RESEARCH ARTICLE

Non-Contact Respiration Rate Measurement From Thermal Images Using Multi-Resolution Window and Phase-Sensitive Processing

JIWON CHOI[1], **KYEONG-TAEK OH**[1], **OYUN KWON**[1], **JUN HWAN KWON**[1], **JEONGMIN KIM**[2], **AND SUN K. YOO**[1]

[1]Department of Medical Engineering, Yonsei University College of Medicine, Seoul 03722, Republic of Korea
[2]Department of Anesthesiology and Pain Medicine, Yonsei University College of Medicine, Seoul 03722, Republic of Korea

Corresponding authors: Sun K. Yoo (sunkyoo@yuhs.ac) and Jeongmin Kim (anesjeongmin@yuhs.ac)

ABSTRACT This paper presents a method for measuring human respiration using a thermal camera. Respiration, being a commonly monitored biomedical signal, has traditionally been measured using contact-based methods, which can cause discomfort and skin damage to patients. With the need for non-contact respiration measurement to prevent infection in the post-COVID-19 era, previous studies encountered challenges in applying non-contact methods in clinical settings due to the distortion of respiratory signals caused by high-dimensional filters. To address these limitations, this paper proposes the use of a thermal camera for non-contact data acquisition and accurate respiratory rate (RR) prediction, even in the presence of noisy respiratory signals. The proposed method leverages a multi-resolution window (MW) and phase-sensitive (PS) processing. Specifically, thermal images are captured using an infrared (IR) thermal camera, and the respiratory signal is extracted from the series of thermal images. The MW method, employing three windows of different sizes in the time domain, is then applied to the extracted respiratory signal. The resulting MW output is transformed into the frequency domain using the Fourier synchro-squeezed transform (FSST). The PS processing involves the multiplication of the converted signal, which is in polar form after a phasor operation. The processed data is utilized to train two bidirectional Long Short-Term Memory (bi-LSTM) networks, and the RR is calculated based on the trained model. To validate the proposed model, a total of 37 surgical patients were involved, with 20 patients used for training and 17 patients for testing. Six deep learning models were designed and their performances were compared. The results indicate that the proposed model outperformed the others, achieving a test classification accuracy of 98.06% and a root mean square error (RMSE) of 0.381. Despite being non-invasive and non-contact, our method demonstrates high accuracy in predicting RR, attributed to the utilization of MW and PS processing. Therefore, it holds potential for application in clinical environments, such as monitoring patients with pulmonary diseases or in intensive care units (ICUs).

INDEX TERMS Thermal imaging camera, multi-resolution window, phase-sensitive processing, feature extraction, deep learning, respiration rate.

The associate editor coordinating the review of this manuscript and approving it for publication was Cristian A. Linte.

I. INTRODUCTION

Blood pressure, heart rate, body temperature, and respiration rate (RR) are representative vital signs, which are important elements for healthcare monitoring and prediction of diseases. Among others, RR is closely related to other vital signs [1] and can predict pulmonary diseases [2]. Therefore, after the coronavirus infection (COVID-19), a lot of study on RR prediction is being conducted.

The unit of RR is breath per minute (bpm), which is calculated by counting respiratory cycles per minute. RR ranges from 12 to 20 bpm in adults and 30 to 60 bpm in neonates [3]. In hospitals, medical staff measures RR passively by counting abdominal and chest movements [4]. However, such a passive method is inefficient in a busy medical field, and since it is based on visible movements, it is difficult to accurately measure respiration. Therefore, it is necessary to develop a system capable of measuring continuous respiration with high accuracy that can be used in hospitals.

In general, RR measurement approaches are classified into contact methods and non-contact methods. Contact methods require wearing a chest belt, combining several sensors, or indirectly using a pulse oximeter to acquire data [5], [6], [7]. While contact methods exhibit high accuracy, their implementation on surgical patients or patients suffering from infectious viral diseases is difficult. In particular, in the case of surgical patients, even contact-type invasive intubation systems have poor reliability due to noise caused by movement or sensor contact failure [5]. Also, the method of chest impedance presents a limitation when detecting episodes of apnea or using electro surgical unit (ESU). Therefore, recently, to solve problems and not harm patients, studies in a non-contact method have been proposed [6], [7], [8], [9], [10], [11], [12], [13], [14], [15]. As a method of acquiring respiration data in a contactless method, there are some cases using radar sensors [6], [7], [8]. Radar sensors can obtain data even in the lack of illumination and exhibit high precision despite being a non-contact method. However, they react sensitively to the external environment. So, it can only measure data accurately when the chest is exposed without the patient's clothing. In another way, respiration can be measured in a non-contact method using the camera [9], [10], [11], [12], [13], [14], [15]. The camera method is very convenient and has the advantage of measuring respiration using a smartphone or an inexpensive camera sensor. In particular, the infrared (IR) sensors [11] or IR thermal camera [12], [13], [14], [15]-based non-contact measurement methods have many been proposed. Using the IR thermal camera, it is possible to measure even in the dark condition without any light, and the RR can be estimated from the temperature changes in the nasal region [16], [17]. It can be also classify patients with symptoms of COVID-19 by detecting fever [18]. Emotion classification is also possible [19]. Therefore, in this paper, a thermal imaging camera was used to acquire data in a non-contact manner.

This study aims to develop a system that predicts RR with high accuracy, even though it is a non-contact method, so it can be used in a virus situation and in hospitals. In addition, there is no harm to the patient and no discomfort in wearing sensors, and the medical staff does not have to count respiration. The proposed method automatically selects the region of interest (ROI) in the thermal image and then integrates the image based on the time axis to extract the respiratory signal from the pixels with the temperature change due to respiration. After, the respiration rate is calculated using signal processing and deep learning. The contributions in this work are:

- We propose multi-resolution window (MW) processing. The respiratory signal extracted from the thermal image has poor resolution features. So, the number of features for frequency at time resolution was increased by using three different windows.
- By using phase-sensitive (PS) processing, the performance of the deep learning model we developed was improved. After the time-frequency transform [20], [21], [22], [23], phasor operation was applied for the frequency band that includes respiration.
- Surgical patients' data was used to confirm the feasibility of the development system in the clinical settings.

The structure of this paper is as follows: Section II arranges related studies with the method. Section III describes the data acquisition environment and the design of the proposed deep learning model. Section IV presents the predicted results for each model using clinical data. The results of the two models proposed in this study were compared with four previously proposed deep learning models in the field of biomedical signal classification. Section V discusses the significance of the proposed methods based on the results. Finally, Section VI concludes the article.

II. RELATED WORK

As mentioned earlier, thermal imaging cameras, which measure temperature, were suitable for measuring respiration. Because the inspiratory temperature was lower than the expiratory temperature at room temperature. Therefore, in this paper, we will limit our analysis to the thermal camera system.

Abbas et al. [12] developed and analyzed a system for non-contact respiration monitoring using thermal imaging in a neonatal intensive care unit (ICU) in the hospital. Their method was tested in a clinical setting using the thermal camera, and respiration was monitored using wavelet transform of time-frequency analysis. However, it was only tested on 5-minute data of 5 people. In addition, it was only conducted on premature infants. So further studies were needed in a larger number of people and under different care setups.

Rumiński [13] was tested on a relatively large number of 16 people. However, the accuracy of RR varied significantly from person to person, and since the respiration rate was calculated based on the periodicity of respiration, the accuracy of the respiration rate was lower when there were sudden changes in the respiration pattern or drift for signal. Furthermore, since a 4th Butterworth filter was used to

remove the baseline, distortion of the respiratory signal was a concern.

Alkali et al. [14] and Shu et al. [15] followed the most common way of estimating RR using the thermal camera. First, the respiratory signal was extracted based on the sum of pixels in the nose area, and noise was removed using a Butterworth filter. RR was calculated based on the max index in the frequency-power domain by applying a fast Fourier transform (FFT) to the respiratory signal in the time domain. However, since the parameter for calculating RR from the max index was different for each person and retuning was required for new data. Therefore, recently, transformed Tompkins algorithm or deep learning method were widely used. Table 1 summarized the advantages and disadvantages of each method.

TABLE 1. Advantages and disadvantages of related work.

Method		Pros and Cons
Abbas et al.	Pros	- Tested on patients in clinical environment
	Cons	- Tested only for as few as 5 subjects - Tested only on infants in the neonatal ICU
Rumiński	Pros	- Tested with lots of data from 16 subjects
	Cons	- Fitting required per subjects - Low accuracy when there is a sudden change in the respiration pattern - Signal distortion by using high order filter
Alkali et al.	Pros	- Find nose ROI with segmentation method - Low noise in raw respiratory signal
	Cons	- Tested only for as few as 5 subjects - Calculate RR from max index (need retuning)
Shu et al.	Pros	- Find face ROI with YOLO (low signal noise)
	Cons	- Tested only 1 person - Calculate RR from max index (need retuning)

Recently, when predicting respiration in a non-contact method, two algorithms were mostly used to calculate RR from a respiration signal with a large signal-noise ratio (SNR). The first was a threshold peak detection algorithm based on the Tompkins algorithm [24], [25], [26]. If the signal does not have a rising or falling trend, the peak can be easily found using the threshold method. The second method used the deep learning model based on Long Short-Term Memory (LSTM). After increasing the features of a one-dimensional signal using Fourier transform or wavelet transform, the deep learning method based on LSTM learned feature maps. Recently, instead of the time-frequency transform, the one-dimension Convolutional Neural Network (1dCNN) was used to design the feature maps.

The first way, Yang et al. [26] used peak detection algorithm based on Pan & Tompkins, which uses multiple filters to detect features based on a few computations. It first detects feature regions in the processed signal and then searches for them in the original signal, which limits its utilization for rapid real-time classification and prediction. The second way of classifying signal features using deep learning has been used more for electrocardiogram classification than

respiration. McSharry et al. [27] used the LSTM model to learn the time-sequence characteristics of signals and classify the features. Pham [28] improved the detection accuracy of signal-based features by performing classification based on the forward time step and the backward time step of the data. Deep learning models enable real-time processing due to their fast classification speed but have relatively low accuracy. When waveform features are prominent, as in an electrocardiogram signal, they can be classified adequately using LSTM. However, in the case of respiration, the rate and volume of respiration vary from person to person and change over time for the same person. Consequently, the features are not clearly distinguished throughout the signal, resulting in low classification accuracy. Therefore, Yildirim [29] and Meng et al. [30] added a wavelet transform to their preprocessing to amplify features and improve accuracy. Further, recently, the 1dCNN-LSTM model [31], [32], [33], [34], which adds a convolutional layer to the deep learning model, has been used frequently. However, both above solutions increase the model size significantly and only improve accuracy marginally.

III. PROPOSED METHODOLIGY

In this work, our method used the bi-LSTM model to predict RR and Fourier synchro-squeezed transform (FSST) to improve prediction accuracy. Furthermore, to prevent signal distortion caused by high-order filters, we proposed MW and PS processing. All data used were acquired in the operating room. The integral image with only the temperature changes due to respiration was used to extract the respiratory signal. The flowchart of the methodology is depicted in Fig. 3, and each step is elaborated in detail below.

A. EXPERIMENTAL ENVIRONMENT

Fig. 1 shows the experimental setup. The IR thermal camera (FLIR T-420, FLIR Systems, Inc.) detects temperature with an accuracy of $\pm 2\%$ in the range of -20 to $650°C$. The image capture rate is 10Hz, and the resolution is 320 by 240 pixels, and the field of view is $25°$ by $19°$. The distance between the patient and camera was 50 to70cm. To make the nose box setup convenient in preprocessing, the face in the image was placed in the center. In addition, to compare results, the RR was measured using simplified electrical impedance
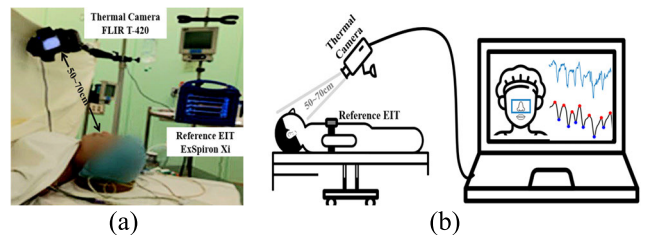

FIGURE 1. Experimental environment—Simultaneous data acquisition with IR thermal imaging camera and EIT. (a) Data acquisition environment in the operating room. (b) A schematic diagram of the experimental setup.

tomography (EIT) device (ExSpiron Xi, SENZIME, Inc.) at the same time as the thermal camera. The limits of the acquisition environment of the data set are based on operating room regulations: temperature 20 to 24°C, humidity 50 to 55%.

B. PREPROCESSING

The raw respiratory signals can be obtained from temperature changes in the nasal region in thermal images. First, the integral images were made by applying the integral operation from raw thermal images, like Fig. 2. Because the integration operation for calculating the difference between the current sample image and the previous sample image was applied, only pixel values with temperature changes due to respiration existed in the integral image. Second, the inner corners of the eyes based on pixels with high temperature were automatically found. The nose box was set to a position where temperature changes between the inner eye canthus [35], [36]. Third, for each frame, the raw respiratory signals were extracted by adding the pixels within the nose box. Last, a 2nd Butterworth filter was used to remove the artifacts of the raw respiratory signal. The designed filter is a low-pass filter, and the cutoff frequency is 1Hz.
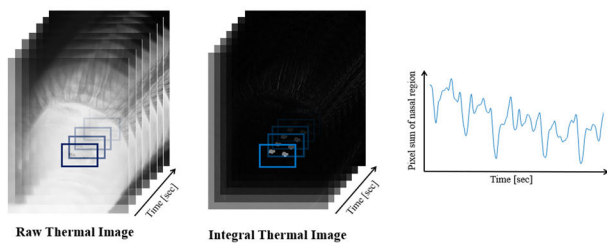

FIGURE 2. Raw respiratory signal extraction—A series of processes to make integral images from thermal images and extract respiratory signals from the nasal region.

C. MULTI-RESOLUTION WINDOW SLICING

The point where the temperature of the nose part is lowest during inhalation is called the true valley, and the point where the temperature of the nose part is highest during expiration is called the true peak. Except for real samples, which are feature points of the breathing signal, samples that look like valleys are false valleys, and samples that look like peaks are false peaks. Our method uses all samples that look like features and quickly classifies them through deep learning. All feature samples exist at positions where the first derivative of the signal is zero. Therefore, to detect true valley samples with a deep learning model, MW processing is applied to all positions where the gradient of the respiratory signal is 0.

In general, the MW is a technique for restoring the resolution by applying several kernels with different sizes in the low-resolution image [37]. However, in this study, the MW used in a 2D image was applied to a 1D signal. MW used three windows of different sizes in the time domain based on the adult's respiratory period. So, the small window (S-window) covers almost 3 seconds and includes more than

half cycle of normal respiration. And the medium window (M-window) covers 6 seconds and includes more than 1 cycle of normal respiration. Last, the large window (L-window) covers 12 seconds and includes more than 2 cycles of normal respiration. As shown in Figure 3, since the MW was set so that the feature sample with zero gradients was in the center, the window between samples overlapped if there were many feature samples in a short time. After MW processing, Min-Max normalization was performed for each window. The normalization range was set from −1 to 1.

D. PHASE-SENSITIVE DATA PROCESSING

After the MW processing and normalization, the data were transformed from time to frequency domain using a Fourier synchro-squeezed transform (FSST) [20], [21], [22], [23]. The analysis of the FSST was performed using the Kaiser window, and the window design was performed using the MATLAB R2021b FSST function tool. After FSST, the frequency band of interest was a band where respiration exists, and a low frequency of 1 Hz excluding DC was selected.

PS processing was used to increase the number of time and frequency information in the signal. When using time-frequency transforms such as Fourier transforms or wavelet transforms, the data was transformed into a complex form with amplitude and phase spectrum. In the past, only amplitude information was used in many cases, but recently, both amplitude and phase information in image data have been used to improve accuracy [38]. After FSST used in the proposed method was also in the form of a complex number $Z = a + bi$—real part a, imaginary part b. Therefore, to increase the information amount of the respiratory signal, which has less frequency information compared to other bio-signals, phasor operation was applied to the frequency band of interest. In other words, the data to which equation (1) was applied was used as the input data of the deep learning model.

$$\sqrt{a^2 + b^2} \times \angle tan^{-1}\left(\frac{b}{a}\right) \qquad (1)$$

E. CLASSIFICATION DEEP LEARNING MODEL

37 surgical patients' data were used for training and test the deep learning model. All data were obtained with informed written consent from the participants. The data from 37 subjects were randomly divided (Table 3). 20 patients for training were 18 males, 2 females. 17 patients for the test were 16 males, 1 female.

To verify the superiority of the proposed techniques, six deep learning models were designed by applying different preprocessing methods, and training was conducted using data from 20 subjects. Six models are Convolutional and recurrent networks (1dCNN+LSTM) [31], Single window LSTM (SWL) [27], Single window bi-LSTM (SWbL) [28], Single window FSST bi-LSTM (SWbL+ FSST) [20], Multi window FSST bi-LSTM (MWbL+FSST), Multi window phase sensitive bi-LSTM (MWbL+PS).
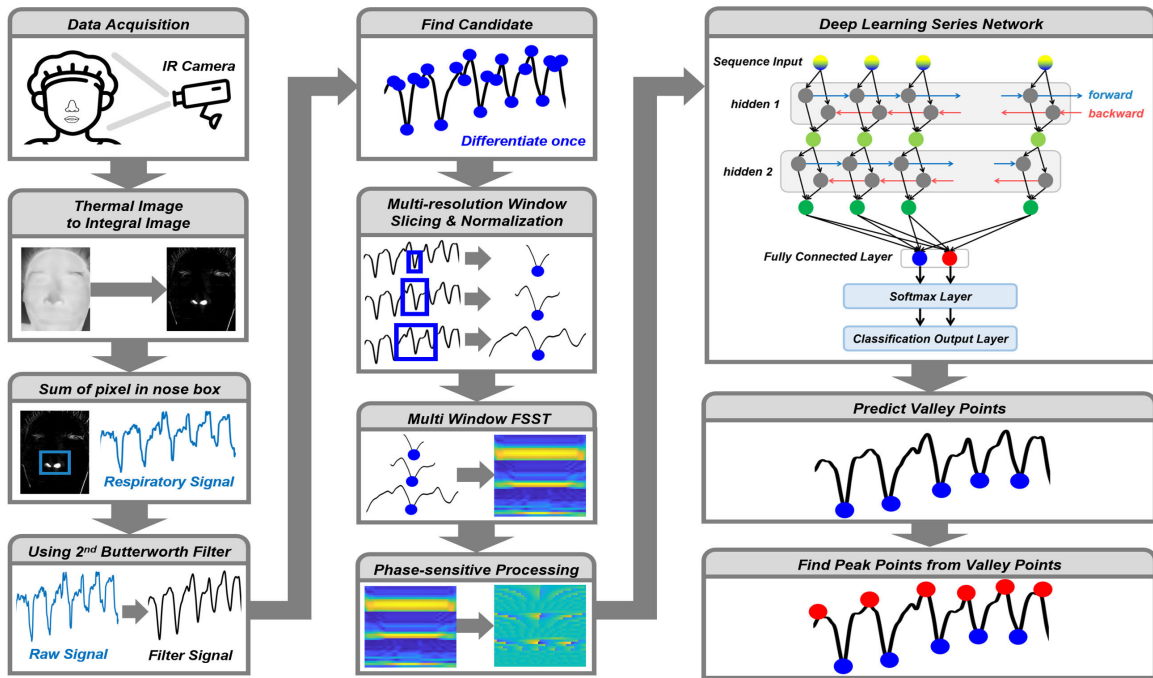
FIGURE 3. Data flow of proposed model.

Fig. 3 shows the flow of the proposed model, MWbL+PS. MWbL+PS consists of two bi-LSTM layers. The bi-LSTM layers consisted of 128 and 93 hidden units, and the mini-batch size was set to 128. Based on experience, the number of train epochs was set to 10 to avoid overfitting. The output of the final LSTM network was transmitted to a fully connected layer with two nodes—one representing a true-valley and the other representing a false-valley. The cost function of the softmax layer used cross entropy. The classification weights of the output layer were set to inversely proportional to the number of each class.

The other five models were the same process as MWbL+PS until the preprocessing of selecting feature samples. MWbL+FSST was almost the same process as MWbL+PS, and the difference was that PS processing was not used for MWbL+FSST. MWbL+FSST was learned by creating feature maps using only the amplitude values of data composed of complex numbers after FSST. SWbL+FSST was a similar process to MWbL+FSST but applied only S-window in the window slicing step. SWbL was not using FSST. After preprocessing, one S-window was applied, and the time domain data were used as input to the deep learning model. SWL used LSTMs instead of bi-LSTMs in the process for SWbL. Last, 1dCNN+LSTM used an M-window that includes 1 cycle of normal respiration for adults after preprocessing. 1dCNN+LSTM used multiple CNNs instead of FSST to extract features and was trained with 1 LSTM.

F. ANALYSIS METHODS

17 patients' data were used for the test. To confirm the performance of the deep learning model, a confusion matrix was constructed, and the real label and the predicted label were compared. The EIT data measured with the correct answer is not the waveform of the signal but consists of the RR calculated inside the gold standard device. Therefore, the answer label was assigned to the extracted waveform based on the EIT's RR. The true positives (TP) were true-valleys for both answer labels and prediction labels, and true negatives (TN) were all false-valleys (no-valleys). The performance of the model was verified respiration-by-respiration using accuracy, precision, recall, and F1-score with the following equations (2), (3), (4), (5). In addition, the correlation between the true-RR measured by EIT and the RR predicted by the proposed models was confirmed, and a Bland-Altman analysis was performed.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F1score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

IV. RESULTS

A. DEMOGRAPHIC INFORMATION

37 surgical Table 2 shows the demographics of a dataset in which 37 people were randomly divided into 20 and 17 people. A t-test was performed for age, height, weight, BMI, acquisition time, and gender to confirm that there was no statistical difference between the training and test data ($p > 0.05$).

TABLE 2. Patient demographics.

	Train		Test		p-value
	Avg.	Std.	Avg.	Std.	
Age (years)	70.71	9.22	67.27	9.17	0.382
Height (cm)	167.00	8.26	168.00	7.32	0.712
Weight (kg)	70.79	13.74	68.06	9.01	0.503
BMI (kg/m^2)	25.25	8.67	24.06	2.31	0.260
Acquisition Time (min)	9.52	1.25	10.91	3.91	0.152
Gender (M/F)	18 / 2	-	16 / 1	-	0.658

TABLE 3. Model evaluation metrics.

	Train Data	Test Data			
	Accuracy	Accuracy	Precision	Recall	F1-score
1dCNN+LSTM	96.09	95.44	96.97	98.53	97.74
SWL	93.80	93.50	97.00	96.48	96.74
SWbL	93.78	93.67	97.01	96.65	96.83
SWbL+FSST	94.86	94.04	96.43	97.65	97.04
MWbL+FSST	99.13	97.99	98.54	99.46	99.00
MWbL+PS	99.53	98.06	98.48	99.60	99.03

TABLE 4. Correlation coefficient and mean-difference for each model.

	1dCNN +LSTM	SWL	SWbL	SWbL +FSST	MWbL +FSST	MWbL +PS
Pearsons's r	0.963	0.901	0.913	0.941	0.982	0.992
R-square	0.928	0.812	0.834	0.885	0.964	0.983
Residual sum	7.656	20.896	16.925	11.454	4.201	2.176
RMSE	0.714	1.180	1.062	0.874	0.529	0.381
Mean	0.582	1.171	1.029	0.812	0.646	0.262
+1.96 SD	2.196	3.659	3.380	2.819	1.773	1.015
-1.96 SD	-1.033	-1.316	-1.323	-1.196	-0.481	-0.490

B. PERFORMANCE OF VALLEY DETECTION

Table 3 shows the performance of six deep learning models designed for comparison and verification. For training data, MWbL+PS had the best accuracy, followed by MWbL+FSST, 1dCNN+LSTM, SWbL+FSST, SWbL, and SWL. Similarly, for the test data, the proposed MWbL+PS model has the best evaluation values of accuracy, precision, recall, and F1-score, followed by MWbL+FSST and 1dCNN+LSTM.

Fig. 4 shows the selected feature samples, the results predicted by the six models, and the ground truth for 1 minute of test data. There were many points where the gradient was zero in the denoised signal, as shown in (c). Using the feature samples found in (c), the deep learning models predicted classes for true valleys or false valleys, and the peaks were set to positions where the magnitude of the signal was max between two valleys based on the predicted valley points. (d) using 1dCNN + LSTM incorrectly predicted noise as a true valley in the temperature rise section due to exhalation. (e), (f), (g), and (h) incorrectly predicted the noise during exhalation as in (c), and the noise during apnea before transitioning from inhalation to exhalation was incorrectly predicted as the true valley. Therefore, only the result of MWbL+PS was equivalent to the ground truth.

C. CORRELATION FOR RR

Fig. 5 is a correlation plot and Bland-Altman plot comparing the RR measured by EIT and the RR predicted by each deep learning model. Table 4 summarizes the results of Fig. 5. To compare RR, we used 2 minutes of data among the 10 minutes of each patient's data in the test and the average RR obtained from EIT. The results indicate a strong correlation between the proposed contactless approach and the contact approach. For MWbL+PS, the Pearson correlation coefficient r was 0.992, the R-square was 0.983, close to 1, and the root mean square error (RMSE) was 0.381. The correlation of
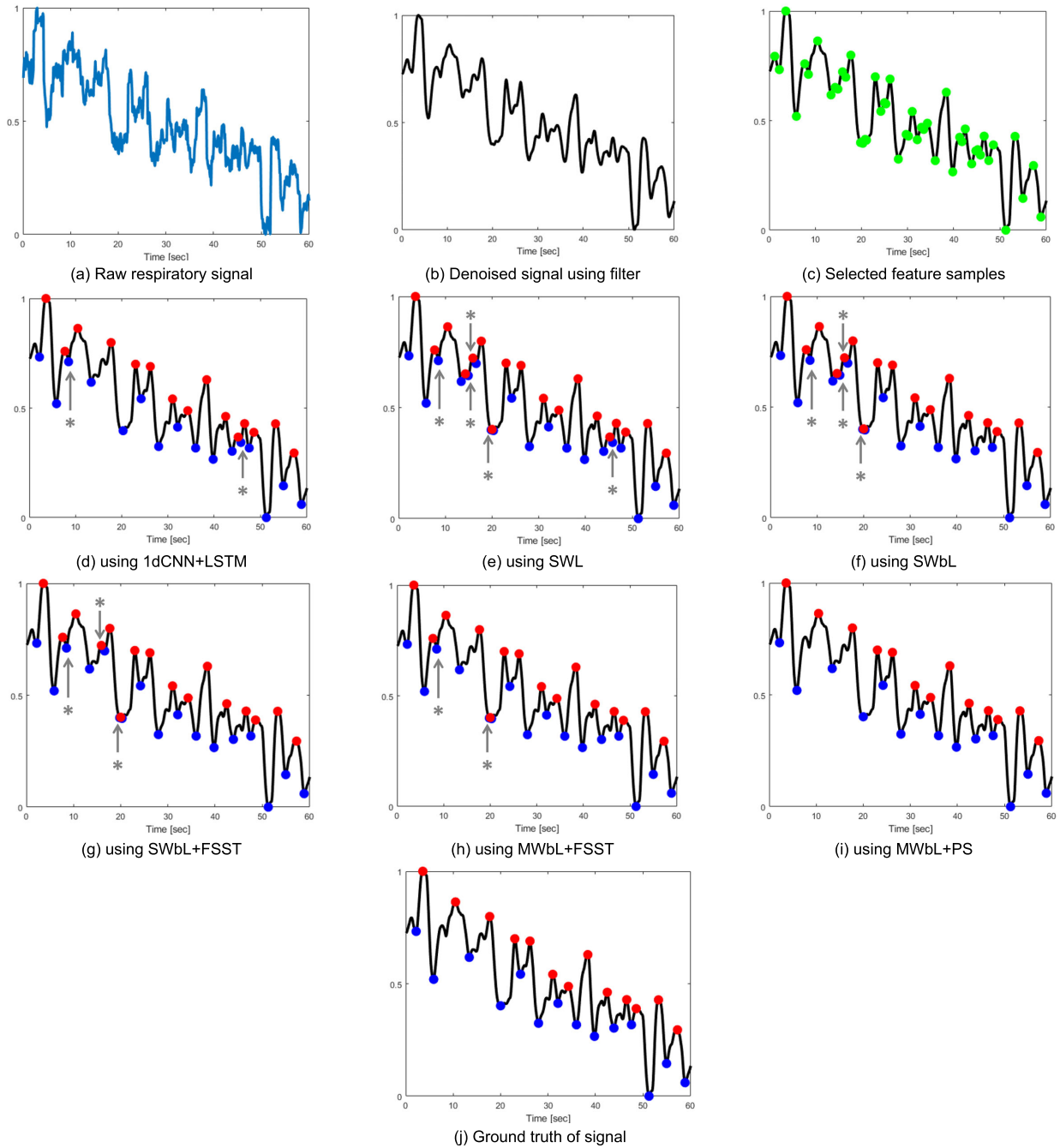
FIGURE 4. Feature detection results of each model—(a) is the raw signal extracted from the region of interest (ROI) around the nose. (b) is a signal from which noise is removed by applying a 2nd Butterworth filter to the signal depicted in (a). (c) shows all points (green points) where the gradient is 0 from (b). Window processing is applied to green points. (d) shows the test results of the 1dCNN+LSTM model. (e), (f), (g), (h), and (i) show the prediction results of SWL, SWbL, SWbL+FSST, MWbL+FSST, and MWbL+PS. (j) is the ground truth of the signal guessed from the RR measured by EIT. In (d) to (j), the blue points are the valley region of the respiratory signal, and the red points are the peak region of the respiratory signal, and the peaks or valleys with gray arrows are incorrect prediction errors.

MWbL + PS and EIT was the highest, followed by MWbL + FSST, 1dCNN + LSTM, SWbL + FSST, SWbL, and SWL.

V. DISCUSSIONS
A. SURGICAL PATIENTS' DATA
The data used for training and test are the thermal face images of 37 people acquired in the environment of Fig. 1. The experimental environment was designed to reflect the real clinical setting where spontaneous respiration was maintained but gradually impaired. A total of 37 patients, including 6 transurethral resections of the prostate (TURP) and 31 transurethral resections of the bladder (TURB), breathed spontaneously under spinal anesthesia. All data consisted of patients aged 46 to 77 years,
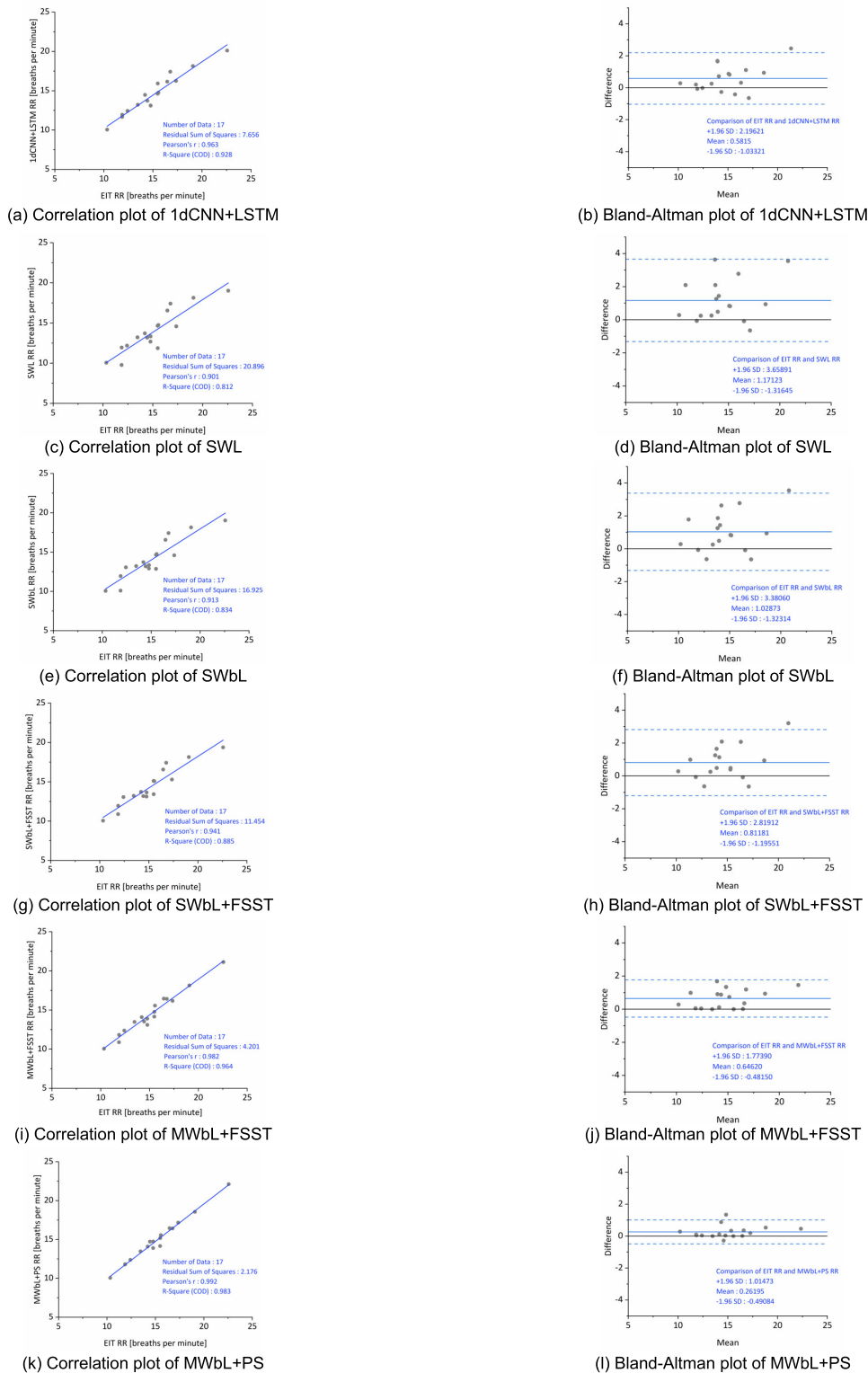
FIGURE 5. Correlation plot and Bland-Altman plot of each model—(a) and (b) present a correlation plot and Bland-Altman plot for 1dCNN+LSTM. (c) and (d) present plots for SWL. (e) and (f) present plots for SWbL. (g) and (h) present plots for SWbL+FSST. (i) and (j) present plots for MWbL+FSST. Finally, (k) and (l) present plots for MWbL+PS, which uses MW and PS processing.

with a mean age of 60 years (Table 2), and patients who underwent spinal anesthesia were recruited without age restrictions. However, TURB and TURP were commonly performed on patients with benign prostatic hyperplasia or bladder tumors, so there are many patients over the age of 50.

In general, in the case of surgical patients, RR is acquired through a patient monitor or a sensor in a ventilator. But the accuracy is relatively low due to problems, such as the movement of the transducer and poor contact with the sensor. Therefore, non-contact respiration measurement based on thermal imaging camera and deep learning can lead to high accuracy. Since our method is a non-contact method, there is no patient discomfort due to sensor attachment, and infection can be prevented. In addition, since respiration can be easily measured under the condition of Fig. 1, there is no need to directly count the patient's respiration, which is advantageous in terms of convenience for medical staff.

B. PREPROCESSING

In a resting state, the respiratory period in adults is 12 to 20 breaths per minute, and abnormal respiration is defined as less than 12 or more than 25 breaths per minute [3]. Therefore, the cut-off frequency of the 2nd Butterworth filter was set to 1Hz, assuming the situation of 60 breaths per minute to remove only noise without removing respiratory features. In particular, in prior research, there was a problem of signal distortion by using a high-order Butterworth filter or multiple filters such as Chebyshev, Bessel, etc [14], [15]. However, since the proposed method applied MW and PS processing, only little amount of high-frequency noise was removed using a low-order Butterworth filter without using multiple filters.

After removing the noise components, the positions where the 1st derivative value was zero were selected. In the selected feature samples, there were a small number of true valley samples that were respiratory features and many false valley samples that were not features. So, since the data was imbalanced, the performance was evaluated based on the F1-score, a commonly used metric for evaluating imbalanced classes (Table 3).

C. MULTI-RESOLUTION WINDOW SLICING

In this study, the MW was used to classify the feature regions of the respiratory signals effectively from the perspective of the model. Based on the adult's respiratory period, the three windows were designed to cover at least a half cycle, one cycle, and two cycles of steady-state respiration. The S-window clearly showed the features of true valleys. M-window was added to distinguish between false and true valleys' features. And L-Window was used to capture the periodic features of true-valleys. After MW processing, the Min-Max normalize was performed, for fast convergence of the deep learning model.

D. PHASE-SENSITIVE DATA PROCESSING

After applying MW, the data was converted to the time-frequency domain using FSST to increase the number of frequency information. According to a paper comparing frequency conversion methods [22], FSST uses an additional argument $e^{j2\pi \eta t}$ to enable transformation without increasing computational complexity. Further, since the phase $\varphi_k (t)$ of

FSST is the form of a time derivative $(d\varphi\_k(t))/dt$, corresponding to an instantaneous frequency, the transformed data involves more phase features than other frequency transformations. After transformation, the frequency band for feature extraction was set equal to the Butterworth filter's cut-off frequency and used after removing the DC component from bands from 0 to 1Hz.

In general, only amplitude components are used, or amplitude and phase components are used in different feature lines to improve accuracy—the top lines of the feature map were composed of amplitude, and the bottom lines were composed of phase. However, when using only the amplitude components, there was a concern that underfitting may be caused for respiratory signals with few features. In another way, when creating feature maps using different feature lines, an overfitting problem for phase components may occur in which new data cannot be properly predicted, because too many features were extracted for one sample. Therefore, we solved the problem with PS processing that performed phasor operation.

E. CLASSIFICATION DEEP LEARNING MODEL

The primary contribution of this study is the classification of respiratory feature regions with high accuracy using simple deep learning model architecture. The proposed method outperformed the recently utilized signal classification method, 1dCNN+LSTM. 1dCNN+LSTM consists of 4 CNN blocks and 1 LSTM block. In the CNN architecture, convolution, normalization, and dropout are repeated twice, and convolutional residual blocks are connected to each CNN block. In contrast, the proposed deep learning models have simple architecture, comprising two bi-LSTM layers, a fully connected layer, softmax, and classification, as depicted in the deep learning series network in Fig. 3. The number of hidden units is a key factor in LSTM. Generally, the number of hidden units was set to 100 to 300 based on the features of the given data [39], [40], [41]. Therefore, the first hidden layer was set to 128, which was similar to the size of the L-window. And the second hidden layer was set to 93, which was a multiple of the size of the S-window and close to 100. The output layer of the final model is the classification output layer. To prevent overfitting and underfitting of imbalanced data, the weight values that are inversely proportional to the number of samples in each class were set.

The experimental results showed that the MWbL+PS model performs the best, with an accuracy of 98.06 % (Table 3). The error of 1.94% was when the nasal region was not visible within the image frame. For example, RR cannot be predicted in a dynamic situation in which the subject's head was completely turned left/right or when the medical staff's hand covered the nose.

The performance improvement of the proposed MW processing was confirmed by comparing SWbL+FSST and MWbL+FSST. Table 3 showed that the accuracy and F1-score of MWbL+FSST were higher than SWbL+FSST. Next, the performance improvement of the PS processing was

confirmed by comparing MWbL+FSST and MWbL+PS. Table 3 showed that the accuracy and F1-score of MWbL+PS were higher than MWbL+FSST. However, in signal processing, phasor operations are known to enhance classification performance [42], [43]. Also, the MWbL+FSST could not predict all true-valleys, but the MWbL+PS correctly predicted all true-valleys (Fig. 4). Furthermore, in the case of surgical patients, it was dangerous if respiration was not accurately measured, so there was a difference in the performance from a clinical aspect. Therefore, the results indicated PS processing improved classification performance. In addition, via the comparison of the models' predicted RR and the ground-truth RR, MWbL+PS showed the highest Pearson's r value at 0.992, and the RMSE and residual sum value of MWbL+PS were the smallest among the six models (Fig. 5, Table 4). Therefore, it is proved that the performance of MWbL+PS was the best.

In related work, the most extensively used model on this topic is the 1dCNN+LSTM. Instead of complex processing, 1dCNN+LSTM adds a 1d-convolutional layer before the LSTM neural network to extract features, resulting in higher classification accuracy than SWbL+FSST (Table 3). However, the performance of 1dCNN+LSTM is lower than models with MW processing. Both the accuracy and F1-score of MWbL+FSST or MWbL+PS are higher than 1dCNN+LSTM. Therefore, the proposed model detects respiration features better than 1dCNN+LSTM.

F. LIMITATIONS AND FUTURE WORKS
There are three limitations in this study. First, we only consider integral values obtained from the nasal region to use pixels with temperature changes due to respiration. We excluded the mouth region because only 17% of the world's population breathes orally [44] and anesthetized patients move their mouths a lot. So, in future work, RR can be more accurately estimated by extracting raw signals, not only from the nose but also from the mouth. However, the proposed algorithm cannot measure RR when the respiration region is covered by an object.

Second, the proposed method defined the nasal region by finding the pixel with the low temperature between two pixels with the highest temperatures. In other words, our method used the ROI box by the manual. Therefore, noise caused by movement is included in the signal. In future work, the performance can be improved by adding an object detection and tracking algorithm. However, since the proposed algorithm is a non-contact method, measurement is possible only in static situations, including slight motion. Measurements are not possible in dynamic situations with large patient movements. In addition, if the distance between the subject and the camera was too far, the number of pixels occupied by the subject in the image was small and the resolution was very low. Therefore, the setting in Fig.1 was recommended for the measurement environment.

Third, environmental limitations of the data exist. Only the respiration of patients aged 40 to 70s under spinal anesthesia

was verified. Also, all data were acquired under the operating room regulations, the temperature of 20 to 24° and humidity of 50 to 55%. After, it will be possible to use it in daily life if it is additionally confirmed whether it operates outside of a limited environment such as an operating room or an ICU and whether it operates in the respiration of various ages.

However, despite these three limitations, the proposed method was still highly practical. First, the proposed method can be applied in virus situations as it measured RR using a non-contact manner. Especially, it can be used to measure the respiration of patients within isolation wards, making it useful for infection control. Second, it was more beneficial than contact-based systems in busy clinical environments. If the device of the proposed method is improved to display the respiratory waveform and RR and to add the alarm for abnormal respiration, medical staffs in busy clinical environments can conveniently record and monitor patients' continuous respiration. Therefore, the proposed method is expected to be utilized in virus situations, monitoring equipment in ICU with frequent cases of apnea, gastrointestinal endoscopy, and children undergoing MRI or CT [12], [45].

TABLE 5. Common issues in related work and the performance advantages of the proposed method.

| Common issues in related work	- Only validated for a small number of subjects and a short duration of time - Signal distortion issues - Not possible in cases of signal issues (ex. signal drift)
Advantages of proposed way	- Validated over the long period of time for many subjects - Not used multi-filters, No signal distortion (advantages of using MW, PS processing) - measured possible in cases of signal issues
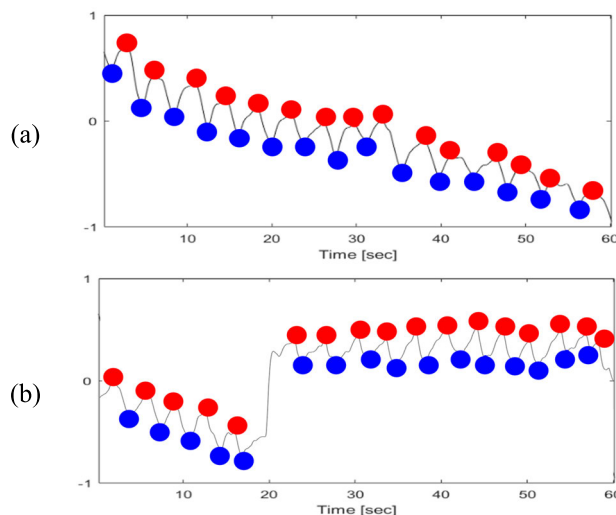
FIGURE 6. Predicted results of MWbL+PS in cases with signal issues—in cases of (a) signal with trend, (b) signal with drift.

G. PERFORMANCE IMPROVEMENTS

Related work using IR sensors or thermal camera had been proven only for short-time data and only monitoring was possible (Table 1, Table 5). However, since this study used 10 minutes of respiration data of patients during surgery, the monitoring period was relatively long. Nevertheless, the model predicted RR with very high accuracy even in situations where there was drift in the respiratory signals (Fig. 6). We achieved excellent performance by developing MW processing, which had not been used in signal classification and prediction. In addition, the performance was improved by increasing the number of frequency information of data using FSST and adding PS processing.

VI. CONCLUSION

In this paper, we measure continuous respiration using thermal facial images in a non-contact manner. The nasal region is defined based on the inner eye canthus, which is the area with the highest temperature in the thermal facial image. Only the pixels with temperature changes due to respiration, identified using integral images, are utilized. Previous methods suffered from distorted respiratory signals and the inability to accurately measure respiration due to the use of high-dimensional filters to remove noise from the signal. However, our method analyzes the signal in the time domain with various resolutions using MW and increases the frequency information of the signal using PS, thereby avoiding signal distortion. Additionally, we addressed the issue of data imbalance by adjusting the class weights in the classification layer of the bi-LSTM-based deep learning model. So far, there has been no case of testing long-time clinical data using deep learning in related work. The proposed method is suitable for verifying clinical feasibility because there has been no case of testing long-term clinical data using deep learning in related work. As a result, we achieved a performance improvement of 4.39% on test data. It is also 2.62% better than CNN and LSTM models combined. Therefore, the results were promising, and the error was less than 2%.

The versatility of proposed method extends beyond respiratory analysis. It can function as a key component in various non-contact biosignal measurement and analysis systems. Potential applications include but are not limited to healthcare system. The method outlined in this paper offers a flexible and innovative approach to emotion analysis, leveraging both respiratory signal and thermal imaging technology. Its potential for integration into comprehensive emotion analysis systems holds promise for advancing our understanding of human emotions and enhancing various applications across multiple domains.

A. AUTHOR CONTRIBUTIONS

All authors read and approved the final manuscript. Jiwon-Choi, the first author of this paper, played a role throughout the research and paper and was responsible for designing the algorithm, analyzing the results, and drafting the manuscript.

Kyeong Taek Oh, as the co-first author of this paper, designed the research concept, materialized the multi-window concept which is the key to the paper, and performed a critical revision of the manuscript. OyunKwon, as the third author of this paper, conducted data collection and preprocessing and designed a sample extraction method that is the input of the deep learning model. JunhwanKwon, as the fourth author of this paper, was responsible for acquiring patient data in the operating room and generated synchronized thermal images and EIT data in mat files along with patient information for ease of data utilization. JeongminKim, as the corresponding author of this paper, made significant contributions to data acquisition, analysis, and interpretation by establishing a clinical design and data collection environment and presenting clinical opinions on the study. Sun K. Yoo, as the corresponding author of this paper, provided overall guidance for the research and paper and approved the final version of the manuscript.

B. AVAILABILITY OF MATERIALS AND DATA

Informed consent was obtained from all individual participants included in the study. However, since the data are thermal face images, the datasets generated and/or analyzed during the current study are not publicly available due to privacy reasons—Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, and Korea Personal Information Protection Act (K-PIPA). But it is available from the corresponding author upon reasonable request.

C. DECLARATION OF ETHICS APPROVAL

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The data are acquired following the Good Clinical Practice principles, and approval was obtained from the Institutional Review Board of Severance Hospital (Ref. 1-2016-0008). All data used are recorded on ClinicalTrails.Gov under No. NCT02993497 and written informed consent is obtained from all patients.

D. DECLARATION OF COMPETING INTEREST
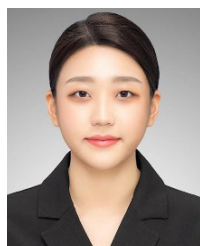
No conflict of interest.

REFERENCES

[1] J. H. Ticona, V. M. Zaccone, and I. M. McFarlane, "Community-acquired pneumonia: A focused review," *Amer. J. Med. Case Rep.*, vol. 9, no. 1, pp. 45–52, Nov. 2020, doi: 10.12691/ajmcr-9-1-12.

[2] T. A. McCartan, A. P. Worrall, R. Ó. Conluain, F. Alaya, C. Mulvey, E. MacHale, V. Brennan, L. Lombard, J. Walsh, M. Murray, R. W. Costello, and G. Greene, "The effectiveness of continuous respiratory rate monitoring in predicting hypoxic and pyrexic events: A retrospective cohort study," *Physiol. Meas.*, vol. 42, no. 6, Jun. 2021, Art. no. 065005, doi: 10.1088/1361-6579/ac05d5.

[3] A. Rowden. *What is a Normal Respiratory Rate Based on Your Age?*. Accessed: Dec. 13, 2022. [Online]. Available: https://www.medicalnewstoday.com/articles/324409

[4] S. Rolfe, "The importance of respiratory rate monitoring," *Brit. J. Nursing*, vol. 28, no. 8, pp. 504–508, Apr. 2019, doi: 10.12968/bjon.2019.28.8.504.

[5] S. Madsen, J. Baczuk, K. Thorup, R. Barton, N. Patwari, and J. T. Langell, "A noncontact RF-based respiratory sensor: Results of a clinical trial," *J. Surgical Res.*, vol. 203, no. 1, pp. 1–5, Jun. 2016, doi: 10.1016/j.jss.2016.03.018.

[6] A. Pentari, D. Manousos, T. Kassiotis, G. Rigas, and M. Tsiknakis, "Respiration and heartbeat rates estimation using IR-UWB non-contact radar sensor recordings: A pre-clinical study," in *Proc. EDBT/ICDT Joint Conf.*, 2023, pp. 1–5. [Online]. Available: https://ceur-ws.org/Vol-3379/HeDAI_2023_paper404.pdf

[7] D. Lee, "Cross correlation based signal classification for monitoring system of abnormal respiratory status," *J. Korea Academia-Ind. Cooperation Soc.*, vol. 21, no. 5, pp. 7–13, 2020, doi: 10.5762/KAIS.2020.21.5.7.

[8] J. Lee and S. K. Yoo, "Radar-based detection of respiration rate with adaptive harmonic quefrency selection," *Sensors*, vol. 20, no. 6, p. 1607, Mar. 2020, doi: 10.3390/s20061607.

[9] H. Wang and S. Zhang, "Non-contact human respiratory rate measurement under dark environments by low-light video enhancement," *Biomed. Signal Process. Control*, vol. 85, Aug. 2023, Art. no. 104874, doi: 10.1016/j.bspc.2023.104874.

[10] G. Boccignone, A. D'Amelio, O. Ghezzi, G. Grossi, and R. Lanzarotti, "An evaluation of non-contact photoplethysmography-based methods for remote respiratory rate estimation," *Sensors*, vol. 23, no. 7, p. 3387, Mar. 2023, doi: 10.3390/s23073387.

[11] X. He, R. Goubran, and F. Knoefel, "IR night vision video-based estimation of heart and respiration rates," in *Proc. IEEE Sensors Appl. Symp. (SAS)*, Mar. 2017, pp. 1–5.

[12] A. K. Abbas, K. Heimann, K. Jergus, T. Orlikowsky, and S. Leonhardt, "Neonatal non-contact respiratory monitoring based on real-time infrared thermography," *Biomed. Eng. OnLine*, vol. 10, no. 1, p. 93, 2011, doi: 10.1186/1475-925X-10-93.

[13] J. Rumiński, "Analysis of the parameters of respiration patterns extracted from thermal image sequences," *Biocybernetics Biomed. Eng.*, vol. 36, no. 4, pp. 731–741, 2016, doi: 10.1016/j.bbe.2016.07.006.

[14] A. H. Alkali, R. Saatchi, H. Elphick, and D. Burke, "Facial tracking in thermal images for real-time noncontact respiration rate monitoring," in *Proc. Eur. Model. Symp.*, Nov. 2013, pp. 265–270.

[15] S. Shu, H. Liang, Y. Zhang, Y. Zhang, and Z. Yang, "Non-contact measurement of human respiration using an infrared thermal camera and the deep learning method," *Meas. Sci. Technol.*, vol. 33, no. 7, Jul. 2022, Art. no. 075202, doi: 10.1088/1361-6501/ac5ed9.

[16] P. Jagadev and L. I. Giri, "Human respiration monitoring using infrared thermography and artificial intelligence," *Biomed. Phys. Eng. Exp.*, vol. 6, no. 3, Mar. 2020, Art. no. 035007, doi: 10.1088/2057-1976/ab7a54.

[17] M. Fu, W. Weng, W. Chen, and N. Luo, "Review on modeling heat transfer and thermoregulatory responses in human body," *J. Thermal Biol.*, vol. 62, pp. 189–200, Dec. 2016, doi: 10.1016/j.jtherbio.2016.06.018.

[18] Z. Zeng, G. Mei, T. Liao, and Y. Huang, "The temperature difference method for screening patients with COVID-19 fever symptoms," *J. Phys., Conf. Ser.*, vol. 2226, no. 1, Mar. 2022, Art. no. 012010, doi: 10.1088/1742-6596/2226/1/012010.

[19] B.-L. Jian, C.-L. Chen, M.-W. Huang, and H.-T. Yau, "Emotion-specific facial activation maps based on infrared thermal image sequences," *IEEE Access*, vol. 7, pp. 48046–48052, 2019, doi: 10.1109/ACCESS.2019.2908819.

[20] J. Choi, O. Kwon, J. Kwon, K.-T. Oh, and S. K. Yoo, "Development of signal feature extraction system for ECG-based heart disease classification," *J. Korea Multimedia Soc.*, vol. 26, no. 1, pp. 75–83, Jan. 2023, doi: 10.9717/kmms.2023.26.1.075.

[21] B. Rajeshwari, A. Sinha, A. Sengupta, M. Patra, K. P. Sahoo, and N. Ghosh, "Synchrosqueezed transform based click event segmentation in phonocardiogram of mitral valve prolapse," in *Proc. 44th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2022, pp. 1997–2000.

[22] F. Auger, P. Flandrin, Y.-T. Lin, S. Mclaughlin, S. Meignen, T. Oberlin, and H.-T. Wu, "Time-frequency reassignment and synchrosqueezing: An overview," *IEEE Signal Process. Mag.*, vol. 30, no. 6, pp. 32–41, Nov. 2013.

[23] J. Yan, S. Laflamme, P. Singh, A. Sadhu, and J. Dodson, "A comparison of time-frequency methods for real-time application to high-rate dynamic systems," *Vibration*, vol. 3, no. 3, pp. 204–216, Aug. 2020, doi: 10.3390/vibration3030016.

[24] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 3, pp. 230–236, Mar. 1985, doi: 10.1109/TBME.1985.325532.

[25] H. Sedghamiz. (2014). *MATLAB Implementation of Pan Tompkins ECG QRS Detector*. [Online]. Available: https://www.researchgate.net/publication/313673153_Matlab_Implementation_of_Pan_Tompkins_ECG_QRS_detect

[26] F. Yang, S. He, S. Sadanand, A. Yusuf, and M. Bolic, "Contactless measurement of vital signs using thermal and RGB cameras: A study of COVID 19-related health monitoring," (in English), *Sensors*, vol. 22, no. 2, p. 627, Jan. 2022, doi: 10.3390/s22020627.

[27] P. E. McSharry, G. D. Clifford, L. Tarassenko, and L. A. Smith, "A dynamical model for generating synthetic electrocardiogram signals," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 3, pp. 289–294, Mar. 2003, doi: 10.1109/TBME.2003.808805.

[28] T. D. Pham, "Time–frequency time–space LSTM for robust classification of physiological signals," *Sci. Rep.*, vol. 11, no. 1, pp. 1–11, Mar. 2021, doi: 10.1038/s41598-021-86432-7.

[29] Ö. Yildirim, "A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification," *Comput. Biol. Med.*, vol. 96, pp. 189–202, May 2018, doi: https://doi.org/10.1016/j.compbiomed.2018.03.016.

[30] Y. Meng, L. Lin, Z. Qin, Y. Qu, Y. Qin, and Y. Li, "Biosignal classification based on multi-feature multi-dimensional WaveNet-LSTM models," *J. Commun.*, vol. 17, no. 5, pp. 399–404, 2022, doi: 10.12720/jcm.17.5.399-404.

[31] J. Zhao, X. Mao, and L. Chen, "Speech emotion recognition using deep 1D & 2D CNN LSTM networks," *Biomed. Signal Process. Control*, vol. 47, pp. 312–323, Jan. 2019, doi: 10.1016/j.bspc.2018.08.035.

[32] M. N. Dar, M. U. Akram, S. G. Khawaja, and A. N. Pujari, "CNN and LSTM-based emotion charting using physiological signals," *Sensors*, vol. 20, no. 16, p. 4551, Aug. 2020, doi: 10.3390/s20164551.

[33] L. A. Abdullah and M. S. Al-Ani, "CNN-LSTM based model for ECG arrhythmias and myocardial infarction classification," *Adv. Sci., Technol. Eng. Syst. J.*, vol. 5, no. 5, pp. 601–606, 2020, doi: 10.25046/aj050573.

[34] A. Peimankar and S. Puthusserypady, "DENS-ECG: A deep learning approach for ECG signal delineation," *Expert Syst. Appl.*, vol. 165, Mar. 2021, Art. no. 113911, doi: 10.1016/j.eswa.2020.113911.

[35] A. Basu, A. Routray, R. Mukherjee, and S. Shit, "Infrared imaging based hyperventilation monitoring through respiration rate estimation," *Infr. Phys. Technol.*, vol. 77, pp. 382–390, Jul. 2016, doi: 10.1016/j.infrared.2016.06.014.

[36] A. H. Alkali, R. Saatchi, H. Elphick, and D. Burke, "Thermal image processing for real-time non-contact respiration rate monitoring," *IET Circuits, Devices Syst.*, vol. 11, no. 2, pp. 142–148, Mar. 2017, doi: 10.1049/iet-cds.2016.0143.

[37] D. Qiu, Y. Cheng, X. Wang, and X. Zhang, "Multi-window back-projection residual networks for reconstructing COVID-19 CT super-resolution images," *Comput. Methods Programs Biomed.*, vol. 200, Mar. 2021, Art. no. 105934, doi: 10.1016/j.cmpb.2021.105934.

[38] A. Anaya-Isaza and M. Zequera-Diaz, "Fourier transform-based data augmentation in deep learning for diabetic foot thermograph classification," *Biocybernetics Biomed. Eng.*, vol. 42, no. 2, pp. 437–452, Apr. 2022, doi: 10.1016/j.bbe.2022.03.001.

[39] H. Kang, S. Yang, J. Huang, and J. Oh, "Time series prediction of wastewater flow rate by bidirectional LSTM deep learning," (in English), *Int. J. Control, Autom. Syst.*, vol. 18, no. 12, pp. 3023–3030, Dec. 2020, doi: 10.1007/s12555-019-0984-6.

[40] Q. Wang, Y. Yu, H. O. A. Ahmed, M. Darwish, and A. K. Nandi, "Open-circuit fault detection and classification of modular multilevel converters in high voltage direct current systems (MMC-HVDC) with long short-term memory (LSTM) method," *Sensors*, vol. 21, no. 12, p. 4159, Jun. 2021, doi: 10.3390/s21124159.

[41] M. Rahman, D. Islam, R. J. Mukti, and I. Saha, "A deep learning approach based on convolutional LSTM for detecting diabetes," *Comput. Biol. Chem.*, vol. 88, Oct. 2020, Art. no. 107329, doi: 10.1016/j.compbiolchem.2020.107329.

[42] A. Kumar, S. Singh Negi, N. Kishor, and K. Uhlen, "Signal processing and classification of synchro-phasor data," in *Proc. 18th Medit. Electrotech. Conf. (MELECON)*, Apr. 2016, pp. 1–6.

[43] G. Rajshekhar and P. Rastogi, "Multiple signal classification technique for phase estimation from a fringe pattern," *Appl. Opt.*, vol. 51, no. 24, p. 5869, 2012, doi: 10.1364/AO.51.005869.

[44] Y. Izuhara, H. Matsumoto, T. Nagasaki, Y. Kanemitsu, K. Murase, I. Ito, T. Oguma, S. Muro, K. Asai, Y. Tabara, K. Takahashi, K. Bessho, A. Sekine, S. Kosugi, R. Yamada, T. Nakayama, F. Matsuda, A. Niimi, K. Chin, and M. Mishima, "Mouth breathing, another risk factor for asthma: The Nagahama study," *Allergy*, vol. 71, no. 7, pp. 1031–1036, Jul. 2016, doi: 10.1111/all.12885.

[45] C. B. Pereira, X. Yu, T. Goos, I. Reiss, T. Orlikowsky, K. Heimann, B. Venema, V. Blazek, S. Leonhardt, and D. Teichmann, "Noncontact monitoring of respiratory rate in newborn infants using thermal imaging," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 4, pp. 1105–1114, Apr. 2019, doi: 10.1109/TBME.2018.2866878.

JUN HWAN KWON received the B.S. degree in biomedical engineering from Konkuk University, South Korea, in 2016, and the M.S. and Ph.D. degrees in biomedical engineering from the Yonsei University College of Medicine, Seoul, South Korea, in 2018 and 2023, respectively. Since 2016, he has been a Research Assistant with the Severance Hospital, Seoul. Since 2023, he has been a Postdoctoral Researcher with the Department of Medical Engineering, Yonsei University College of Medicine. His current research interests include medical engineering, image processing, bio-signal, and thermal cameras.

JIWON CHOI received the B.S. degree in biomedical engineering from Konkuk University, South Korea, in 2020, and the M.S. degree in biomedical engineering from the Yonsei University College of Medicine, Seoul, South Korea, in 2023. In 2020, she was a Researcher with the Inter-University Semiconductor Research Center, Seoul National University. From 2023 to the present, she is a Researcher in the Korea Smart Healthcare Association (GOSHA), Seoul. Her current research interests include circuit design, embedded systems, bio-signal processing, artificial intelligence, and deep learning.

JEONGMIN KIM received the M.D., M.S., and Ph.D. degrees in medicine from the Yonsei University College of Medicine, Seoul, South Korea, in 2007, 2011, and 2015, respectively. Since 2013, she has been a Professor with the Department of Anesthesiology and Pain Medicine, Severance Hospital, Yonsei University College of Medicine.

KYEONG-TAEK OH received the B.S. degree in computer engineering from Korea Polytechnic University, South Korea, in 2015, and the M.S. and Ph.D. degrees in biomedical engineering from the Yonsei University College of Medicine, Seoul, South Korea, in 2017 and 2023, respectively. Since 2015, he has been a Research Assistant with the Severance Hospital, Seoul. Since 2023, he has been a Postdoctoral Researcher with the Department of Medical Engineering, Yonsei University College of Medicine. His current research interests include artificial intelligence, deep learning, pattern recognition, and medical imaging.

OYUN KWON received the B.S. degree in physics from Kyung-Hee University, Seoul, South Korea, in 2019, and the M.S. degree in medical device engineering and management from the Yonsei University College of Medicine, Seoul, in 2021, where he is currently pursuing the Ph.D. degree in biomedical engineering. His current research interests include bio-signal processing, artificial intelligence, deep learning, and u-Health.

SUN K. YOO received the B.S., M.S., and Ph.D. degrees in electrical engineering from Yonsei University, Seoul, South Korea, in 1981, 1985, and 1989, respectively. Since 1995, he has been a Professor with the Department of Medical Engineering, Yonsei University College of Medicine, Seoul. His current research interests include u-Health, medical imaging, smart devices, bio-signal processing, pattern recognition, and human sensibility engineering.