WS	Dyspnea	Coryza	Runny nose	Fever
	IE-WS	Runny nose	Dyspnea	Coryza	Fever
	EDG-WS	Fever	Dyspnea	Gender	Runny nose
Brazilian dataset-2	SV-WS	Fever	Runny nose	Coryza	Taste
	IE-WS	Fever	Gender	Cough	Runny nose
	EDG-WS	Runny nose	Dyspnea	Gender	Cough
Brazilian dataset-3	SV-WS	Runny nose	Fever	Coryza	Taste
	IE-WS	Fever	Runny nose	Gender	Dyspnea
	EDG-WS	Fever	Dyspnea	Gender	Cough

classify COVID-19 or not are diverse in different datasets. Tables 6 and 7 show the order of symptoms in large laboratory COVID-19 datasets and clinical datasets, respectively.

3.2.2. Symptom weight function evaluation

3.2.2.1. Evaluation in the small COVID-19 datasets. The experiments are divided into two parts according to the size of the datasets mentioned in 3.1. Table 8 shows the accuracy and other metrics of small COVID-19 data sets calculated from symptom weight functions taken from three laboratory and clinical symptom datasets from three provinces in China.

In comparison to the different symptom weight functions in ISW-LM, Table 3 that the results of the two types of datasets are different. The accuracy of weight functions calculated in the laboratory COVID-19 datasets is up to 97.1711%, with precision and recall rates reaching 100% and over 99.75%, respectively, while the F1 Score is above 99.87%.

The above results are obtained by SV-WS in the Symptom-3 dataset. For the clinical COVID-19 datasets, the results are evenly distributed. Taking the clinical COVID-19 dataset of Anhui Province as an example, the highest accuracy rate is 81.8182%, the precision rate is over 83.33%, the recall rate is up to 85.6618%, and the F1 Score has a top value of 84.1919%. However, the accuracy rate is lower than 55% for datasets with a single symptom, such as the Symptom-1 and Symptom-2 datasets (see Fig. 6).

Moreover, the performance metrics mentioned above can be further appreciated in the confusion matrix shown in Fig. 7, which demonstrates that most of the COVID-19 classes are properly identified, however, a few are incurred in misclassifications. The SV-WS used in the Symptom-3 dataset is proven to be the best, and the EDG-WS is optimal in the Anhui dataset.

To determine the overall performance of the accuracy of the proposed symptom weight functions, the confidence limits of the three symptom weight functions are shown in Fig. 8, using the Symptom-3 and Anhui datasets as examples. It can be seen that the accuracy of the proposed ISW-LM is approximately 97% and 80% in the symptom-3 and Anhui datasets, respectively. The results of Fig. 8 show that the average accuracy, which is given in Table 8, is reliable.

3.2.2.2. Evaluation in the large COVID-19 datasets. Furthermore, the same experiments are implemented on the large COVID-19 datasets. Table 9 shows the performance metrics for symptom weight functions in large COVID-19 datasets.

The results show that the accuracy rate in the laboratory blood test dataset shows a better value is better, with an overall score of 75.5144%. However, the accuracy of the proposed algorithm varies greatly, with a difference of 5.5556%. In addition, the recall and F1 Score of the proposed method still need to be balanced and improved. The analysis shows a high precision, recall, and F1 score, with values of 87.3482%, 80.0714%, and 77.8219%, respectively. Compared with other symptom weight functions in the clinical Israeli dataset-2 dataset, the SV-WS has an optimal performance in terms of accuracy and precision. Besides, the second function emerged as the best weight function, a recall rate 48.6312% and an F1 score of 47.3068%.

Table 7
The order of symptoms in Israeli datasets.

Dataset	Symptom weight function	The order of symptoms			
		First	Second	Third	Fourth
Israeli dataset-1	SV-WS	Headache	Sore throat	Dyspnea	Gender
	IE-WS	Gender	Headache	Cough	Fever
	EDG-WS	Headache	Sore throat	Gender	Dyspnea
Israeli dataset-2	SV-WS	Sore throat	Fever	Headache	Dyspnea
	IE-WS	Gender	Headache	Fever	Sore throat
	EDG-WS	Headache	Sore throat	Dyspnea	Gender

Table 6
The order of symptoms in blood test dataset.

Dataset	Symptom weight function	The order of symptoms			
		First	Second	Third	Fourth
Blood test	SV-WS	Platelets	Kallistatin	Red blood cells	Monocytes count
	IE-WS	Aspartate aminotransferase	Eosinophils count	White blood cells	Lactate dehydrogenase
	EDG-WS	Eosinophils count	Calcium	Nucleic acid testing	Polymerase chain reaction

Table 8
Performance metrics for symptom weight functions in small COVID-19 datasets.

Dataset		Symptom weight function	Performance metric (%)			
			Accuracy	Precision	Recall	F1 Score
Laboratory COVID-19 datasets	Symptom-1	SV-WS	54.0291	49.3269	49.4581	49.3912
		IE-WS	54.3689	53.4857	55.5454	54.4742
		EDG-WS	53.5947	51.5344	48.4104	49.8790
	Symptom-2	SV-WS	53.3750	48.7965	51.1771	49.8673
		IE-WS	54.0625	49.8320	52.1533	50.9333
		EDG-WS	53.9375	52.6599	51.9158	52.1567
	Symptom-3	SV-WS	97.0791	100.0000	99.7545	99.8768
		IE-WS	97.1711	100.0000	98.5267	99.2549
		EDG-WS	96.5271	100.0000	99.1406	99.5670
Clinical COVID-19 datasets	Anhui	SV-WS	76.1363	87.7104	73.3543	77.9652
		IE-WS	81.8182	83.3333	85.3641	84.1919
		EDG-WS	80.1137	84.0852	85.6618	83.9773
	Chongqing	SV-WS	62.3189	76.1905	100.0000	86.4397
		IE-WS	66.3044	81.3665	79.9295	80.3953
		EDG-WS	66.3044	81.6782	74.9806	77.8922
	Hebei	SV-WS	75.4464	83.5556	87.5776	83.6111
		IE-WS	70.0893	86.1111	88.1067	86.1047
		EDG-WS	76.7857	82.6087	81.5045	81.2598

Fig. 7. Confusion matrix for symptom weight functions in small COVID-19 datasets.

As shown in Fig. 9, 54 COVID-19 patients and 63 non-COVID-19 patients in the blood test dataset, whose result is calculated by the EDG-WS, are properly classified. On the other hand, 97.3% of patients in the Israeli dataset-2 dataset are categorized correctly by the SV-WS using functions.

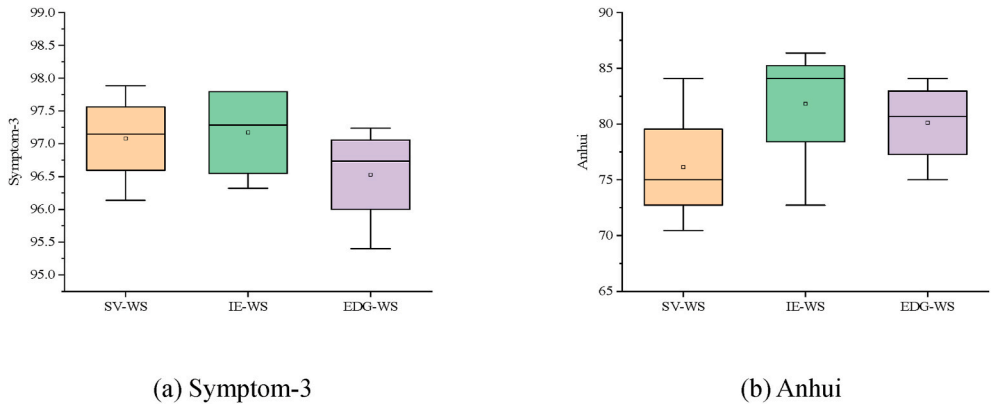
Similarly, the overall performance of the proposed ISW-LM function in the blood test and Israeli dataset-2 datasets is shown in Fig. 10. As shown in Fig. 10, the overall accuracy distribution is approximately 72%, and the average values in Table 9 are all within the confidence limit.

The accuracy of the proposed method in Table 9 is approximately 95% on average and is within the interval shown in Fig. 10. The results show that the accuracy is obtained with a highly convincing probability.

3.2.3. *Analysis of experimental results*

In this paper, the symptom weight function is used to calculate, rank, and diagnose the crucial symptoms related to COVID-19 based on the proposed ISW-LM. The corresponding function and symptoms are selected for the specific datasets by comparing the accuracy of the three functions in the ISW-LM.

Combined together, Table 3, Table 4, Table 8, and Fig. 7 show that in the small laboratory COVID-19 datasets, fever, body pain, age, and dyspnea are the most crucial symptoms. The specific datasets are analyzed by taking Symptom-3 and Anhui data as examples. The ISW-LM predicted with a high accuracy that sore throat, travel abroad, dyspnea, and fever are the most significant symptoms for the Symptom-3 data. Similarly, the key symptoms for diagnosing COVID-19 in Anhui



(a) Symptom-3 (b) Anhui

Fig. 8. Confidence limits of accuracy in the Symptom-3 and Anhui datasets.

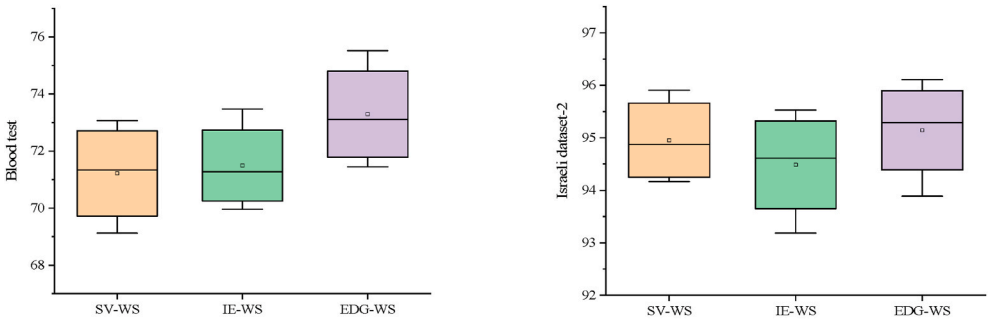
Table 9
Performance metrics for symptom weight function in large COVID-19 datasets.

Dataset		Symptom weight function	Performance metric (%)			
			Accuracy	Precision	Recall	F1 Score
Laboratory COVID-19 dataset	Blood test	SV-WS	70.3195	87.3482	17.1630	28.5386
		IE-WS	69.9588	71.1580	74.4211	72.6900
		EDG-WS	75.5144	75.7425	80.0714	77.8219
Clinical COVID-19 dataset	Israeli dataset-2	SV-WS	95.9081	52.8562	27.7115	36.3588
		IE-WS	95.5282	46.5812	48.6312	47.3068
		EDG-WS	95.6849	49.7243	44.6894	45.6310

Fig. 9. Confusion matrix for symptom weight functions in large COVID-19 datasets.

Province are fever, age, cough, and isolation. Furthermore, the accuracy of datasets containing simple symptoms still needs to be improved in the accurate selection of key symptoms.

As shown in Table 6, Table 7, Table 9, and Fig. 9, the top fourth highest-ranking symptoms for the blood test dataset are eosinophil count, calcium, nucleic acid testing, and polymerase chain reaction. In the large clinical COVID-19 datasets, the Israeli dataset-2 dataset is taken as an example. The most important symptoms for the diagnosis of COVID-19 include headache, sore throat, dyspnea, and sex. Meanwhile, more attention should be given to the generalization capabilities of the



(a) Blood test (b) Israeli dataset-2

Fig. 10. Confidence limits of accuracy in blood test and Israeli dataset-2 datasets.

proposed ISW-LM, including the diagnosis of hematological and epidemiological symptoms in large clinical COVID-19 datasets.

3.3. Comparison with state-of-the-art methods

To better evaluate the proposed ISW-LM, this experiment is dedicated to comparing the recall metric in which the blood test dataset is used with state-of-the-art methods. Eight algorithms that have performed well in classification have been selected for comparisons, such as DT, RF, KNN, SVM, etc., using the proposed method in Latif, Siddique, et al. [13], called TWRF, as shown in Fig. 11.

Fig. 11 summarizes the recall metric that is measured by nine classification methods in the blood test dataset. This result indicates that the proposed method is superior to other algorithms for this measure, and the ISW-LM gives the highest testing score of 87%. Most classification methods have a recall rate of less than 70%. Therefore, the best-performing models could be usefully applied in clinical scenarios, which verifies that the proposed method can classify COVID-19 effectively.

4. Conclusion

The early detection and diagnosis of COVID-19 patients are critical to preventing the spread of the disease and promptly treating patients. Recent studies have revealed that patients' epidemiological symptoms and routine blood tests can be used to classify and screen for COVID-19. This study proposes an intensive symptom weight learning mechanism called ISW-LM to classify and diagnose COVID-19 patients. Three symptom weight functions are proposed to analyze and evaluate the importance of symptom intensity for a positive diagnosis. These rankings of symptom intensity may aid doctors in identifying potentially infected patients before a formal diagnosis is made. Finally, multiple laboratory and clinical COVID-19 datasets are used to test the validity of the proposed model. By analyzing the results, the model presents the important symptoms that identify COVID-19 in different datasets, in which the most frequent and significant predictive symptoms in most datasets for diagnosing COVID-19 are fever, sore throat, and cough. Different state-of-the-art classification models are also used to compare and verify the effectiveness of the proposed ISW-LM. Experimental results show that the proposed ISW-LM can obtain an accuracy of 97.1711%. Compared with that of other algorithms, the recall rate can also be increased to 87%.

By analyzing and judging the intensity of the symptoms of infected patients, the proposed method can assist doctors in the treatment and reliable early detection of COVID-19, which can save both treatment time and cost. In future work, the proposed method can also be used to diagnose the degree of infection in patients with severe infections or complications of chronic diseases.

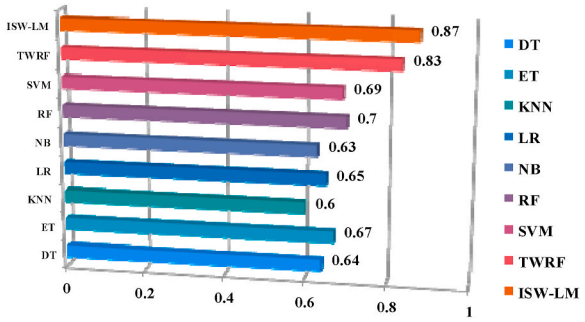


Fig. 11. Comparison with state-of-the-art methods in the recall metric.

Declaration of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by China Postdoctoral Science Foundation under Grant 2021M700676, Natural Science Foundation of Liaoning Province under Grant 2021-MS-272, and Dalian high-level Talents Innovation plan under Grant 2019RQ021.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2022.105615.

References

[1] Who, Coronavirus Disease (Covid-19), 2022. https://www.who.int/emergencies/diseases/novel-coronavirus-2019. (Accessed 17 February 2022).
[2] Stephan Ludwig, Zarbock Alexander, Coronaviruses and SARS-CoV-2: a Brief Overview, Anesthesia and analgesia, 2020.
[3] Raju Vaishya, et al., Artificial intelligence (AI) applications for COVID-19 pandemic. Diabetes & metabolic syndrome, Clin. Res. Rev. 14 (4) (2020) 337–339.
[4] Chenglong Liu, et al., Differentiating novel coronavirus pneumonia from general pneumonia based on machine learning, Biomed. Eng. Online 19 (1) (2020) 1–14.
[5] Punn, Narinder Singh, Sanjay Kumar Sonbhadra, Sonali Agarwal, COVID-19 Epidemic Analysis Using Machine Learning and Deep Learning Algorithms, MedRxiv, 2020.
[6] Miguel Marcos, et al., Development of a severity of disease score and classification model by machine learning for hospitalized COVID-19 patients, PLoS One 16 (4) (2021), e0240200.
[7] Samuel Lalmuanawma, Jamal Hussain, Lalrinfela Chhakchhuak, Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: a review, Chaos, Solit. Fractals 139 (2020) 110059.

[8] Zaid Abdi Alkareem Alyasseri, et al., Review on COVID-19 diagnosis models based on machine learning and deep learning approaches, Expet Syst. 39 (3) (2022), e12759.
[9] Umer Saeed, et al., Machine learning empowered COVID-19 patient monitoring using non-contact sensing: an extensive review, Journal of pharmaceutical analysis 12 (2) (2022) 193–204.
[10] Jinrui Gao, et al., Predictive criteria of severe cases in COVID-19 patients of early stage: a retrospective observational study, J. Clin. Lab. Anal. 34 (10) (2020), e23562.
[11] Abdul Hafeez, et al., A review of COVID-19 (Coronavirus Disease-2019) diagnosis, treatments and prevention, EJMO 4 (2) (2020) 116–125.
[12] Abhirup Banerjee, et al., Use of machine learning and artificial intelligence to predict SARS-CoV-2 infection from full blood counts in a population, Int. Immunopharm. 86 (2020) 106705.
[13] Siddique Latif, et al., Leveraging data science to combat covid-19: a comprehensive review, IEEE Transactions on Artificial Intelligence 1 (1) (2020) 85–103.
[14] Islam, Muhammad Nazrul, et al., A Systematic Review on the Use of AI and ML for Fighting the COVID-19 Pandemic, IEEE Transactions on Artificial Intelligence, 2021.
[15] Hanumanthu Swapnarekha, et al., Role of intelligent computing in COVID-19 prognosis: a state-of-the-art review, Chaos, Solit. Fractals 138 (2020) 109947.
[16] Davide Brinati, et al., Detection of COVID-19 infection from routine blood exams with machine learning: a feasibility study, J. Med. Syst. 44 (8) (2020) 1–12.
[17] Thomas Tschoellitsch, et al., Machine learning prediction of SARS-CoV-2 polymerase chain reaction results with routine blood tests, Lab. Med. 52 (2) (2021) 146–149.
[18] Rachid Zagrouba, et al., Modelling and simulation of COVID-19 outbreak prediction using supervised machine learning, Comput. Mater. Continua (CMC) (2021) 2397–2407.
[19] Mahdi Mahdavi, et al., A machine learning based exploration of COVID-19 mortality risk, PLoS One 16 (7) (2021), e0252384.
[20] Ahmed Hamed, Sobhy Ahmed, Hamed Nassar, Accurate classification of COVID-19 based on incomplete heterogeneous data using a KNN variant algorithm, Arabian J. Sci. Eng. 46 (9) (2021) 8261–8272.
[21] Matjaž Kukar, et al., COVID-19 diagnosis by routine blood tests using machine learning, Sci. Rep. 11 (1) (2021) 1–9.
[22] M. Buvana, K. Muthumayil, Prediction of COVID-19 patient using supervised machine learning algorithm, Sains Malaysiana 50 8 (2021) 2479–2497.
[23] Warda M. Shaban, et al., Accurate detection of COVID-19 patients based on distance biased Naïve Bayes (DBNB) classification strategy, Pattern Recogn. 119 (2021) 108110.
[24] Prabh Deep Singh, et al., A novel ensemble-based classifier for detecting the COVID-19 disease for infected patients, Inf. Syst. Front 23 (6) (2021) 1385–1401.
[25] Md Raihan, et al., Development of Risk-free COVID-19 Screening Algorithm from Routine Blood Test Using Ensemble Machine Learning, 2021 arXiv preprint arXiv: 2108.05660.
[26] Yuri Kravchenko, et al., Machine Learning Algorithms for Predicting the Results of COVID-19 Coronavirus Infection, IT&I Workshops, 2020.
[27] Nan-Nan Sun, et al., A Prediction Model Based on Machine Learning for Diagnosing the Early COVID-19 Patients, medRxiv, 2020.
[28] Jiangpeng Wu, et al., Rapid and Accurate Identification of COVID-19 Infection through Machine Learning Based on Clinical Available Blood Test Results, MedRxiv, 2020.
[29] Pijush Dutta, Shobhandeb Paul, Asok Kumar, Comparative Analysis of Various Supervised Machine Learning Techniques for Diagnosis of COVID-19. Electronic Devices, Circuits, and Systems for Biomedical Applications, Academic Press, 2021, pp. 521–540.
[30] Ibrahim Arpaci, et al., Predicting the COVID-19 infection with fourteen clinical features using machine learning classification algorithms, Multimed. Tool. Appl. 80 (8) (2021) 11943–11957.
[31] Alves, Marcos Antonio, et al., Explaining machine learning based diagnosis of COVID-19 from routine blood tests with decision trees and criteria graphs, Comput. Biol. Med. 132 (2021) 104335.
[32] Lucas M. Thimoteo, et al., Explainable artificial intelligence for COVID-19 diagnosis through blood test variables, Journal of Control, Automation and Electrical Systems (2022) 1–20.
[33] Ahamad, Md Martuza, et al., A machine learning model to identify early stage symptoms of SARS-Cov-2 infected patients, Expert Syst. Appl. 160 (2020) 113661.
[34] Dan Assaf, et al., Utilization of machine-learning models to accurately predict the risk for critical COVID-19, Internal and emergency medicine 15 (8) (2020) 1435–1443.
[35] Maryam AlJame, et al., Ensemble learning model for diagnosing COVID-19 from routine blood tests, Inform. Med. Unlocked 21 (2020) 100449.
[36] L.J. Muhammad, et al., Supervised machine learning models for prediction of COVID-19 infection using epidemiology dataset, SN computer science 2 (1) (2021) 1–13.
[37] Sakifa Aktar, et al., Machine learning approaches to identify patient comorbidities and symptoms that increased risk of mortality in COVID-19, Diagnostics 11 (8) (2021) 1383.
[38] Janmenjoy Nayak, et al., Intelligent system for COVID-19 prognosis: a state-of-the-art survey, Appl. Intell. 51 (5) (2021) 2908–2938.
[39] L.S. Suma, H.S. Anand, Nature inspired optimization model for classification and severity prediction in COVID-19 clinical dataset, J. Ambient Intell. Hum. Comput. (2021) 1–13.
[40] Krishnaraj Chadaga, et al., Clinical and laboratory approach to diagnose COVID-19 using machine learning, Interdiscipl. Sci. Comput. Life Sci. (2022) 1–19.
[41] Krishnaraj Chadaga, et al., Medical diagnosis of COVID-19 using blood tests and machine learning, J. Phys. Conf. 2161 (1) (2022), 012017.
[42] Kwekha-Rashid, Ameer Sardar, Heamn N. Abduljabbar, Bilal Alhayani, Coronavirus disease (COVID-19) cases analysis using machine-learning applications, Appl. Nanosci. (2021) 1–13.
[43] Alam Takbir, Kaggle, 2021. https://www.kaggle.com/datasets/takbiralam/covid19-symptoms-dataset. (Accessed 7 November 2021).
[44] Beijing Advanced Innovation Center for Big Data and Brain Computing, Beihang University, Beijing, 2020, p. 100191. https://github.com/BDBC-KG-NLP/COVID-19-tracker. (Accessed 8 July 2021).
[45] Íris Viana dos Santos Santana, et al., Mendeley Data, 2021. https://data.mendeley.com/datasets/b7zcgmmwx4/4. (Accessed 5 July 2021).
[46] Cabitza, Federico, and Andrea Campagner. The need to separate the wheat from the chaff in medical informatics: introducing a comprehensive checklist for the (self)-assessment of medical AI studies. Int. J. Med. Inf. 153 (2021): 104510.
[47] Israeli ministry of Health. https://data.gov.il/dataset/covid-19, 2021. (Accessed 8 July 2021).
[48] Ta Zhou, Fu-lai Chung, Shitong Wang, Deep TSK fuzzy classifier with stacked generalization and triplely concise interpretability guarantee for large data, IEEE Trans. Fuzzy Syst. 25 (5) (2016) 1207–1221.
[49] Mario Versaci, et al., Joint use of eddy current imaging and fuzzy similarities to assess the integrity of steel plates, Open Phys. 18 (1) (2020) 230–240.
[50] Fa Zhu, et al., A weighted one-class support vector machine, Neurocomputing 189 (2016) 1–10.
[51] Wen Zhang, et al., Feature weighted confidence to incorporate prior knowledge into support vector machines for classification, Knowl. Inf. Syst. 58 (2) (2019) 371–397.
[52] Ibomoiye Domor Mienye, Yanxia Sun, Zenghui Wang, Prediction performance of improved decision tree-based algorithms: a review, Procedia Manuf. 35 (2019) 698–703.
[53] Kapil Juneja, Chhavi Rana, An improved weighted decision tree approach for breast cancer prediction, Int. J. Inf. Technol. 12 (3) (2020) 797–804.
[54] Jiaqiang Zou, Pengfei Li, Modelling of litchi shelf life based on the entropy weight method, Food Packag. Shelf Life 25 (2020) 100509.
[55] Min Zhu, et al., Class weights random forest algorithm for processing class imbalanced medical data, IEEE Access 6 (2018) 4641–4652.
[56] Harsimran Guram, Ashok Sharma, Patch base Segmentation for Classification of Dementia disorder with optimize feature weight and Random forest based approach, Mater. Today Proc. (2021), https://doi.org/10.1016/j.matpr.2020.12.208.
[57] Majdi Mafarja, et al., Binary grasshopper optimisation algorithm approaches for feature selection problems, Expert Syst. Appl. 117 (2019) 267–286.
[58] Tripti Goel, et al., Automatic Screening of Covid-19 Using an Optimized Generative Adversarial Network, Cognitive computation, 2021, pp. 1–16.
[59] Pijush Dutta, Shobhandeb Paul, Asok Kumar, Comparative Analysis of Various Supervised Machine Learning Techniques for Diagnosis of COVID-19. Electronic Devices, Circuits, and Systems for Biomedical Applications, Academic Press, 2021, pp. 521–540.
[60] Sakifa Aktar, et al., Predicting Patient COVID-19 Disease Severity by Means of Statistical and Machine Learning Analysis of Blood Cell Transcriptome Data, 2020, p. 10657, arXiv preprint arXiv:2011.