

LSTM-Based Time Series Forecasting of Pulmonary Function Test for COPD Early Diagnosis

Beomseo Choi*, Hongjun Kim*, Seung Hyun Jeon^o

ABSTRACT

Chronic Obstructive Pulmonary Disease (COPD) is a serious lung disease that makes breathing difficult and cannot be easily detected. Even though early diagnosis technology for COPD using machine learning has been developed, Pulmonary Function Test (PFT) data-based time series prediction studies are still lacking. We use PFT data with insufficient measurement intervals, propose a Long Short-Term Memory (LSTM) to predict PFT values for the future 1Q from the past 2Q, and classify whether COPD occurs or not. The data were interpolated to resolve the imbalanced time period. To confirm the validity of the augmented data, Multivariate Analysis of Variance (MANOVA) was performed, and through the rigorous MANOVA, we proved that there was no significant difference between the original and interpolated data. Mean Absolute Percentage Error (MAPE), recalls, and F1 scores, which are the harmonic mean of precision and recall for classification, were measured for two test scenarios: only the original data and the augmented data. Finally, we found the interpolated data decreased MAPE by almost 7%, however, improved recall and F1 score by almost 22% and 12% for obstructive pulmonary disease, compared with the original data. Besides, we can predict COPD within 3 months, irrelevant to smokers and non-smokers.

Key Words : Chronic Obstructive Pulmonary Disease, Early Diagnosis, Pulmonary Function Test, Long Short-Term Memory, Interpolation

I. Introduction

Chronic Obstructive Pulmonary Disease (COPD) is an irreversible chronic lung disease that narrows the airways over a long period of time. According to the World Health Organization (WHO), COPD is the third leading cause of death globally in 2019 and is responsible for approximately 6% of all deaths. COPD is rare in low-income countries, but it ranks in the top five in all foreign countries. COPD is caused by several risk factors, including exposure to smoking and air pollution. COPD cannot be cured in a short period of time. Early diagnosis of the disease and

prompt treatment play an important role in reducing mortality due to COPD.

Early symptoms of COPD include chronic cough and phlegm, fatigue, and shortness of breath. The problem with early diagnosis of COPD is that it is difficult to detect because the early symptoms are not clear. In modern times, chest X-ray, Pulmonary Function Test (PFT), Chest Computed Tomography (CT), and Arterial Blood Gas Analysis (ABGA) are used to diagnose COPD^[1]. Among them, PFT can easily obtain data in terms of low inspection cost and short inspection time. However, in the prior research for early diagnosis of COPD, authors mainly conducted image analysis such as CT. H. Park et al.

※ This work was supported by the ICT R&D program of MSIT/IITP(2022-0-00849).

♦ First Author : Daejeon University, Department of Computer Engineering, beomseo0707@gmail.com, 학생회원

o Corresponding Author : Daejeon University, Department of Computer Engineering, creemur@dju.kr, 정회원

* Daejeon University, Department of Computer Engineering, hjkim99@dju.kr

논문번호 : 202310-119-A-RU, Received October 10, 2023; Revised November 10, 2023; Accepted November 14, 2023

conducted a study predicting spirometry from CT images. They classified high-risk participants by spirometry values and used a Convolutional Neural Network (CNN)^[2].

Although classification research through machine learning is active, time series analysis research using PFT time series data is still lacking. We use intermittent PFT time series data from multiple tentative patients to predict Forced Vital Capacity (FVC) and Forced Expiratory Volume in 1 second (FEV₁) and types of ventilatory disorders. However, an imbalance of measurement intervals per patient should be solved for time series analysis.

In this paper, we propose a Long Short-Term Memory (LSTM) based COPD prediction framework to diagnose the patient's COPD within the next 1 quarter (Q). First, to solve the problem of the data observation interval's inconsistency, the training data were downsampled based on a Q unit using preprocessing and augmented with fill and interpolation. We test the validity of the augmented data using Multivariate Analysis of Variance (MANOVA), and then confirm that there was no difference between the augmented and the original test data. To verify the MANOVA results, two scenarios with the augmented and original test data are presented. We predict the future 1Q based on the well-refined training data during the past 2Q. Thus, we improve the performance of the augmented version such as Mean Absolute Percentage Error (MAPE) and F1 score as follows: 7% reduction and 4% enhancement, respectively. Among ventilatory disorders, in the case of obstructive, recall and F1 score for the augmented test data improved by 22% and 12%, respectively.

II. Related Works

A lot of research has been conducted to diagnose lung disease using artificial intelligence (AI) technologies^[3]. In this section, we describe the previous research on machine learning and deep learning related to COPD, as well as interpolation for data preprocessing.

2.1 AI Approaches to Predict COPD Diagnosis

L. Beverin et al. showed high performance in predicting lung disease based on machine learning using PFT data. The authors predicted Total Lung Capacity (TLC) using Random Forest (RF)^[4]. The study used PFT data. As a result of the study, the sensitivity, specificity, and F1 score of the algorithm predicting restrictive ventilatory impairment were 83, 92, and 75%, respectively. This study uses similar features to ours. However, they cannot predict COPD using the proposed RF model. However, D. Spathis and P. Vlamos classified COPD using RF^[5], which shows a precision of 97.7%. The authors used PFT and ABGA data. This paper revealed that smoking, FVC, FEV₁, and age are important factors for COPD through the feature importance of RF. They can predict COPD at the time the patient is tested. However, even if the patient does not have COPD at the time of measurement, COPD may appear in the future if the patient's condition is worsening. Since we use PFT time series data, the nearest future onset of COPD can be predicted by considering changes in the patient's condition. There are studies using the LSTM model to study early diagnosis of COPD. V. Nunavath et al. predicted the health status of COPD patients^[6]. The authors used an LSTM model based on ABGA data. The LSTM model was learned based on data during the past 5 days and showed an accuracy of 84.12% in predicting the patient's health status one day in advance. This approach does not distinguish whether patients have COPD or not. D. Perna and A. Tagarelli proposed a learning framework using respiratory sound data and Recurrent Neural Network (RNN)-based LSTM, Bidirectional LSTM (BiLSTM), Gated Recurrent Unit (GRU), and Bidirectional GRU (BiGRU)^[7]. Among the four models, LSTM consistently showed better. Thus, we consider choosing the LSTM-based COPD research.

2.2 Interpolation Approaches to Augment Insufficient COPD Data

Recent research has improved performance through interpolation in insufficient situations of input data. O. O. Abayomi-Alli et al. used biomedical voice

measurement, the Oxford Parkinson Disease dataset, and BiLSTM for early detection of Parkinson's disease^[8]. The research presents interpolation to augment the small dataset. The interpolation methods used were cubic spline and Piecewise Cubic Hermite Interpolating Polynomial (PCHIP). PCHIP creates a cubic Hermite interpolating polynomial from data points in the data interval. Each piece is monotonic and is characterized by smoothly connecting data points between data intervals^[9]. O. O. Abayomi-Alli et al. argue that the main limitation of interpolation is that it produces out-of-range and noisy data. In this paper, instead of using cubic spline interpolation, linear interpolation and PCHIP interpolation were used to solve the problem, because linear interpolation and PCHIP interpolation are both monotonic. H. Watz et al. used interpolation for statistical analysis of COPD^[10]. The authors used data from patients who completed lung capacity measurements daily for 56 days, at least once per week, for COPD postmortem analysis. Missing values in the data were filled in using the linear interpolation, fill, and carry forward methods. FEV₁ continuously and smoothly decreases as life continues^[11]. Thus, two interpolations are presented in this paper. The first is linear interpolation, and the second is PCHIP. Cubic spline interpolation is not considered in this paper because negative numbers may occur. Since FEV₁ does not shake, it cannot be negative, and shows a gradual pattern, we adopt linear interpolation and PCHIP, which are monotonic interpolations.

III. System Model

3.1 Observed Data

The dataset was provided by Chungnam National University Hospital and collected between January 1, 2020, and July 31, 2022. Table 1 shows the features of the proposed framework. One-hot encoding was performed for sex.

PFT proceeds with three stages: inhale with maximum effort, exhale with maximum effort, and breathe in again with maximum effort. During PFT, flow and volume are measured and expressed as a time-volume curve and volume-flow curve. The two

Table 1. The features of the proposed framework.

Features	Interpolation	Units
Age	N/A	Year
Sex		Male and female
Pack year		The number of packs of cigarettes smoked per day multiplied by the number of years of smoking.
Height	Linear interpolation	cm
Weight		kg
FVC	PCHIP	L
FEV ₁		L
FEF _{25-75%}		L/sec
PEF		L/sec

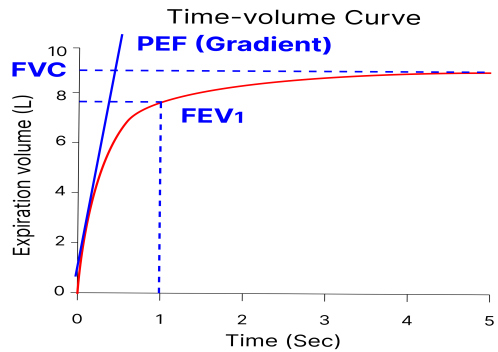


Fig. 1. Time-volume curve results from PFT.

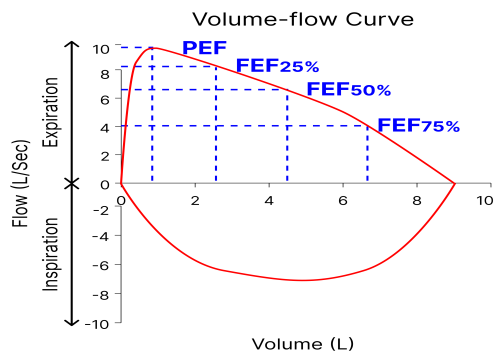


Fig. 2. Volume-flow curve results from PFT.

curves show FVC, FEV₁, Forced Expiratory Flow (FEF), and Peak Expiratory Flow (PEF). FVC refers to the volume of air when you inhale as much as possible and then exhale all the way with maximum effort. FEV₁ refers to the amount of air exhaled with

maximum effort in 1 second after starting to inhale and exhale with maximum effort. $FEF_{25-75\%}$ refers to the average airflow between 25% and 75% of FVC. PEF is the maximum airflow achieved during exhalation with maximal effort. FVC, FEV_1 , and PEF are shown in Fig. 1 showing the time-volume curve. PEF and FEF are shown in Fig. 2 showing the volume-flow curve. FVC% and FEV_1/FVC can be used to distinguish types of ventilatory disorders^[12]. FVC% is FVC (measured)/ FVC (predicted). FEV_1/FVC is FEV_1 (measured)/ FVC (measured). Predictions are calculated by spirometric reference equations. We used Morris's reference equation^[13]. The criteria for classification of ventilation disorders are shown in Fig. 3. Restrictive is a symptom of decreased TLC. Therefore, it shows a decrease in FVC. Obstructive is a symptom of narrowing of the airway. Mixed shows both restrictive and obstructive symptoms.

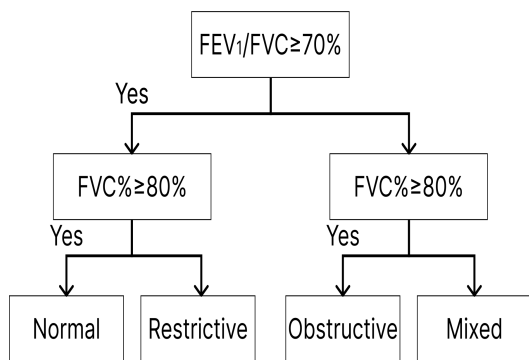


Fig. 3. The criteria for classification of ventilation disorders.

3.2 Preprocessing

This section describes a method for keeping the intervals of time series data constant and a method for handling missing values. The proposed framework is shown in Fig. 4. The dataset has different intervals because PFT data were obtained regardless of smokers and non-smokers. Accordingly, to solve this problem, we assume observations in Q units and perform downsampling at Q intervals. If the PFT frequency increases in the future, downsampling can be performed on a monthly basis shorter than quarterly.

The fill method was used for the sex and pack year

features because there is only one value in one patient. Linear interpolation does not reflect the characteristic of the age, which increases by one year with each birthday. However, since there was no information on the patient's birthday, the patient's birthday was assumed to be January 1st. Therefore, every January 1st, the age is increased by one year. Min-max normalization was performed.

The number of the original and augmented data is 1,408 and 1,446, respectively. The split ratio of the original data and the augmented data is 49:51. We validate the use of augmented data to predict COPD. For the original data group and the augmented data group, pack year, age, height, weight, FVC, FEV_1 , $FEF_{25-75\%}$, and PEF are analyzed using MANOVA. We set alpha to 0.05, and the results are shown in Table 2.

As a result of MANOVA analysis, the p-value was 0.1083, which is larger than the alpha value. The null hypothesis that "the overall vector averages of the two groups are the same" cannot be rejected, and then there is no significant difference between the two groups. To perform MANOVA, multivariate normality and multivariate homoscedasticity must be satisfied. If the absolute value of skewness is greater than 3 or the absolute value of kurtosis is greater than 10, there is a problem with normality^[14]. The skewness of the original data is between approximately -1.22 and 1.45, and the kurtosis is approximately between -0.38 and 2.25. The skewness of the augmented data is between approximately -1.22 and 1.46, and the kurtosis is approximately between

Table 2. The results of MANOVA for raw dataset and augmented dataset. *Num DF* is the Numerator Degrees of Freedom, *Den DF* is the Denominator Degrees of Freedom.

Methods	Statistic Values	F Values	Num DF	Den DF	Pr (>F)
Wilks' lambda	0.99541	1.64	8	2845	0.1083
Pillai's trace	0.0045903	1.64	8	2845	0.1083
Roy's largest root	0.0046115	1.64	8	2845	0.1083
Hotelling's trace	0.0046115	1.64	8	2845	0.1083

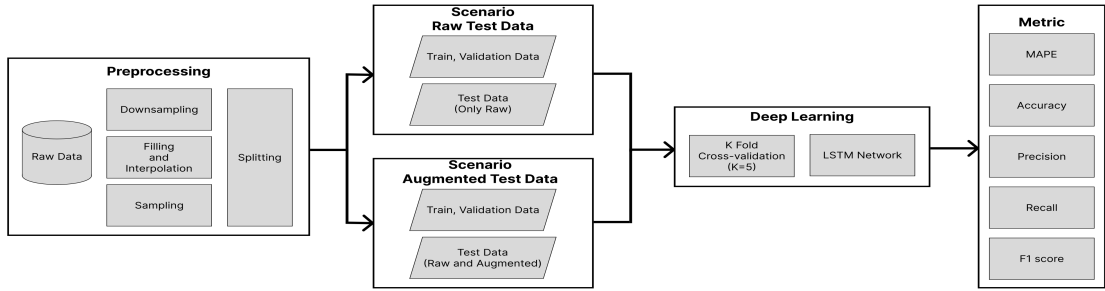


Fig. 4. LSTM-based COPD prediction framework.

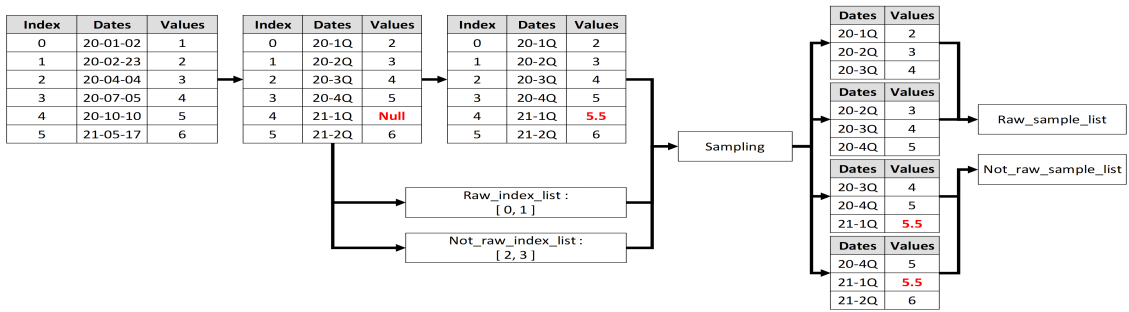


Fig. 5. An example of the process of preprocessing and sampling.

-0.50 and 3.16. Therefore, we can ensure that multivariate normality is satisfied.

Box’s M Test was performed to test multivariate homoscedasticity. We set alpha to 0.001. The p-value of Box’s M Test is 0.02904, which is larger than the alpha value. Since the null hypothesis that variances between multivariate groups are equal cannot be rejected, homoscedasticity is satisfied. The detailed results of the Box’s M Test are shown in Table 3.

Table 3. The results of Box’s M Test for raw dataset and augmented dataset.

Box’s M Test		
Chi-Square	Degrees of Freedom	p-value
53.718	36	0.02904

3.3 Proposed LSTM-Based COPD Forecasting Framework

We aim to put in input data at time T and T-1 and to get output data at time T+1. In other words, we predict future 1Q data with past 2Q data. Fig. 5 shows an example of the preprocessing process. This paper downsamples irregular time series data. For data merged due to downsampling, the value of the last

data is used. To extract samples to be used in LSTM, the sample size was set to 3, which is the sum of the past 2Q and the future 1Q, and sampling was performed by sliding. The entire observed data includes augmented data. To distinguish between augmented data and original data, if there are no missing values in the sample, the starting index of the sample is stored in raw_index_list. Otherwise, it is stored in not_raw_index_list. And then the missing values of the downsampled data are filled by interpolation or filling. The example in Fig. 5 used linear interpolation. Then, sampling is performed using the index information in raw_index_list and not_raw_index_list.

Here, we define two test scenarios and conduct experiments.

- Augmented Test Data (ATD): Both augmented data and original data are used without distinction. 20% of the total is used as test data, and the remaining 80% is used as training and validation data.
- Raw Test Data (RTD): the augmented data is used as training and validation data, and the original data is used as test data.

Even though there was no significant difference between the original data and the augmented data, there was a difference in the samples of ATD and RTD. This difference arises because the method of recognizing raw samples during the sample extraction process is quite restricted. The number and ratio of samples are shown in Table 8 in the Appendix.

We apply a stratified split scheme to split the data into training and validation data and test data in ATD. As RTD only uses raw data as test data, we cannot apply the stratified split for RTD. To split training and validation data, stratified K-fold cross-validation is performed. Stratified techniques can reduce bias when splitting data evenly or evaluating model performance.

Here, we apply K=5, and the average evaluation of each fold was used as the result. MAPE was used as a regression metric. Accuracy, precision, recall, and F1 score were used as classification metrics. Mean Squared Error (MSE) was used as a loss metric.

$$MSE = \frac{1}{n} \sum_{i=1}^n (A_i - F_i)^2, \quad (1)$$

where A_i is an actual value, F_i is a forecast value, and n is a sample size.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{A_i - F_i}{A_i} \right|. \quad (2)$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}, \quad (3)$$

where TP is a true positive, FP is a false positive, TN is a true negative, and FN is a false negative.

$$Precision = \frac{TP}{TP + FP}. \quad (4)$$

$$Recall = \frac{TP}{TP + FN}. \quad (5)$$

F1 score, which is used as the harmonic mean of precision and recall for COPD classification, is expressed by (4) and (5).

$$F1\ score = \frac{2 \times Precision \times Recall}{Precision + Recall}. \quad (6)$$

IV. Experimental Results

This section summarizes the hyperparameters in Table 4 and presents test results of each scenario in Tables 5, 6, and 7.

Table 5 shows the MAPE for each age group for the two test scenarios. The ATD showed a lower error compared to the RTD. The total average MAPE decreased by 7%. Here we describe MAPE with the lower error as a boldface type.

Table 6 shows the accuracies by age group for the two test scenarios. The ATD showed higher accuracy compared to the RTD. Here, we describe higher accuracy as a boldface type.

Table 4. Hyperparameters for ATD and RTD.

Hyperparameters	ATD	RTD
Batch	32	
Epoch	500	1000
Layer	2	
Learning rate	0.001	
Sequence length	2	
Loss function	MSE	
Optimizer	Adam	

Table 5. MAPE by age groups and ventilatory disorders for ATD and RTD.

Groups	ATD		RTD	
	FVC MAPE	FEV ₁ MAPE	FVC MAPE	FEV ₁ MAPE
Total	0.03	0.04	0.1	0.11
20s	0.04	0.04	0.02	0.05
30s	0.03	0.04	0.04	0.04
40s	0.06	0.06	0.06	0.08
50s	0.03	0.03	0.18	0.14
60s	0.03	0.04	0.12	0.11
70s	0.03	0.05	0.11	0.16
80s	0.02	0.03	0.05	0.03
Normal	0.03	0.03	0.08	0.08
Restrictive	0.06	0.07	0.13	0.11
Obstructive	0.02	0.04	0.08	0.11
Mixed	0.05	0.06	0.12	0.21

Table 6. Accuracy of classification by age groups for ATD and RTD.

Age groups	ATD	RTD
Total	0.91	0.86
20s	0.77	0.67
30s	1	1
40s	0.9	0.88
50s	0.87	0.5
60s	0.91	0.89
70s	0.9	0.84
80s	0.96	1

Table 7 shows the precisions, recalls, and F1 scores of the two test scenarios. The ATD showed higher performance compared to the RTD. Here, we describe higher precisions, recalls, and F1 scores as a boldface type.

In the medical field, recall is considered important. This is to reduce cases where actual positive patients are judged negative. Besides, medical data have severe imbalances between classes. Therefore, F1 scores are also effective. In classification, the recall and F1 score of obstructive increased by 22% and 12%, respectively.

Through the results of the ATD and the RTD, the error was reduced when using augmented data as a test. In principle, it is correct to use only the original data as a test, but in the case of the proposed experiment, MANOVA confirmed that there was no

difference between the two groups, and even though the measurement intervals of PFT data are unstable, we prove meaningful results can be sufficiently confirmed for COPD prediction even when augmented data is used.

Compared to ATD, which includes augmented samples as test data, RTD has less test data. Therefore, the difference in the number of test data for the two scenarios led to different results. Additionally, because ATD has low epochs for train and many test data against RTD, and the distributions for training and validation data and test data are similar for each class, ATD achieves better performance than RTD.

V. Conclusion

By examining the outcomes from both the ATD and RTD, we have confirmed that interpolation data, whose availability has been verified by MANOVA for PFT time series data, are reliable and can lead to better performance in MAPE, recall, and F1 score. PFT is relatively simple and inexpensive compared to other tests for early diagnosis of COPD. However, because the health status and severity of each patient were different, the measurement intervals were not consistent. Besides, it was not easy to obtain sufficient PFT data to predict COPD. Nevertheless, by obtaining reliable COPD predictions in terms of recall and F1 scores through PFT data, interpolation can provide medical staff with reliable reference prediction results compared to traditional COPD prediction judgments through the naked eye or relying on expensive tests for seriously ill patients.

Table 7. Precisions, recalls, and F1 scores for ATD and RTD.

Scenarios	Type	Precision	Recall	F1 score
ATD	Normal	0.94	0.92	0.93
	Restrictive	0.90	0.75	0.82
	Obstructive	0.90	0.95	0.93
	Mixed	0.87	0.87	0.87
	Average	0.90	0.87	0.89
RTD	Normal	0.89	0.91	0.90
	Restrictive	0.79	0.87	0.83
	Obstructive	0.91	0.73	0.81
	Mixed	0.80	0.91	0.85
	Average	0.85	0.86	0.85

VI. Appendix

Table 8. The number and ratio of samples used in LSTM. A restrictive sample in the age group '30s' was removed from the ATD. Because there was only one data, the stratified split could not be applied.

Age groups	Type	ATD					RTD				
		The Number of Total	The Number of Training and Validation	The Number of Test	The ratio between Training and Validation (Left) and Test (Right)		The Number of Total	The Number of Training and Validation	The Number of Test	The ratio between Training and Validation (Left) and Test (Right)	
20s	Normal	4	3	1	75	25	4	4	0	100	0
	Restrictive	8	6	2	75	25	8	7	1	87.5	12.5
	Obstructive	6	5	1	83.33	16.67	6	5	1	83.33	16.67
	Mixed	16	13	3	81.25	18.75	16	15	1	93.75	6.25
30s	Normal	20	16	4	80	20	20	18	2	90	10
	Restrictive	0	0	0	0	0	1	0	1	0	100
	Obstructive	0	0	0	0	0	0	0	0	0	0
	Mixed	0	0	0	0	0	0	0	0	0	0
40s	Normal	57	46	11	80.7	19.3	57	55	2	96.49	3.51
	Restrictive	23	19	4	82.61	17.39	23	18	5	78.26	21.74
	Obstructive	13	10	3	76.92	23.08	13	12	1	92.31	7.69
	Mixed	12	9	3	75	25	12	12	0	100	0
50s	Normal	90	72	18	80	20	90	88	2	97.78	2.22
	Restrictive	21	17	4	80.95	19.05	21	21	0	100	0
	Obstructive	48	38	10	79.17	20.83	48	48	0	100	0
	Mixed	23	18	5	78.26	21.74	23	23	0	100	0
60s	Normal	207	165	42	79.71	20.29	207	202	5	97.58	2.42
	Restrictive	64	51	13	79.69	20.31	64	62	2	96.88	3.13
	Obstructive	215	172	43	80	20	215	208	7	96.74	3.26
	Mixed	74	59	15	79.73	20.27	74	73	1	98.65	1.35
70s	Normal	166	133	33	80.12	19.88	166	158	8	95.18	4.82
	Restrictive	75	60	15	80	20	75	73	2	97.33	2.67
	Obstructive	317	254	63	80.13	19.87	317	312	5	98.42	1.58
	Mixed	109	87	22	79.82	20.18	109	103	6	94.5	5.5
80s	Normal	97	78	19	80.41	19.59	97	94	3	96.91	3.09
	Restrictive	4	3	1	75	25	4	4	0	100	0
	Obstructive	153	123	30	80.39	19.61	153	153	0	100	0
	Mixed	61	49	12	80.33	19.67	61	60	1	98.36	1.64

References

- [1] D. D. Sin, "The importance of early chronic obstructive pulmonary disease: A lecture from 2022 asian pacific society of respirology," *Tuberculosis and Respiratory Diseases*, vol. 86, no. 2, pp. 71-81, Apr. 2023. (<https://doi.org/10.4046/trd.2023.0005>)
- [2] H. Park, et al., "Deep learning-based approach to predict pulmonary function at chest CT," *Radiology*, vol. 307, no. 2, pp. 221488, Apr. 2023. (<https://doi.org/10.1148/radiol.221488>)
- [3] S. Gonem, et al., "Applications of artificial intelligence and machine learning in respiratory medicine," *Thorax*, vol. 75, no. 8, pp. 695-701, Aug. 2020. (<https://doi.org/10.1136/thoraxjnl-2020-214556>)
- [4] L. Beverin, et al., "Predicting total lung capacity from spirometry: A machine learning approach," *Frontiers in Med.*, vol. 10, pp. 1174631, May 2023. (<https://doi.org/10.3389/fmed.2023.1174631>)
- [5] D. Spathis and P. Vlamos, "Diagnosing asthma and chronic obstructive pulmonary disease with machine learning," *Health Informatics J.*, vol. 25, no. 3, pp. 811-827, Sep. 2019. (<https://doi.org/10.1177/1460458217723169>)
- [6] V. Nunavath, et al., "Deep neural networks for prediction of exacerbations of patients with chronic obstructive pulmonary disease," *EANN 2018*, pp. 217-228, Bristol, UK, Sep. 2018. (https://doi.org/10.1007/978-3-319-98204-5_18)
- [7] D. Perna and A. Tagarelli, "Deep auscultation: Predicting respiratory anomalies and diseases via recurrent neural networks," in *2019 IEEE 32nd Int. Symp. CBMS*, pp. 50-55, Cordoba, Spain, Jun. 2019. (<https://doi.org/10.1109/CBMS.2019.00020>)
- [8] O. O. Abayomi-Alli, et al., "BiLSTM with data augmentation using interpolation methods to improve early detection of parkinson disease," in *Proc. FedCSIS*, pp. 371-380, Sofia, Bulgaria, Sep. 2020. (<http://dx.doi.org/10.15439/2020F188>)
- [9] F. N Fritsch and J. Butland, "A method for constructing local monotone piecewise cubic interpolants," *SIAM*, vol. 5, no. 2, pp. 300-304, 1984. (<https://doi.org/10.1137/0905021>)
- [10] H. Watz, et al., "Spirometric changes during exacerbations of COPD: A post hoc analysis of the WISDOM trial," *Respiratory Res.*, vol. 19, no. 1, pp. 251, Dec. 2018. (<https://doi.org/10.1186/s12931-018-0944-3>)
- [11] C. Fletcher and R. Peto, "The natural history of chronic airflow obstruction," *Br. Med. J.*, vol. 1, no. 6077, pp. 1645-1648, Jun. 1977. (<https://doi.org/10.1136/bmj.1.6077.1645>)
- [12] Y. S. Sim, et al., "Spirometry and bronchodilator test," *Tuberculosis and Respiratory Diseases*, vol. 80, no. 2, pp. 105-112, Apr. 2017. (<https://doi.org/10.4046/trd.2017.80.2.105>)
- [13] R. O. Crapo, A. H. Morris, and R. M. Gardner, "Reference spirometric values using techniques and equipment that meet ATS recommendations," *Am. Rev. Respiratory Disease*, vol. 123, no. 6, pp. 659-664, Jun. 1981. (<https://doi.org/10.1164/arrd.1981.123.6.659>)
- [14] R. B. Kline, *Principles and Practice of Structural Equation Modeling*, 2nd Ed., New York: Guilford, 2005.

Beomseo Choi



Mar. 2019~Current : B.S. in Computer Engineering, Daejeon University.

<Research Interests> Machine learning.

[ORCID:0009-0006-9371-2825]

Hongjun Kim



Feb. 2014 : Ph.D. School of Electrical Engineering, KAIST, Korea

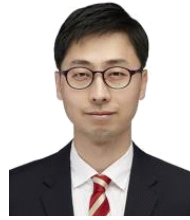
Feb. 2014~Apr. 2014 : Samsung Electronics Company Ltd., Korea

Mar. 2014~Current : Dept. Of Computer Engineering, Daejeon Univ.

<Research Interests> Optimal control, mobile robot, machine learning.

[ORCID:0000-0002-4308-342X]

Seung Hyun Jeon



Feb. 2017 : Ph.D. School of Electrical Engineering, KAIST, Korea

Jul. 2018~Mar. 2023 : KT R&D Center, KT, Korea

Mar. 2023~Current : Dept. Of Computer Engineering, Daejeon Univ.

<Research Interests> Machine learning, blockchain networks, energy consumption models for networks.

[ORCID:0000-0001-7303-4672]