

Article

Prediction of Pulmonary Function Parameters Based on a Combination Algorithm

Ruishi Zhou ^{1,2}, Peng Wang ^{1,3}, Yueqi Li ^{1,2}, Xiuying Mou ^{1,2}, Zhan Zhao ^{1,3}, Xianxiang Chen ^{1,3}, Lidong Du ^{1,3,*}, Ting Yang ^{4,*}, Qingyuan Zhan ^{4,*} and Zhen Fang ^{1,2,3,*}

- ¹ Aerospace Information Research Institute, Chinese Academy of Sciences (AIRCAS), Beijing 100190, China; zhouruishi17@mails.ucas.ac.cn (R.Z.); wangpeng01@aircas.ac.cn (P.W.); liyueqi18@mails.ucas.ac.cn (Y.L.); mouxiuying18@mails.ucas.ac.cn (X.M.); zhaozhan@mail.ie.ac.cn (Z.Z.); chenxx@aircas.ac.cn (X.C.)
- ² School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100190, China
- ³ Personalized Management of Chronic Respiratory Disease, Chinese Academy of Medical Sciences, Beijing 100190, China
- ⁴ Department of Respiratory Medicine, China-Japan Friendship Hospital, Beijing 100029, China
- * Correspondence: lddu@mail.ie.ac.cn (L.D.); zryyyangting@163.com (T.Y.); zhanqy@163.com (Q.Z.); zfang@mail.ie.ac.cn (Z.F.)

Abstract: Objective: Pulmonary function parameters play a pivotal role in the assessment of respiratory diseases. However, the accuracy of the existing methods for the prediction of pulmonary function parameters is low. This study proposes a combination algorithm to improve the accuracy of pulmonary function parameter prediction. Methods: We first established a system to collect volumetric capnography and then processed the data with a combination algorithm to predict pulmonary function parameters. The algorithm consists of three main parts: a medical feature regression structure consisting of support vector machines (SVM) and extreme gradient boosting (XGBoost) algorithms, a sequence feature regression structure consisting of one-dimensional convolutional neural network (1D-CNN), and an error correction structure using improved K-nearest neighbor (KNN) algorithm. Results: The root mean square error (RMSE) of the pulmonary function parameters predicted by the combination algorithm was less than 0.39L and the R^2 was found to be greater than 0.85 through a ten-fold cross-validation experiment. Conclusion: Compared with the existing methods for predicting pulmonary function parameters, the present algorithm can achieve a higher accuracy rate. At the same time, this algorithm uses specific processing structures for different features, and the interpretability of the algorithm is ensured while mining the feature depth information.

Keywords: combination algorithm; support vector machines; extreme gradient boosting; one-dimensional convolutional neural network; improved K-nearest neighbor



Citation: Zhou, R.; Wang, P.; Li, Y.; Mou, X.; Zhao, Z.; Chen, X.; Du, L.; Yang, T.; Zhan, Q.; Fang, Z. Prediction of Pulmonary Function Parameters Based on a Combination Algorithm. *Bioengineering* **2022**, *9*, 136. <https://doi.org/10.3390/bioengineering9040136>

Academic Editors: Jijiang Yang, Jianqiang Li and Larbi Boubchir

Received: 24 February 2022

Accepted: 23 March 2022

Published: 25 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent years, artificial intelligence (AI) and machine learning (ML) have rapidly evolved in various fields, including healthcare. These methods can help detect diseases, improve pathological classification, and predict disease patterns and epidemiology, a prime example of which is ML-based algorithms developed during the COVID-19 pandemic [1,2]. In addition, the authors of [3] created a system that developed and trained a neural network model for the diagnosis of diabetes mellitus in pregnant women and the accuracy of the trained network was over 92%. In [4], a fuzzy expert system was proposed for diagnosing and analyzing human diseases. The system not only indicates if the disease is present but also indicates the level at which the disease is present. It is notable that this approach for diagnosing human diseases has an accuracy and reliability of 97%. The authors of [5], developed an expert system for oral ulcers that focuses on four common oral ulcers. In addition, the study of medical image data in [6], used CT images for the segmentation and

classification of small hepatocellular carcinoma, and achieved good results. The accuracy of the segmentation was 0.9049, and the accuracy of the classification was 0.838.

The above studies are all able to achieve good results in their corresponding fields, but not in the field of chronic respiratory diseases.

At the same time, chronic respiratory diseases, including chronic obstructive pulmonary disease (COPD) and asthma, are significantly increasing in regard to morbidity and mortality worldwide. They can affect individuals of all age groups and cause over 3 million deaths each year according to World Health Organization data [7,8]. Therefore, it is important to apply AI technology to the assessment, diagnosis, and treatment stages of chronic respiratory diseases, thereby reducing patient morbidity and mortality.

The assessment of the patient's pulmonary function parameters is a prerequisite for the prevention and treatment of chronic respiratory diseases. Today, spirometry is one of the most widely-used techniques to assess pulmonary function [9]. Unfortunately, spirometry has strict end-of-test criteria and poor patient compliance leads to the low accuracy of pulmonary function parameters. Schermer et al. found that 50% of the spirometry methods were inaccurate in terms of pulmonary function parameters [10]. For this reason, many scholars have conducted studies on the prediction of lung function parameters. Sharan et al. investigated the prediction of lung function parameters using coughing sounds [11], while Ioachimescu et al. (2020) performed partial lung function prediction based on age, sex, race, height, and weight and using an artificial neural network (ANN) algorithm [12]. Miyoshi et al. (2020) developed regression equations to estimate forced vital capacity (FVC) and forced expiratory volume in one second (FEV1) [13]. Chen et al. developed an FEV1 and FVC prediction model based on multi-output support vector regression [14].

Meanwhile, volumetric capnography has emerged as a technique for pulmonary function assessment that helps to solve the problem of inaccurate prediction of pulmonary function parameters and has wide application prospects [15]. Jarenbäck et al. (2018) obtained an index of efficiency of tidal ventilation with respect to CO₂ exchange (efficiency index, EFFi) in volumetric capnography and tested the hypothesis that EFFi may be used for the diagnostics and grading of COPD [16]. Kellerer et al. (2020) conducted a systematic analysis of the relationship between capnovolumetric and conventional lung function parameters to help in the interpretation of capnovolumetric parameters [17]. Although these authors conducted preliminary studies on volumetric capnography, they did not use volumetric capnography data for the specific prediction of pulmonary function parameters. They only elaborated on the correlation between volumetric capnography and pulmonary function parameters such as FVC and FEV1.

Therefore, in this paper, a combination algorithm based on volumetric capnography data is proposed for the first time to solve the problem of the accuracy of pulmonary function parameters prediction, thus improving the accuracy of pulmonary function parameter prediction.

The novelty and contributions of this study are as follows.

- (1) This paper is the first to propose the use of volumetric capnography data for the prediction of pulmonary function parameters, which is more accessible and less demanding for testers than other studies.
- (2) The algorithm proposed in this paper combines the advantages of traditional machine learning algorithms for processing high-dimensional medical features and deep learning for learning low-dimensional sequence features, to improve the accuracy of pulmonary function parameter prediction.
- (3) This paper provides a reference paradigm for other medical data processing by handling high-dimensional features and low-dimensional features in medical data.

In the subsequent sections, we first establish the signal acquisition system and compensate the system signal using an adaptive control algorithm. Then, the proposed combination algorithm is described in detail and the performance of different algorithms is compared. The experimental results show that the combination algorithm proposed in this paper has

high accuracy in the prediction of pulmonary function parameters and can significantly improve the quality of pulmonary function assessment.

2. Materials and Methods

2.1. Signal Acquisition System

To enable better volumetric capnography data acquisition, we built a homemade signal acquisition system. We tested the performance of the system with a PWG-33 pulmonary waveform generator and a carbon dioxide concentration verification platform, and the system’s measurement error was within 5%. The signal acquisition system is shown in Figure 1. It mainly contains handheld multi-sensor devices and a user interface, and the system signal is compensated by an adaptive adjustment algorithm.

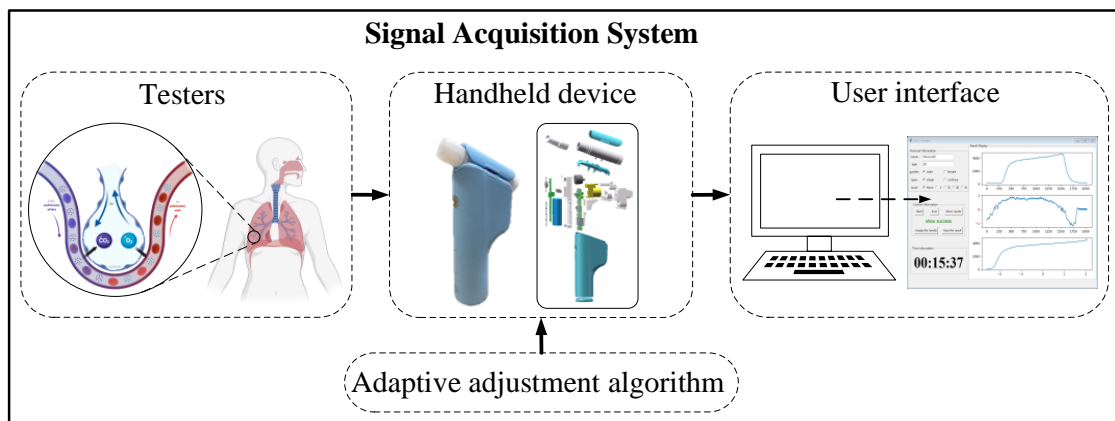


Figure 1. Signal acquisition system.

2.1.1. Devices and User Interface

The embedded system of the handheld device is shown in Figure 2, including a microprocessor module, power management module, 4G communication module, display module, keyboard control module, a sensor acquisition module [18].

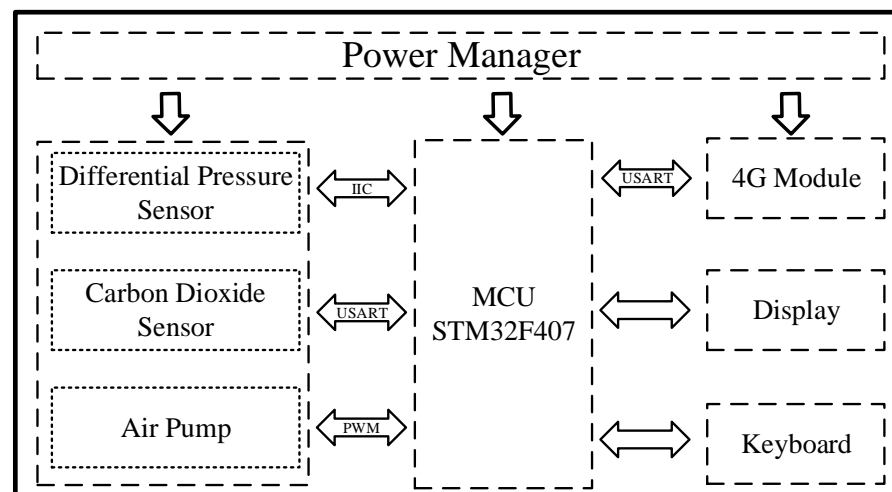


Figure 2. Embedded system of the handheld device. The microcontroller unit (MCU) communicates with the differential pressure sensor via Inter-Integrated Circuit (IIC) protocol, with the carbon dioxide sensor via Universal Synchronous/Asynchronous Receiver/Transmitter (USART) protocol, and the air pump control via pulse-width modulation (PWM) wave.

The microprocessor module is used to perform adaptive control algorithm and data processing, the power management module is responsible for the power supply of the

entire device, the 4G communication module is responsible for signal transmission, the display module and the keyboard control module are used to interact with the user, and the sensor acquisition module is used for data acquisition.

The user interface mainly includes a personal information window, a control information window, a time information window, and a result display window. Firstly, the user fills in the contents of the personal information window as prompted. Then, different buttons are selected for interaction in the control information window. When the start button is clicked, the user interface plays a guided breathing tone and collects the user’s breathing information. At the same time, the time information window displays and records the breath time information. After breathing, the user can click the show button to display the results. The respiratory flow rate, the respiratory carbon dioxide concentration, and the volumetric capnography will be displayed in the result display window. Finally, the user can click the save results button to save the collected information in CSV format locally for subsequent processing and analysis.

2.1.2. Adaptive Adjustment Algorithm

During respiration, the flow rate will always change. As a result, inconsistencies in the characteristic parameters of volumetric capnography can occur under different respiratory conditions. Therefore, the signal acquisition system uses an adaptive adjustment algorithm based on minimum prediction error to ensure the accuracy of volumetric capnography measurements. The algorithm can ensure the consistency of the characteristic parameters of volumetric capnography, and control the flow in advance to reduce the lag in the flow control due to the ability to predict the flow according to the actual flow. The algorithm is similar to the idea of minimizing the local structure error [19–21].

The adaptive adjustment algorithm is mainly divided into three stages, the prediction of respiratory flow at the next moment, the calculation of forecast error, and the adaptive adjustment of smoothing parameters.

- Prediction of respiratory flow at the next moment.

Set the initial sampling flow f_0 , the initial smoothing parameter α_0 and the prediction window size N . Get the actual respiratory flow $F_{i-N+1}, F_{i-N+2}, \dots, F_{i-1}, F_i$ at the previous N moments through the sampling of the differential pressure sensor. According to the traditional exponential smoothing algorithm, predict the respiratory flow at the next moment \hat{F}_{i+1} :

$$\hat{F}_{i+1} = \alpha_0 F_i + \alpha_0(1 - \alpha_0)F_{i-1} + \alpha_0(1 - \alpha_0)^2 F_{i-2} + \dots \tag{1}$$

- Calculation of forecast error.

Get the actual respiratory flow F_{i+1} at time $i + 1$, and record the difference in flow E_{i+1} between the predicted flow \hat{F}_{i+1} at $i + 1$ time and the actual respiratory flow F_{i+1} :

$$E_{i+1} = \hat{F}_{i+1} - F_{i+1} \tag{2}$$

Set the error calculation window size W , calculate the mean value of the difference in the sliding window $\overline{E_w}$:

$$\overline{E_w} = \sum_{x=i-w}^i \frac{E_x}{W} \tag{3}$$

- Adaptive adjustment of smoothing parameters.

According to E_{i+1} , $\overline{E_w}$, the self-adjustment coefficient β , the smoothing parameter α , adjust the smoothing parameter α :

$$\alpha_{i+1} = \alpha_i \left(1 + \beta \left(\frac{E_{i+1}}{\overline{E_w}} - 1 \right) \right) \tag{4}$$

Control the sampling flow f_{i+1} according to the updated smoothing parameter α_{i+1} :

$$f_{i+1} = \alpha_{i+1}f_i + \alpha_{i+1}(1 - \alpha_{i+1})f_{i-1} + \alpha_{i+1}(1 - \alpha_{i+1})^2f_{i-2} + \dots \tag{5}$$

In different breathing situations, the adaptive algorithm can adjust the sampling flow rate, solve the problem of inconsistencies in the characteristic parameters of the volumetric capnography under different respiratory flows, and ensure the accuracy of the volumetric capnography [22]. The details of the adaptive adjustment algorithm are shown in Algorithm 1.

Algorithm 1: Adaptive Adjustment Algorithm.

Input: initial sampling flow f_0 , initial smoothing parameter α_0 , the prediction window size N , error calculation window size W , the self-adjustment coefficient β .

Output: smoothing parameters α_{i+1} , sampling flow f_{i+1}

while obtaining the actual respiratory F_i **do**

for $\text{len}(F) < N$ **do**

$F = F.add(F_i)$

predict the respiratory flow at the next moment \hat{F}_{i+1}

$\hat{F}_{i+1} = \alpha_0F_i + \alpha_0(1 - \alpha_0)F_{i-1} + \alpha_0(1 - \alpha_0)^2F_{i-2} + \dots$

obtain actual breathing flow at the $i + 1$ time point F_{i+1}

calculation of forecast error

$E_{i+1} = \hat{F}_{i+1} - F_{i+1}$

calculate the mean value of the difference in the sliding window $\overline{E_w}$

$\overline{E_w} = \sum_{x=i-w}^i \frac{E_x}{W}$

adaptive adjustment of smoothing parameters α_{i+1} and sampling flow f_{i+1}

$\alpha_{i+1} = \alpha_i \left(1 + \beta \left(\frac{E_{i+1}}{\overline{E_w}} - 1 \right) \right)$

$f_{i+1} = \alpha_{i+1}f_i + \alpha_{i+1}(1 - \alpha_{i+1})f_{i-1} + \alpha_{i+1}(1 - \alpha_{i+1})^2f_{i-2} + \dots$

end

2.2. Combination Algorithm

For the traditional single-structure pulmonary function regression algorithm, the mining of data information is limited by the structural design. Traditional machine learning algorithms can mine relationships in high-dimensional data (medical features, etc.) very well and provide good explanations, but the accuracy needs to be improved [23]. Deep learning algorithms can mine deep relationships from low-dimensional data (sequence data), but cannot provide good explanations [24].

Inspired by the combinatorial structure [25], we propose the combination algorithm, which mainly consists of three parts: a medical feature regression structure, a sequence feature regression structure, and an error correction structure. The medical features are processed by the traditional machine learning algorithm and the sequence data are processed by the deep learning algorithm. Finally, the results of the two are effectively combined to improve the accuracy rate. The algorithm ensures both a good interpretation of high-dimensional medical features and the full utilization of low-dimensional sequence data.

2.2.1. Medical Feature Regression Structure

We constructed a two-layer medical feature regression structure using support vector machines (SVM) and extreme gradient boosting (XGBoost) algorithms to take full advantage of the medical data in the volumetric capnography and also based on the a priori knowledge of the airflow limitation cutoffs of GOLD2020 [26].

For the prior medical knowledge of the airflow limitation cutoff point in GOLD2020, we first need to build a first-level classification task to determine whether a patient has airflow limitation or not. For this classification task, the input features are high-dimensional

data such as medical features and demographic features, for which the support vector machine (SVM) algorithm has a good processing effect. SVM is an optimal margin-based classification technique in machine learning [27].

Since the SVM algorithm is used for the binary classification task, the results obtained are too sparse with only two possibilities. Therefore, to provide more information for the secondary prediction, we used the sigmoid function to make the output result of SVM probabilistic, which contains more information and is denser. The Sigmoid function is calculated as:

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (6)$$

We combined the output probability values of the SVM with the original features as the input features for the second-level prediction to perform the regression prediction of the FEV1 and FVC parameters. For the regression prediction task, we chose the XGBoost algorithm for secondary regression in order to show the importance and interpretability of each feature on the regression results.

XGBoost was recently proposed by Chen and Guestrinis [28]. It is based on the original framework of gradient boosting, for a given data set with n examples and m features $\mathcal{D} = \{(\mathbf{x}_i, y_i)\} (|\mathcal{D}| = n, \mathbf{x}_i \in \mathbb{R}^m, y_i \in \mathbb{R})$, and uses K additive trees to approximate the output \hat{y}_i as the following:

$$\hat{y}_i = \phi(\mathbf{x}_i) = \sum_{k=1}^K f_k(\mathbf{x}_i), f_k \in \mathcal{F} \quad (7)$$

where f_k is an independent classification and regression tree (CART) at each of the k steps, which map the input variables \mathbf{x}_i to y_i . $\mathcal{F} = \{f(\mathbf{x}) = w_{q(\mathbf{x})}\} (q: \mathbb{R}^m \rightarrow T, w \in \mathbb{R}^T)$ is the space of all regression trees.

To learn the set of functions used in the model, we minimized the following regularized objective.

$$\mathcal{L}(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \quad (8)$$

$$\text{where } \Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (9)$$

The training loss function l and the regularization term Ω make up the regularized objective function. The difference between the predicted value \hat{y}_i and the value y_i is measured by the training loss function l . The regularization term Ω assesses the model's complexity and helps to smooth the final learned weight to avoid overfitting.

In addition, XGBoost includes two key techniques: shrinkage and column subsampling. At each stage of boosting, the shrinkage algorithm scales the newly supplied weights, reducing the effect of each tree and overfitting. To speed up the training process, column subsampling only selects a random subset of input characteristics while creating a tree [29].

The medical feature regression structure is shown in Figure 3.

First, the first-level prediction structure was constructed using SVM with FEV1/FVC = 0.7 as the threshold to classify and predict the airway obstruction condition and obtain the corresponding probability values [30]. Subsequently, the probability results of the first-level prediction were combined with the original feature as the input features of the second-level prediction structure. XGBoost was used to construct the secondary prediction structure, and the prediction results of the pulmonary function parameters were then output. During the implementation of the algorithm, we used a heuristic search for the selection of hyperparameters to achieve optimal results [31].

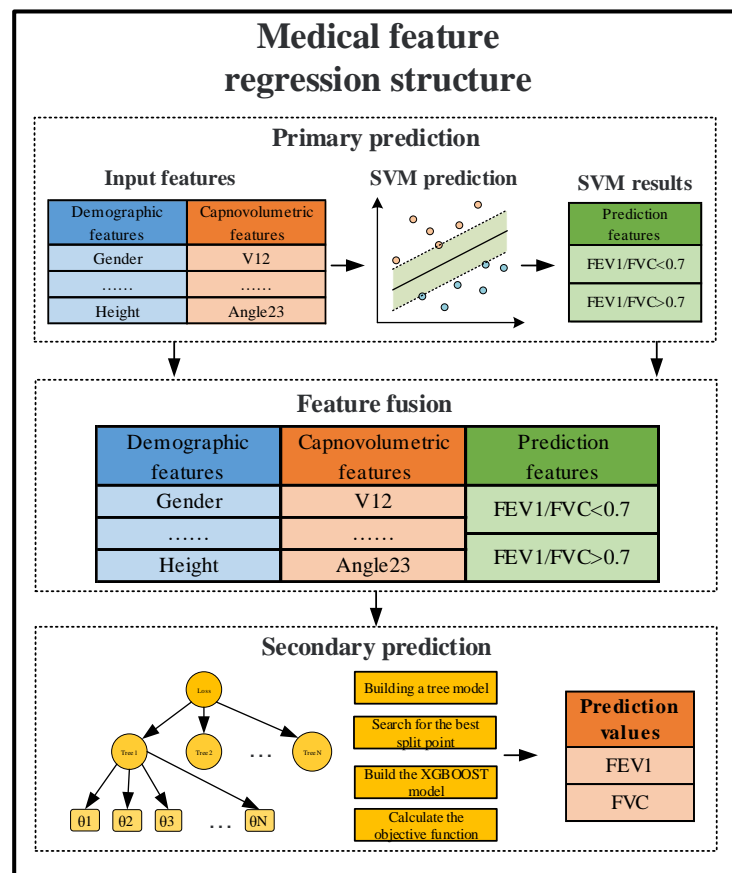


Figure 3. Structural flow chart of the medical feature regression structure.

2.2.2. Sequence Feature Regression Structure

To be able to make full use of the volumetric capnography data, we want to mine information from the low-dimensional raw sequence signals, in addition to using traditional machine learning algorithms for the high-dimensional medical and demographic features.

For the original low-dimensional CO₂ sequence data, deep learning networks have better learning abilities. When choosing the deep learning network structure, we considered that the original sequence information is only one-dimensional in depth; therefore, if we use network structures such as long short-term memory (LSTM) and gated recurrent units (GRU), it will increase the computational effort when performing the data processing with no effective improvement. Therefore, we chose the simpler one-dimensional convolutional neural network (1D-CNN), which is widely used in medical sequence signals [32–34].

Given a sequence of CO₂, $C_{1:n} = C_1, \dots, C_n$, a 1D convolution of width- k is the result of moving a sliding window of size s over the sequence, and applying the same convolution filter or kernel to each window in the sequence, i.e., a dot-product between the concatenation of the vectors in a given window and a weight vector u , which is then often followed by a non-linear activation function g . We chose the rectified linear unit (ReLU) function as the activation function to ensure the updating ability of the network when performing gradient backpropagation.

$$g(x) = \max(0, x) \tag{10}$$

The convolution filter is applied to each window, resulting in scalar values r_i , each for the i window:

$$r_i = g(x_i \cdot u) \in R \tag{11}$$

In practice, one typically applies more filters, $u_1, u_2 \dots, u_l$, which can then be represented as a vector multiplied by a matrix U and with the addition of a bias term b :

$$r_i = g(x_i \cdot U + b) \tag{12}$$

The network structure is shown in Figure 4.

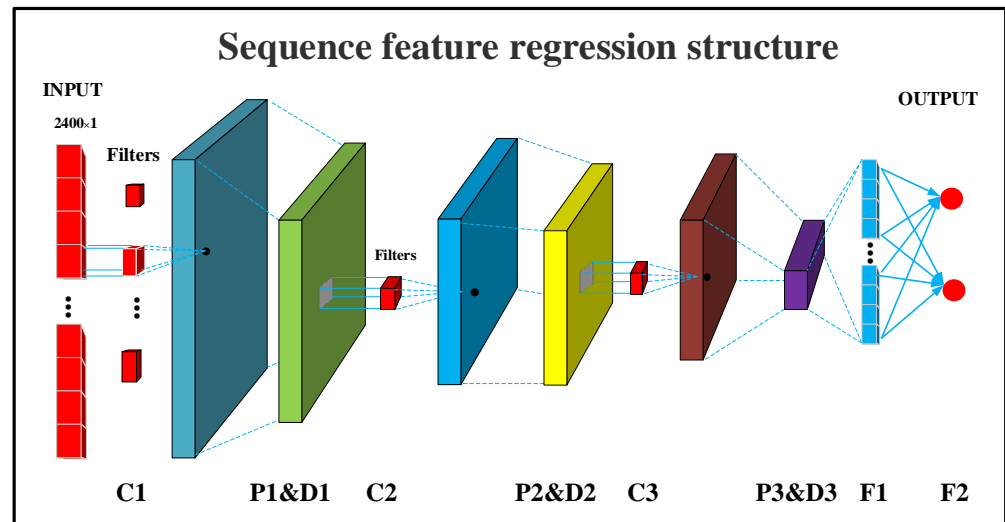


Figure 4. Structural flow chart of the sequence feature regression structure.

The network parameters are shown in Table 1.

Table 1. Configurations of sequence feature regression structure.

Layers (Type)	Output Size	Param
C1 (Conv1D)	(None, 2396, 32)	192
P1 (MaxPooling1D)	(None, 479, 32)	0
D1 (Dropout)	(None, 479, 32)	0
C2 (Conv1D)	(None, 475, 64)	10,304
P2 (MaxPooling1D)	(None, 95, 64)	0
D2 (Dropout)	(None, 95, 64)	0
C3 (Conv1D)	(None, 91, 32)	10,272
P3 (MaxPooling1D)	(None, 91, 32)	0
D3 (Dropout)	(None, 91, 32)	0
F1 (Flatten)	(None, 576)	0
F2 (Dense)	(None, 2)	1154

When building the network, we followed the general structure including a convolution layer, pooling layer and a dropout layer [35]. The convolution layer can be used to extract features from the sequence information, the pooling layer can be used to reduce the training parameters, and the dropout layer can be used to avoid training overfitting. We stacked the generic structure three times to ensure that the output vector has a larger receptive field [36]. After the stacked structure, we added a flattened layer to flatten the vector. Finally, we added a fully connected layer to map the flattened vectors to the FEV1 and FVC parameters and chose the mean square error as the loss function. When choosing the size and number of convolutional kernels, we chose a 1×5 convolutional kernel size, considering that a smaller convolutional kernel filter can help to improve computational efficiency and extract clearer features [37]. The number of convolutional kernels is generally a power of 2. In this case, we chose 32, 64, and 32 convolutional kernels, respectively.

2.2.3. Error Correction Structure

After processing high-dimensional information by traditional machine learning algorithms and low-dimensional information by deep learning algorithms, we need to combine the results of both algorithms organically to combine the respective advantages of both algorithms. To ensure the operability and interpretability of the synthesis results, we chose the improved K-nearest neighbor (KNN) regression algorithm as the output of the final results [38,39].

The traditional KNN algorithm is mainly used for classification problems. For two points $x = (x_1, x_2, \dots, x_n)$ and $y = (y_1, y_2, \dots, y_n)$ on an n -dimensional real vector space R^n , we can define a more generalized distance L_p between the two points, that is, the Minkowski distance as

$$L_p(x, y) = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{\frac{1}{p}} \tag{13}$$

Here, we used the spatial distance when $p = 2$, i.e., the Euclidean distance.

$$L_2(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \tag{14}$$

The traditional KNN algorithm divides the test samples into classes of the k-nearest samples in the n-dimensional space. We made certain improvements to the KNN algorithm to fit the regression problem here.

For any training sample, with the medical feature regression structure and the sequences feature regression structure, we get the output of a four-dimensional vector

$$Train_x_i = (FEV1_med_i, FVC_med_i, FEV1_seq_i, FVC_seq_i) \tag{15}$$

The true output of that sample is

$$Train_y_i = (FEV1_i, FVC_i) \tag{16}$$

For the test sample, the same four-dimensional output can be obtained with the two structures mentioned above.

$$Test_x_i = (FEV1_med_i, FVC_med_i, FEV1_seq_i, FVC_seq_i) \tag{17}$$

In the four-dimensional space, we calculated the distance between the test sample and all training samples. We chose the K closest distance training samples and took the true output of those K training samples and we calculated the mean of their pulmonary function parameters as the final output of our test samples.

$$Test_y_i = \left(\frac{1}{k} \sum_{i=1}^K FEV1_i, \frac{1}{k} \sum_{i=1}^K FVC_i \right) \tag{18}$$

With the improved KNN algorithm, we fully consider the output predicted by each traditional machine learning algorithm and deep learning algorithm, then integrate the results to arrive at the final prediction result. The improved KNN algorithm has advantages such as simple computation and strong interpretation.

The details of the combination algorithm flowchart are shown in Algorithm 2.

Algorithm 2: Combination Algorithm.

Input: Test set $X_{test} = \{x_{med}^{(i)}, x_{seq}^{(i)}\}_{i=1, \dots, n}$, $x_{med}^{(i)}$ is a medical feature vector, $x_{seq}^{(i)}$ is the sequence feature vector

Output: $Y_{test} = \{y^{(i)}\}_{i=1, \dots, n}$, $y^{(i)}$ is the pulmonary function parameter vector

for $i \leq n$ **do**

Medical Feature Regression Structure

$x_{med}^{(i)}$ through SVM model to obtain $y_{SVM}^{(i)}$

Fusion of features from $x_{med}^{(i)}$ and $y_{SVM}^{(i)}$ to obtain $x_{xgboost}^{(i)}$

$x_{xgboost}^{(i)}$ through XGBoost model to obtain $y_{med}^{(i)}$

Sequence feature regression structure

$x_{seq}^{(i)}$ through 1D-CNN model to obtain $y_{seq}^{(i)}$

Error correction structure

By splicing the vectors $y_{med}^{(i)}$ and $y_{seq}^{(i)}$, we obtain the vector $x_{KNN}^{(i)}$

$x_{KNN}^{(i)}$ through the KNN model to obtain $y^{(i)}$

end

3. Results

3.1. Regression Evaluation Index

Given the sensitivity of medical data to maximum error, we propose a comprehensive error evaluation index comprehensive percentage error (CPE), which integrates the maximum percentage error, mean absolute percentage error, and root mean square percentage error and is more suitable for the prediction evaluation of medical data [40].

The maximum percentage error (MPE) computes the maximum residual error percentage, a metric that captures the worst-case error between the predicted value and the true value. The mean absolute percentage error (MAPE) is an evaluation metric for regression problems. This metric is sensitive to relative errors. It is for example not changed by a global scaling of the target variable. The root mean square percentage error (RMSPE) is a measure of the deviation between the predicted value and the true value. They are defined as follows:

$$MPE(x_i, y_i) = \max\left(\frac{|x_i - y_i|}{y_i}\right) \times 100 \tag{19}$$

$$MAPE(x_i, y_i) = \frac{1}{n} \sum_{i=1}^n \frac{|x_i - y_i|}{\max(\epsilon, |y_i|)} \times 100 \tag{20}$$

$$RMSPE(x_i, y_i) = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2}}{\bar{y}_i} \times 100 \tag{21}$$

where x_i is the predicted value, y_i is the true value, n is the number of samples, \bar{y}_i is the mean of the true values, and ϵ is an arbitrary small, yet strictly positive number to avoid undefined results when y_i is zero.

The comprehensive percentage error (CPE), which is a combination of the maximum percentage error, mean absolute percentage error, and relative standard deviation, reflects the overall error of the regression results. The smaller the comprehensive error percentage, the better the regression results.

$$CPE(x_i, y_i) = \frac{1}{3} \times (MPE(x_i, y_i) + MAPE(x_i, y_i) + RMSPE(x_i, y_i)) \tag{22}$$

Therefore, the accuracy rate (ACC) considering the combined comprehensive percentage error is

$$ACC(x_i, y_i) = 1 - CPE(x_i, y_i) \tag{23}$$

3.2. Datasets

We performed volumetric capnography experiments and spirometry experiments on 1007 subjects (472 females and 535 males, aged 17–70 years). The sampling frequency of the volumetric capnography was 200 Hz, the sampling time of the data was greater than 20 s, and the length of the sequence of CO₂ acquisition was greater than 4000 sampling points. Pulmonary function experiments were performed to obtain the pulmonary function parameters, FEV1 and FVC. A total of more than 1007 subjects were obtained with three types of characteristic data. The volumetric capnography features are described in Table 2. The statistical analysis table is shown in Table 3.

Table 2. The description of volumetric capnography features.

Variable	Description	Units
C12	Carbon dioxide concentration at the boundary of phase 1 and phase 2	mmHg
C23	Carbon dioxide concentration at the boundary of phase 2 and phase 3	mmHg
V12	Volume at the boundary of phase 1 and phase 2	mL
V23	Volume at the boundary of phase 2 and phase 3	mL
V2	The volume of phase 2	mL
V3	The volume of phase 3	mL
S2	Slope of phase 2	mmHg/L
S3	Slope of phase 2	mmHg/L
S3/S2	The ratio of slopes of phases 3 and 2	/
Angle23	The angle between phases 2 and 3	°

Table 3. Data description table.

Amount	Category	Variable	Units	Values
Data	Demographics	Male	%	53.1%
		Age	years	56 (14)
		Height	cm	166 (9)
		Weight	kg	69 (14)
		BMI	kg·m ⁻²	24.94 (4.20)
	Volumetric capnography	C12	mmHg	2.49 (0.80)
		C23	mmHg	27.22 (4.76)
		V12	mL	276 (58)
		V23	mL	757 (157)
		V2	mL	480 (128)
		V3	mL	2061 (903)
		S2	mmHg/L	74.63 (25.19)
		S3	mmHg/L	5.44 (3.37)
	S3/S2	/	0.08 (0.04)	
	Angle23	°	168.26 (5.81)	
Spirometric	FEV1	l	2.53(0.86)	
	FVC	l	3.49(0.99)	

To improve contrast and produce a balanced database, a 10-fold cross-validation strategy was applied 10 times to decrease generalization error in the training set. Figure 5 depicts the schematic diagram of the 10-fold cross-validation.

The dataset is first divided into 10 equally-sized, mutually-exclusive subsets: $Data = d_1 \cup d_2 \cup \dots \cup d_{10}$. $d_i \cap d_j$ is empty. Each subsection maintains the consistency of the data distribution, which is obtained through hierarchical sampling from the data. $d_1, d_2 \dots d_{10}$ is used as the test set to obtain 10 test results, and the average value of the 10 test results is used as a cross-validation result.

In this study, cross-validation was performed ten times, with the results of the ten cross-validations being averaged as the final result to assess the algorithm’s performance.

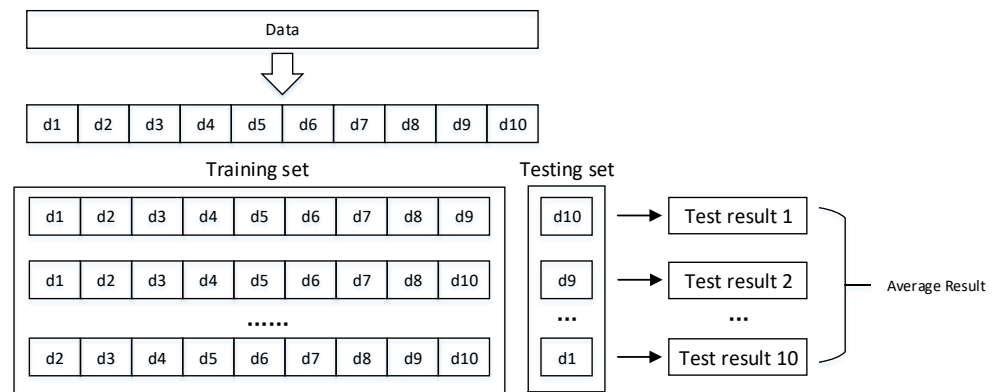


Figure 5. Schematic of 10-fold cross-validation method repeated 10 times.

3.3. Results of the Algorithm

3.3.1. Single-Structure Algorithm Results

To further assess the performance of our proposed combined algorithm, we compared single-structure machine learning and deep learning algorithms for pulmonary function prediction.

We used a single-structure conventional machine learning algorithm for pulmonary function prediction. Demographic and medical features were processed by SVM and XGBoost algorithms, and the results were evaluated using the relevant regression evaluation metrics.

The results of the conventional machine learning algorithm are shown in Table 4. For different pulmonary function parameters, FVC was a better predictor than FEV1 for R^2 and ACC metrics. This is consistent with the medical phenomenon whereby FEV1 measurements are dependent on the effort of the tester and have poor predictive accuracy. In the RMSE index, FVC is 0.48 L greater than FEV1’s 0.43 L, and since it is an absolute value, FVC is greater than FEV1, making the absolute value of RMSE of FVC greater than FEV1. Overall, the traditional machine learning algorithm was able to do a better job of processing the medical features and getting the expected results. The maximum in the R^2 metric was 0.79 and the maximum ACC was up to 79%. The result, however, still has room for improvement.

Table 4. Results of the conventional machine learning algorithm.

Type	Pulmonary Function Parameters	RMSE (L)	R^2	ACC
SVM + XGBoost	FEV1	0.43	0.78	73.90%
	FVC	0.48	0.79	79.18%

The features’ importance is shown in Figure 6. We can see that for the regression prediction of FEV1 versus FVC, the ranking of feature importance differs between the two. However, the two most important characteristics of both are demographic features, which is consistent with reality. As age and body size change, the human pulmonary function also undergoes significant changes.

The fitting curves and error percentages of conventional machine learning algorithms are shown in Figures 7 and 8. Figure 8 shows that the average error of FEV1 prediction is 15.71%, and the average error of FVC is only 12.26%. Additionally, the median error for both is less than their mean error, indicating that the pulmonary function prediction results are acceptable for most testers. However, we also see that there are individual outliers in the results predicted by FVC, which indicates that the performance of FVC prediction is poor for a small number of results and there is still room for optimization.

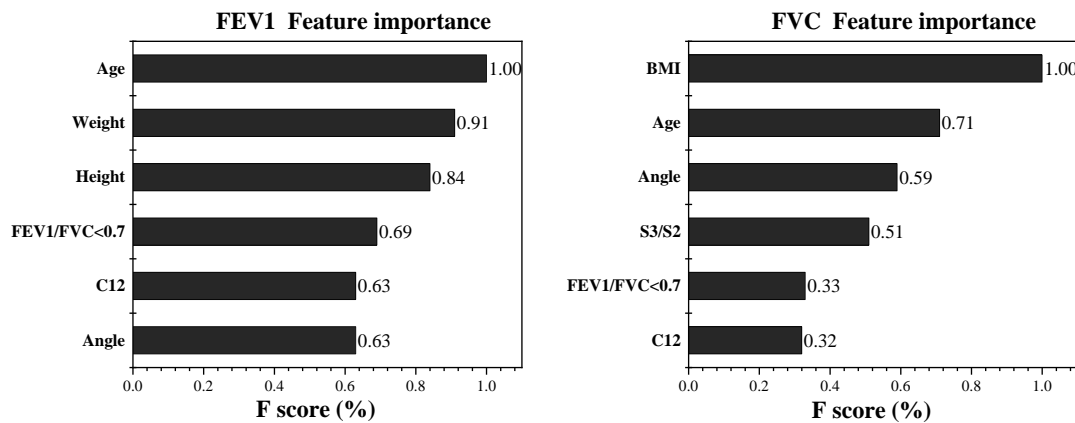


Figure 6. The features’ importance in conventional machine learning algorithms.

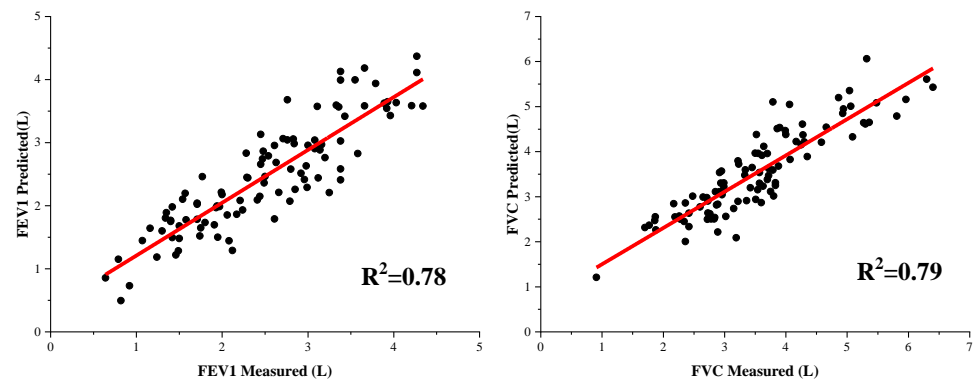


Figure 7. The fitting curves of the conventional machine learning results.

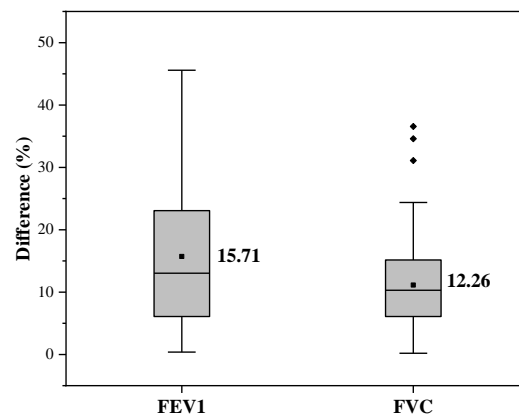


Figure 8. Box plot of error percentages of conventional machine learning.

We used a single-structured deep learning algorithm for pulmonary function prediction. The CO₂ sequences are processed by a one-dimensional convolutional neural network and the results are evaluated using relevant evaluation metrics.

The results of the deep learning algorithm are shown in Table 5. In all evaluation metrics, the prediction of FVC is better than FEV1. Especially in the RMSE metric, FVC is 0.61 L, less than the 0.66 L of FEV1, which is different from common medical knowledge. Because the deep learning algorithm is end-to-end learning, which only mines the raw sequence data for regression and does not provide any medical prior knowledge, there may be results that contradict prior medical knowledge. This result shows that FVC is better than FEV1 in terms of raw sequence distribution, so all the results obtained by regression

with a deep learning algorithm are better than FEV1, and probably since deep learning algorithms require larger datasets, they do not show good performance for the dataset used in this paper.

Table 5. Results of the deep learning algorithm.

Type	Pulmonary Function Parameters	RMSE (L)	R ²	ACC
1D-CNN	FEV1	0.66	0.57	65.09%
	FVC	0.61	0.73	74.76%

Figure 9 shows the training curves of the 1D-CNN network. We use a learning rate that is initially 0.03 and decays as the epoch increases. As can be seen from Figure 9, the network is trained normally and the model loss gradually decreases within 100 epochs without any phenomenon such as overfitting.

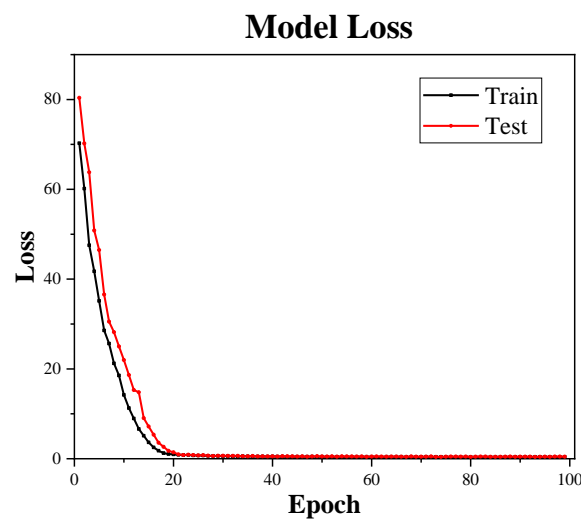


Figure 9. Training curves of the 1D-CNN network.

The fitting curves and error percentages of the deep learning algorithm results are shown in Figures 10 and 11. As can be seen in Figure 10, the prediction performance of FEV1 needs to be improved, with an R² of 0.57. Figure 11 shows that the average error of FEV1 prediction is 21.52%, and the average error of FVC is 14.19%. As can be seen from the figure, there are no outliers in the predicted results using the deep learning algorithm, which indicates that the deep learning algorithm can tap into the depth of information in the data and fit all the data distributions as much as possible, but its prediction performance needs to be improved.

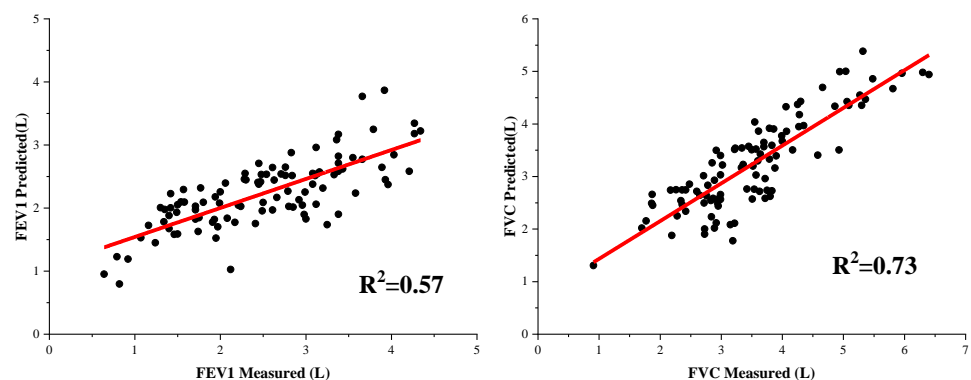


Figure 10. The fitting curves of the deep learning results.

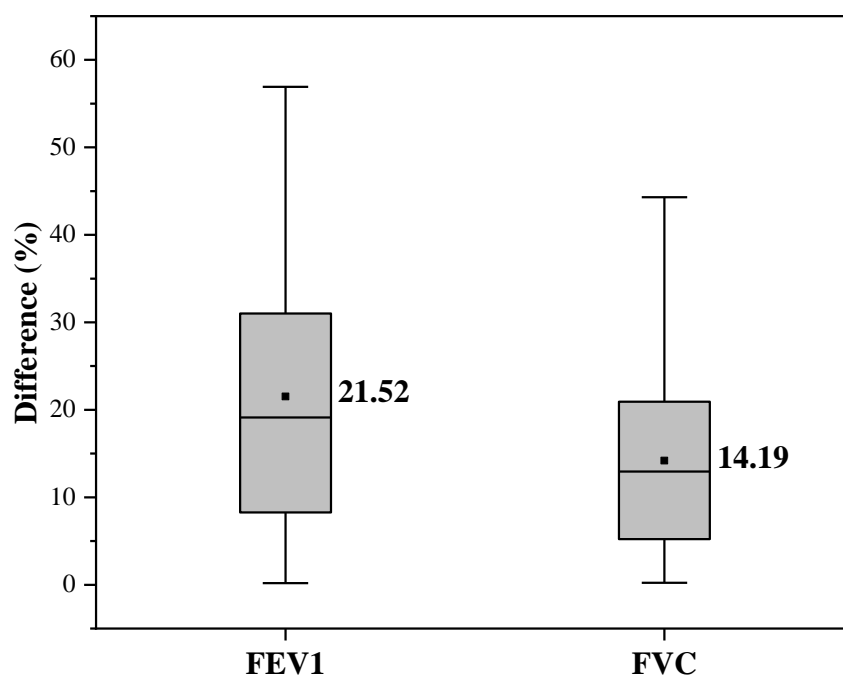


Figure 11. Box plot of error percentages of deep learning.

3.3.2. Combination Algorithm Results

We used combination algorithms to regress medical features using traditional machine learning algorithms and sequence features using deep learning algorithms, and finally, we combined the results through an error correction structure to perform a quadratic regression.

Table 6 shows the results of the combination algorithm. On the ACC index, both FEV1 and FVC were able to reach 80% and above, and the prediction of FVC was 85%, which is a good result. On the R^2 index, both FEV1 and FVC reached 0.85 and above. This indicates a good fit of the results. On the RMSE metric, the FVC prediction result of 0.39 is greater than that of 0.35 for FEV1, which may be because the combined algorithm combines the output of traditional machine learning algorithms and includes prior medical knowledge in the execution process.

Table 6. Results of the combination algorithm.

Type	Pulmonary Function Parameters	RMSE (L)	R^2	ACC
Combination algorithm	FEV1	0.35	0.85	80.79%
	FVC	0.39	0.86	85.77%

The fitting curves and error percentage of the combination algorithm’s results are shown in Figures 12 and 13. As can be seen from Figure 13, the average errors of both FEV1 and FVC are within 10%, which indicates that the pulmonary function prediction by the combined algorithm has high accuracy among most testers and has the potential for wide application. However, there is an outlier in the FEV1 prediction, which indicates that there is still room for optimization of the combined algorithm prediction for individual testers. In terms of the overall distribution of errors, the combined algorithm achieved a good result for both FEV1 and FVC.

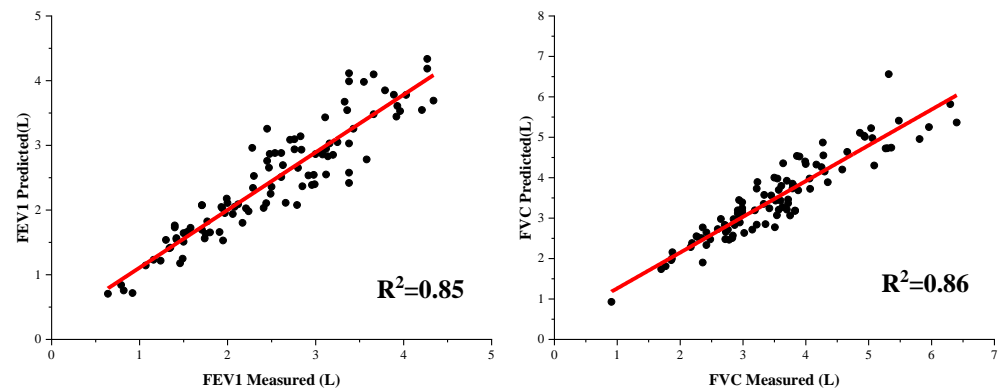


Figure 12. The fitting curves of the combination algorithm results.

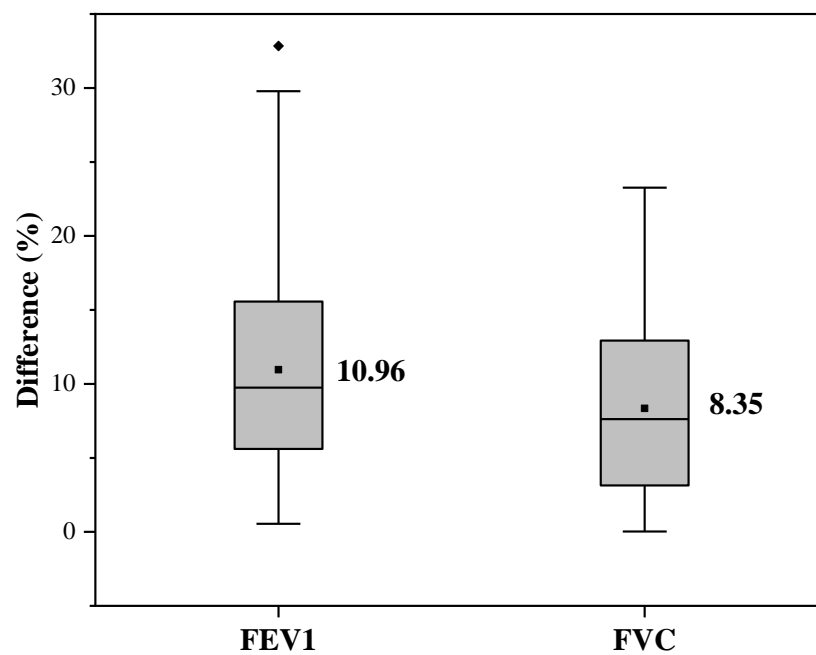


Figure 13. Box plot of error percentages of combination algorithm.

3.3.3. Comparison of Algorithms

- Comparison of experimental results

To further demonstrate the superiority of this combination algorithm, we compared the results of the single-structure algorithm with the combination algorithm on the same data set, as shown in Table 7.

Table 7. Results of the different algorithms.

Parameter Types	Algorithm Types	RMSE (L)	R ² (P)	MPE	MAPE	RMSPE	ACC
FEV1	SVM + XGBoost	0.43	0.78 (<0.01)	45.58%	15.71%	17.01%	73.90%
	1D-CNN	0.66	0.57 (0.02)	56.91%	21.51%	26.30%	65.09%
	combination algorithm	0.35	0.85 (<0.01)	32.84%	10.96%	13.83%	80.79%
FVC	SVM + XGBoost	0.48	0.79 (<0.01)	36.57%	12.26%	13.64%	79.18%
	1D-CNN	0.61	0.73 (<0.01)	44.30%	14.19%	17.22%	74.76%
	combination algorithm	0.39	0.86 (<0.01)	23.27%	8.35%	11.06%	85.77%

In Table 7, it can be seen that for the same pulmonary function parameter (FEV1 or FVC), the combined algorithm has the best prediction performance, followed by the traditional machine learning algorithm, and finally, the deep learning algorithm. The poor performance of the deep learning algorithm may be influenced by the small data set. Meanwhile, the processing method of the deep learning algorithm on the raw sequence information still needs further research. Traditional machine learning algorithms using medical features for pulmonary function parameter prediction are better overall but still need improvement in regard to MPE metrics. The combined algorithm performs optimally on this dataset, and all RMSE metrics less than 0.39 L. The R^2 metrics are greater than 0.85, indicating that the predicted values are strongly correlated with the actual values. The combined learning algorithm incorporates medical feature information and sequence feature information to a certain extent, which can reduce the MAPE and MPE metrics of the predicted values and means that the prediction results have a greater application range.

- Comparison with state-of-the-art performance

As shown in Table 8, our results were further compared with the relevant literature. In terms of the number of people in the dataset, we used data from 1007 subjects, second only to the 3567 in [12]. With regard to the R^2 index, this study obtained results greater than 0.85, which exceeds the results in the literature [11,13,14]. In terms of the RMSE metrics, our results also go beyond those in the literature [11,14]. In summary, the algorithm in this paper achieves high performance in the prediction of the relevant pulmonary function parameters.

Table 8. Performance comparison with other works.

Author	Subjects	Methodology	Result
Sharan et al. [11]	322	Linear and nonlinear regression models	A root mean square error (and correlation coefficient) for standard spirometry parameters FEV1, FVC, and FEV1/FVC of 0.593 L (0.810), 0.725 L (0.749), and 0.164 L (0.547).
Ioachimescu et al. [12]	3567	Regular linear or optimized regression, ANN models	The AEX could become an essential tool in assessing respiratory impairment.
Miyoshi et al. [13]	683	Multivariate linear regression analysis	Actual and estimated VC, FVC, and FEV1 values showed significant correlations (all $r > 0.8$ and $p < 0.001$) in all groups.
Chen et al. [14]	143	M-SVR	The mean squared errors were lower than $0.15 l^2$, and the decision coefficients (R^2) were higher than 0.40.
Ours	1007	SVM, XGBoost, 1D-CNN, KNN	The root mean squared errors (RMSE) were lower than 0.39 L. The coefficient of determinations (R^2) was higher than 0.85. The comprehensive percentage error (CPE) was lower than 20%.

4. Conclusions

In this paper, an algorithm combining traditional machine learning and deep learning was proposed to address the problem of the low accuracy of pulmonary function parameters in assessing respiratory diseases. The algorithm processes medical features by SVM and XGBoost algorithms to ensure the interpretability of the algorithm. The one-dimensional convolutional network is also used to analyze the CO_2 series to fully explore the deep features in the sequence, and then the improved KNN algorithm is used to combine the results both simply and effectively to improve the accuracy of pulmonary function parameters. This algorithm can significantly improve the accuracy of pulmonary function parameter prediction in the assessment stage of respiratory diseases.

The proposed combined algorithm was compared with the single-structure algorithm and showed improvement in all regression metrics. The root mean square error (RMSE) of the pulmonary function parameters predicted by the combination algorithm was less than 0.39 L and the R^2 was determined to be greater than 0.85 through a ten-fold cross-validation

experiment. The algorithm was compared with other algorithms for pulmonary function parameter prediction, and the method was able to better utilize medical and serial features to achieve significant results. In addition, unlike most methods, the method proposed in this paper utilizes carbon dioxide volume data, which can be a better alternative to spirometry.

However, the algorithm proposed in this paper also has some limitations. The performance of the proposed algorithm needs to be improved when extracting information on sequence features. Additionally, this paper mainly focuses on the field of pulmonary function parameter prediction, and further research is needed to apply the algorithm to other fields. There are still some problems that need to be overcome in the course of further research. For example, the number of testers in this dataset is limited, so more samples are needed to validate the performance of the algorithm. Additionally, multidimensional test data can be incorporated for a more accurate prediction of pulmonary function parameters from multimodal data.

Author Contributions: Conceptualization, R.Z.; investigation, T.Y. and Q.Z.; software, P.W. and Y.L.; writing—original draft preparation, X.M.; writing—review and editing, Z.Z., X.C., L.D. and Z.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Project 2020YFC2003703, 2020YFC1512304, 2018YFC2001101, 2018YFC2001802, CAMS Innovation Fund for Medical Sciences (2019-I2M-5-019), and National Natural Science Foundation of China (Grant 62071451).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not applicable.

Acknowledgments: We wish to express our gratitude to doctors from the China-Japan Friendship Hospital for their help during the recording and analysis of patient data. We would like to thank the anonymous reviewers for their constructive comments and recommendations.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Aung, Y.Y.M.; Wong, D.C.S.; Ting, D.S.W. The promise of artificial intelligence: A review of the opportunities and challenges of artificial intelligence in healthcare. *Br. Med. Bull.* **2021**, *139*, 4–15. [[CrossRef](#)] [[PubMed](#)]
2. Chen, J.; See, K.C. Artificial Intelligence for COVID-19: Rapid Review. *J. Med. Internet Res.* **2020**, *22*, e21476. [[CrossRef](#)] [[PubMed](#)]
3. Alade, O.M.; Sowunmi, O.Y.; Misra, S.; Maskeliūnas, R.; Damaševičius, R. A neural network based expert system for the diagnosis of diabetes mellitus. In *International Conference on Information Technology Science*; Springer: Cham, Switzerland, 2017.
4. Azeez, N.A.; Towolawi, T.; Vyver, C.V.D.; Misra, S.; Adewumi, A.; Damaševičius, R.; Ahuja, R. A fuzzy expert system for diagnosing and analyzing human diseases. In *International Conference on Innovations in Bio-Inspired Computing and Applications*; Springer: Cham, Switzerland, 2018.
5. Azeez, N.A.; Oyeniran, S.O.; Vyver, C.V.D.; Misra, S.; Ahuja, R.; Damasevicius, R.; Maskeliunas, R. Diagnosing oral ulcers with Bayes model. In *International Conference on Innovations in Bio-Inspired Computing and Applications*; Springer: Cham, Switzerland, 2018.
6. Hui, B.; Liu, Y.; Qiu, J.; Cao, L.; Ji, L.; He, Z. Study of texture segmentation and classification for grading small hepatocellular carcinoma based on CT images. *Tsinghua Sci. Technol.* **2020**, *26*, 199–207. [[CrossRef](#)]
7. Tan, C.L.; Chan, Y.; Candasamy, M.; Chellian, J.; Madheswaran, T.; Sakthivel, L.P.; Patel, V.K.; Chakraborty, A.; MacLoughlin, R.; Kumar, D.; et al. Unravelling the molecular mechanisms underlying chronic respiratory diseases for the development of novel therapeutics via in vitro experimental models. *Eur. J. Pharmacol.* **2022**, *919*, 174821. [[CrossRef](#)]
8. Hashimoto, N.; Wakahara, K.; Sakamoto, K. The importance of appropriate diagnosis in the practical management of chronic obstructive pulmonary disease. *Diagnostics* **2021**, *11*, 618. [[CrossRef](#)]
9. Barnes, T.; Fromer, L. Spirometry use: Detection of chronic obstructive pulmonary disease in the primary care setting. *Clin. Interv. Aging* **2011**, *6*, 47–52. [[CrossRef](#)]
10. Schermer, T.R.; Jacobs, J.E.; Chavannes, N.; Hartman, J.; Folgering, H.T.; Bottema, B.J.; van Weel, C. Validity of spirometric testing in a general practice population of patients with chronic obstructive pulmonary disease (COPD). *Thorax* **2003**, *58*, 861–866. [[CrossRef](#)]
11. Sharan, R.V.; Abeyratne, U.R.; Swarnkar, V.R.; Claxton, S.; Hukins, C.; Porter, P. Predicting spirometry readings using cough sound features and regression. *Physiol. Meas.* **2018**, *39*, 095001. [[CrossRef](#)]

12. Ioachimescu, O.C.; Stoller, J.K.; Garcia-Rio, F. Area under the expiratory flow-volume curve: Predicted values by artificial neural networks. *Sci. Rep.* **2020**, *10*, 16624. [[CrossRef](#)]
13. Miyoshi, S.; Katayama, H.; Matsubara, M.; Kato, T.; Hamaguchi, N.; Yamaguchi, O. Prediction of spirometric indices using forced oscillometric indices in patients with asthma, COPD, and interstitial lung disease. *Int. J. Chronic Obstr. Pulm. Dis.* **2020**, *15*, 1565. [[CrossRef](#)]
14. Chen, J.; Yang, Z.; Yuan, Q.; Xiong, D.-X.; Guo, L.-Q. Prediction models for pulmonary function during acute exacerbation of chronic obstructive pulmonary disease. *Physiol. Meas.* **2020**, *41*, 125010. [[CrossRef](#)] [[PubMed](#)]
15. Kremeier, P.; Böhm, S.H.; Tusman, G. Clinical use of volumetric capnography in mechanically ventilated patients. *Int. J. Clin. Monit. Comput.* **2019**, *34*, 7–16. [[CrossRef](#)] [[PubMed](#)]
16. Jarenbäck, L.; Tufvesson, E.; Ankerst, J.; Bjermer, L.; Jonson, B. The Efficiency Index (EFFi), based on volumetric capnography, may allow for simple diagnosis and grading of COPD. *Int. J. Chronic Obstr. Pulm. Dis.* **2018**, *13*, 2033. [[CrossRef](#)] [[PubMed](#)]
17. Kellerer, C.; Schneider, A.; Klütsch, K.; Husemann, K.; Sorichter, S.; Jörres, R.A. Correspondence between Capnovolumetric and Conventional Lung Function Parameters in the Diagnosis of Obstructive Airway Diseases. *Respiration* **2020**, *99*, 389–397. [[CrossRef](#)] [[PubMed](#)]
18. Rosli, M.H.B.M.; Kumarasamy, R.; Malarvili, M.B. Design of Device to Monitor Asthma Severity Using Mainstream Technology while Administering Medication. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *884*, 012010. [[CrossRef](#)]
19. Wu, H.-C.; Gupta, N.; Mylavarapu, P.S. Blind multiridge detection for automatic nondestructive testing using ultrasonic signals. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **2006**, *53*, 1902–1911. [[CrossRef](#)]
20. Mood, A.M.; Graybill, F.A.; Boes, D.C. *Introduction to the Theory of Statistics*, 3rd ed.; McGraw-Hill Education: New York, NY, USA, 1974.
21. Gao, J.; Li, F.; Wang, B.; Liang, H. Unsupervised nonlinear adaptive manifold learning for global and local information. *Tsinghua Sci. Technol.* **2021**, *26*, 163–171. [[CrossRef](#)]
22. Kiasadegh, M.; Emdad, H.; Ahmadi, G.; Abouali, O. Transient numerical simulation of airflow and fibrous particles in a human upper airway model. *J. Aerosol Sci.* **2019**, *140*, 105480. [[CrossRef](#)]
23. Arora, P.; Boyne, D.; Slater, J.J.; Gupta, A.; Brenner, D.R.; Druzdzel, M.J. Bayesian Networks for Risk Prediction Using Real-World Data: A Tool for Precision Medicine. *Value Health* **2019**, *22*, 439–445. [[CrossRef](#)]
24. Karim, R.; Beyan, O.; Zappa, A.; Costa, I.G.; Rebholz-Schuhmann, D.; Cochez, M.; Decker, S. Deep learning-based clustering approaches for bioinformatics. *Brief. Bioinform.* **2020**, *22*, 393–415. [[CrossRef](#)]
25. Boehnke, N.; Hammond, P.T. Power in Numbers: Harnessing Combinatorial and Integrated Screens to Advance Nanomedicine. *JACS Au* **2021**, *2*, 12–21. [[CrossRef](#)] [[PubMed](#)]
26. Halpin, D.M.G.; Criner, G.J.; Papi, A.; Singh, D.; Anzueto, A.; Martinez, F.J.; Agusti, A.A.; Vogelmeier, C.F. Global Initiative for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease. The 2020 GOLD Science Committee Report on COVID-19 and Chronic Obstructive Pulmonary Disease. *Am. J. Respir. Crit. Care Med.* **2021**, *203*, 24–36. [[CrossRef](#)] [[PubMed](#)]
27. Chauhan, V.K.; Dahiya, K.; Sharma, A. Problem formulations and solvers in linear SVM: A review. *Artif. Intell. Rev.* **2018**, *52*, 803–855. [[CrossRef](#)]
28. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016.
29. Meng, Y.; Yang, N.; Qian, Z.; Zhang, G. What makes an online review more helpful: An interpretation framework using XGBoost and SHAP values. *J. Theor. Appl. Electron. Commer. Res.* **2020**, *16*, 466–490. [[CrossRef](#)]
30. Cherkassky, V.; Ma, Y. Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Netw.* **2004**, *17*, 113–126. [[CrossRef](#)]
31. Aghezzaf, B.; El Fahim, H. Iterated local search algorithm for solving the orienteering problem with soft time windows. *SpringerPlus* **2016**, *5*, 1–36. [[CrossRef](#)]
32. Tanabe, T.; Ye, J.; Suzuyama, T.; Kobayashi, T.; Yamaguchi, Y.; Yasuda, M. Potential for improving the local realization of coordinated universal time with a convolutional neural network. *Rev. Sci. Instrum.* **2019**, *90*, 125111. [[CrossRef](#)]
33. Abo-Tabik, M.; Costen, N.; Darby, J.; Benn, Y. Towards a Smart Smoking Cessation App: A 1D-CNN Model Predicting Smoking Events. *Sensors* **2020**, *20*, 1099. [[CrossRef](#)]
34. Ma, D.; Shang, L.; Tang, J.; Bao, Y.; Fu, J.; Yin, J. Classifying breast cancer tissue by Raman spectroscopy with one-dimensional convolutional neural network. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2021**, *256*, 119732. [[CrossRef](#)]
35. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *NIPS* **2012**, *60*, 84–90. [[CrossRef](#)]
36. Zhuang, J.; Dong, Y.; Bai, H.; Zuo, P.; Cheng, J. Auto-Selecting Receptive Field Network for Visual Tracking. *IEEE Access* **2019**, *7*, 157449–157458. [[CrossRef](#)]
37. Zhou, R.; Wang, C.; Zhang, P.; Chen, X.; Du, L.; Wang, P.; Zhao, Z.; Du, M.; Fang, Z. ECG-based biometric under different psychological stress states. *Comput. Methods Programs Biomed.* **2021**, *202*, 106005. [[CrossRef](#)] [[PubMed](#)]
38. Wu, X.Y.; Wang, S.H.; Zhang, Y.D. Survey on theory and application of k-Nearest-Neighbors algorithm. *Comput. Eng. Appl.* **2017**, *53*, 1–7.

-
39. Pu, G.; Wang, L.; Shen, J.; Dong, F. A hybrid unsupervised clustering-based anomaly detection method. *Tsinghua Sci. Technol.* **2021**, *26*, 146–153. [[CrossRef](#)]
 40. Burmyakov, A. Schedulability Analysis of Multiprocessor Real-time Systems Using Pruning. Ph.D. Thesis, Universidade do Porto, Porto, Portugal, 2016.