Detection of Respiratory Infections Using RGB-Infrared Sensors on Portable Device
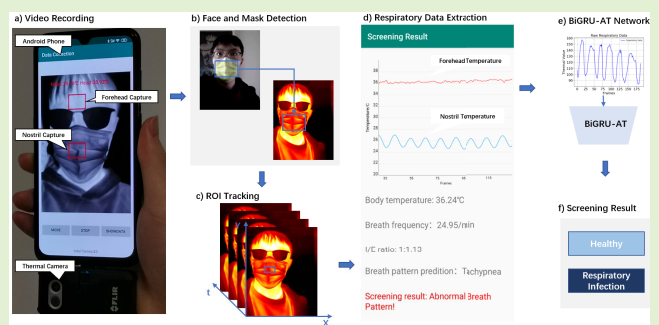
Zheng Jiang [iD], *Graduate Student Member, IEEE*, Menghan Hu [iD], *Member, IEEE*, Zhongpai Gao [iD], Lei Fan, Ranran Dai, Yaling Pan, Wei Tang, Guangtao Zhai [iD], and Yong Lu

***Abstract*—Coronavirus Disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronaviruses 2 (SARS-CoV-2) has become a serious global pandemic in the past few months and caused huge loss to human society worldwide. For such a large-scale pandemic, early detection and isolation of potential virus carriers is essential to curb the spread of the pandemic. Recent studies have shown that one important feature of COVID-19 is the abnormal respiratory status caused by viral infections. During the pandemic, many people tend to wear masks to reduce the risk of getting sick. Therefore, in this paper, we propose a portable non-contact method to screen the health conditions of people wearing masks through analysis of the respiratory characteristics from RGB-infrared sensors. We first accomplish a respiratory data capture technique for people wearing masks by using face recognition. Then, a bidirectional GRU neural network with an attention mechanism is applied to the respiratory data to obtain the health screening result. The results of validation experiments show that our model can identify the health status of respiratory with 83.69% accuracy, 90.23% sensitivity and 76.31% specificity on the real-world dataset. This work demonstrates that the proposed RGB-infrared sensors on portable device can be used as a pre-scan method for respiratory infections, which provides a theoretical basis to encourage controlled clinical trials and thus helps fight the current COVID-19 pandemic. The demo videos of the proposed system are available at: https://doi.org/10.6084/m9.figshare.12028032.**

***Index Terms*— COVID-19 pandemic, deep learning, dual-mode tomography, health screening, recurrent neural network, respiratory state, SARS-CoV-2, thermal imaging.**

Zheng Jiang, Zhongpai Gao, Lei Fan, and Guangtao Zhai are with the Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai 200240, China, and also with the Key Laboratory of Artificial Intelligence, Ministry of Education, Shanghai 200240, China (e-mail: zhaiguangtao@sjtu.edu.cn).

Menghan Hu is with the Shanghai Key Laboratory of Multidimensional Information Processing, East China Normal University, Shanghai 200062, China, and also with the Key Laboratory of Artificial Intelligence, Ministry of Education, Shanghai 200240, China.

Ranran Dai is with the Department of Pulmonary and Critical Care Medicine, Ruijin Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 200240, China.

Yaling Pan and Yong Lu are with the Department of Radiology, Ruijin Hospital Luwan Branch, School of Medicine, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: 18917762053@163.com).

Wei Tang is with the Department of Respiratory Disease, Ruijin Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 200240, China.

Digital Object Identifier 10.1109/JSEN.2020.3004568

I. INTRODUCTION

TO TACKLE the outbreak of the COVID-19 pandemic, early control is essential. Among all the control measures, efficient and safe identification of potential patients is the most important part. Existing researches show that the human physiological state can be perceived through breathing [1], which means respiratory signals are vital signs that can reflect human health conditions to a certain extent [2]. Many clinical literature suggests that abnormal respiratory symptoms may be important factors for the diagnosis of some specific diseases [3]. Recent studies have found that COVID-19 patients have obvious respiratory symptoms such as shortness of breath fever, tiredness, and dry cough [4], [5]. Among those symptoms, atypical or irregular breathing is considered as one of the early signs according to the recent research [6]. For many people, early mild respiratory symptoms are difficult to be recognized. Therefore, through the measurement of respiration conditions, potential COVID-19 patients can be screened to some extent. This may play an auxiliary diagnostic role, thus helping find potential patients as early as possible.

Traditional respiration measurement requires attachments of sensors to the patient's body [7]. The monitor of respiration is measured through the movement of the chest or abdomen.

Contact measurement equipment is bulky, expensive, and time-consuming. The most important thing is that the contact during measurement may increase the risk of spreading infectious diseases such as COVID-19. Therefore, the non-contact measurement is more suitable for the current situation. In recent years, many non-contact respiration measurement methods have been developed based on image sensors, doppler radar [8], depth camera [9] and thermal camera [10]. Considering factors such as safety, stability and price, the measurement technology of thermal imaging is the most suitable for extensive promotion. So far, thermal imaging has been used as a monitoring technology in a wide range of medical fields such as estimations of heart rate [11] and breathing rate [12]–[14]. Another important thing is that many existing respiration measurement devices are large and immovable. Given the worldwide pandemic, the portable and intelligent screening equipment is required to meet the needs of large-scale screening and other application scenarios in a real-time manner. For thermal imaging-based respiration measurement, nostril regions and mouth regions are the only focused regions since only these two parts have periodic heat exchange between the body and the outside environment. However, until now, researchers have barely considered measuring thermal respiration data for people wearing masks. During the epidemic of infectious diseases, masks may effectively suppress the spread of the virus according to recent studies [15], [16]. Therefore, to develop the respiration measurement method for people wearing masks becomes quite practical. In this study, we develop a portable and intelligent health screening device that uses thermal imaging to extract respiration data from masked people, which is then used to do the health screening classification via deep learning architecture.

In classification tasks, deep learning has achieved state-of-the-art performance in most research areas. Compared with traditional classifiers, classifiers based on deep learning can automatically identify the corresponding features and their correlations rather than extracting features manually. Recently, many researchers have developed detection methods of COVID-19 cases through medical imaging techniques such as chest X-ray imaging and chest CT imaging [17]–[19]. These studies have proved that deep learning can achieve high accuracy in detection of COVID-19. Based on the nature of the above methods, they can only be used for the examination of highly suspected patients in hospitals, and may not meet the requirements for the larger-scale screening in public places. Therefore, this paper proposes a scheme based on breath detection via a thermal camera.

For breathing tasks, deep learning-based algorithm can also better extract the corresponding features such as breathing rate and inhale-to-exhale ratio, and make more accurate predictions [20]–[23]. Recently, many researchers made use of deep learning to analyze the respiratory process. Cho *et al.* used a convolutional neural network (CNN) to analyze human breathing parameters to determine the degree of nervousness through thermal imaging [24]. Romero *et al.* applied a language model to detect acoustic events in sleep-disordered breathing through related sounds [25]. Wang *et al.* utilized deep learning and depth camera to classify abnormal respiratory patterns in real-time and achieved excellent results [9].

In this paper, we propose a remote, portable and intelligent health screening system based on respiratory data for pre-screening and auxiliary diagnosis of respiratory diseases like COVID-19. To be more practical in situations where people often choose to wear masks, the breathing data capture method for people wearing masks is introduced. After extracting breathing data from the videos obtained by the thermal camera, a deep learning neural network is performed to work on the classification between healthy and abnormal respiration conditions. To verify the robustness of our algorithm and the effectiveness of the proposed equipment, we analyze the influence of mask type, measurement distance and measurement angle on breathing data collection.

The main contributions of this paper are threefold. First, we combine the face recognition technology with dual-mode imaging to accomplish a respiratory data extraction method for people wearing masks, which is quite essential for the current situation. Based on our dual-camera algorithm, the respiration data is successfully obtained from masked facial thermal videos. Subsequently, we propose a classification method to judge abnormal respiratory states with a deep learning framework. Finally, based on the two contributions mentioned above, we have implemented a non-contact and efficient health screening system for respiratory infections using the collected data from the hospital, which may contribute to finding the possible cases of COVID-19 and keeping the control of the second spread of SARS-CoV-2.

II. METHOD

A brief introduction to the proposed respiration condition screening method is shown below. We first use the portable and intelligent screening device to get the thermal and the corresponding RGB videos. During the data collection, we also perform a simple real-time screening result. After getting the thermal videos, the first step is to extract respiration data from faces in thermal videos. During the extraction process, we use the face detection method to capture people's masked areas. Then a region of interest (ROI) selection algorithm is proposed to get the region from the mask that stands for the characteristic of breath most. Finally, we use a bidirectional GRU neural network with an attention mechanism (BiGRU-AT) model to work on the classification task with the input respiration data. A key point in our method is to collect respiration data from facial thermal videos, which has been proved to be effective by many previous studies [26]–[28].

A. Overview of the Portable and Intelligent Health Screening System for Respiratory Infections

Our data collection obtained by the system is shown in Fig. 1. The whole screening system includes a FLIR one thermal camera, an Android smartphone and the corresponding application we have written, which is used for data acquisition and simple instant analysis. Our screening equipment, whose main advantage is portable, can be easily applied to measure abnormal breathing in many occasions of instant detection.
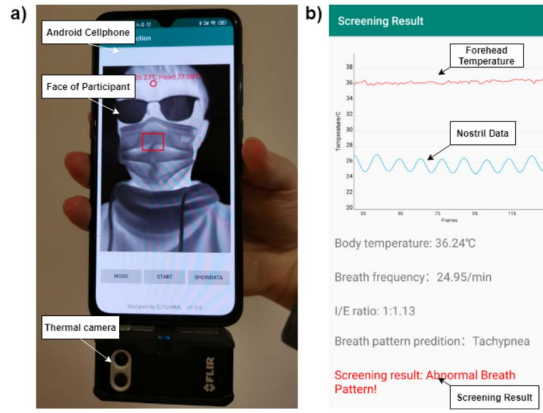
Fig. 1.　Overview of the portable and intelligent health screening system for respiratory infections: a) device appearance; b) analysis result of the application. (Notice that the system can simultaneously collect body temperature signals. In the current work, this body temperature signal is not considered in the model and is only used as a reference for the users.)

As shown in Fig. 1, the FLIR one thermal camera consists of two cameras, an RGB camera and a thermal camera. We collect the face videos from both cameras and use face recognition method to get the nostril area and forehead area. The temperatures of the two regions are calculated in time series and shown in the screening result page in Fig. 1(b). The red line stands for the body temperature and the blue line stands for breathing data. From the breathing data, we can predict the respiratory pattern of the test case. Then, a simple real-time screening result is given directly in the application according to the extracted features shown in Fig. 1. We use the raw face videos collected from both RGB camera and thermal camera as the data for further study to ensure accuracy and higher performance.

B. Detection of Masked Region From Dual Mode Image

When continuous breathing activities perform, there is a fact that periodic temperature fluctuations occur around the nostril due to the inspiration and expiration cycles. Therefore, respiration data can be obtained by analyzing the temperature data around the nostril based on the thermal image sequence. However, when people wear masks, many facial features are blocked because of this. Merely recognizing the face through thermal image will lose a lot of geometric and textural facial details, resulting in recognition errors of the face and mask parts. In order to solve this problem, we adopt the method based on two parallel located RGB and infrared cameras for face and mask region recognition. The masked region of the face is first captured in the RGB camera, then such a region is mapped to the thermal image with a related mapping function.

The algorithm for masked face detection is based on the pyramidbox model created by Tang et al. [29]. The main idea is to apply tricks like the Gaussian pyramidbox in deep learning to get the context correlations as further characteristics. The face image is first used to extract features of different scales using the Gaussian pyramid algorithm. For those high-level contextual features, a feature pyramid network is proposed to further excavate high-level contextual features. Then, the output and those low-level features are combined in low-level feature pyramid layers. Finally, the result is obtained after another two layers of deep neural network. For faces that a lot of features are lost due to the cover of a mask, such a context-sensitive structure can obtain more feature correlations and thus improve the accuracy of face detection. In our experiment, we use the open-source model from the paddle hub which is specially trained for masked faces to detect the face area on RGB videos.

The next step is to extract the masked area and map the area from RGB video to thermal video. Since the position of the mask on the human face is fixed, after obtaining the position coordinates of the human face, we obtain the mask area of the face by scaling down in equal proportions. For a detected face with width w, and height h, the location of left-up corner is defined as $(0, 0)$, the location of right-bottom corner is then (w, h). The corresponding coordinate of the two corners of the mask region is declared as $(w/4, h/2)$ and $(3w/4, 4h/5)$. Considering that the background to the boundary of the mask will produce a large contrast with the movement, which is easy to cause errors, we choose the center area of the mask through this division. Then the selected area is mapped from the RGB image to thermal image to obtain the masked region in thermal videos.

C. Extract Respiration Data From ROI

After getting the masked region in thermal videos, we need to get the region of interest (ROI) that represents breathing features. Recent studies often characterize breathing data through temperature changes around the nostril [11], [30]. However when people wear masks, there exists another problem that the nostrils are also blocked by the masks, and when people wearing different masks, the ROI may be different. Therefore, we perform an ROI tracking method based on maximizing the variance of thermal image sequence to extract a certain area on the masked region of the thermal video which stands for the breath signals most.

Due to the lack of texture features in masked regions compared to human faces, we judge the ROI from the temperature change of the thermal image sequence. The main idea is to traverse the masked region in the thermal images and find a small block with the largest temperature change as the selected ROI. The position of a certain block is fixed in the masked region among all the frames since the nostril area is fixed on the face region. We do not need to consider the movement of the block since our face recognition algorithm can detect the mask position in each frame's thermal image. For a certain block with height m, and width n, we define the average pixel intensity at frame t as:

$$\bar{s}(t) = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} s(i, j, t) \tag{1}$$

For thermal images, $\bar{s}(t)$ represents the temperature value at frame t. For every block we obtained, we calculate their $\bar{s}(t)$ on time line. Then, for each block n, the total variance of the list of average pixel intensity with T frames $\sigma_s^2(n)$ is
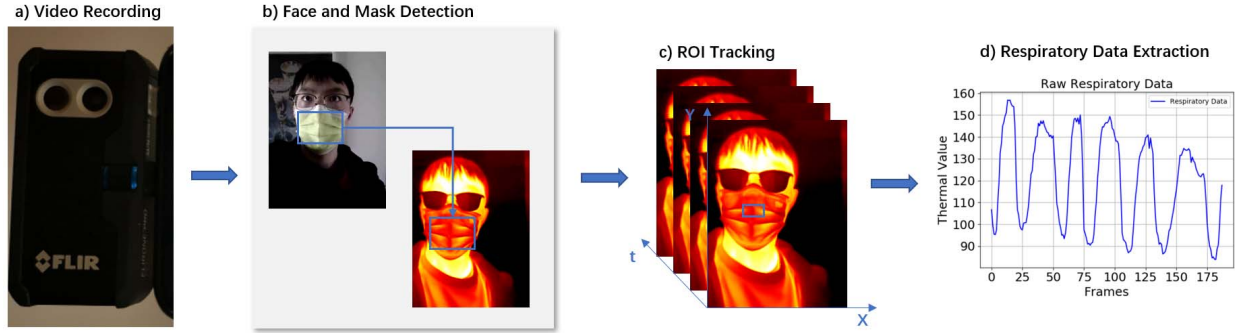
Fig. 2. The pipeline of the respiration data extraction: a) record the RGB video and thermal video through a FLIR one thermal camera; b) use face detection method to detect face and mask region in the RGB frames and then map the region to the thermal frames; c) capture the ROIs in the thermal frames of mask region by tracking method; d) extract the respiration data from the ROIs.
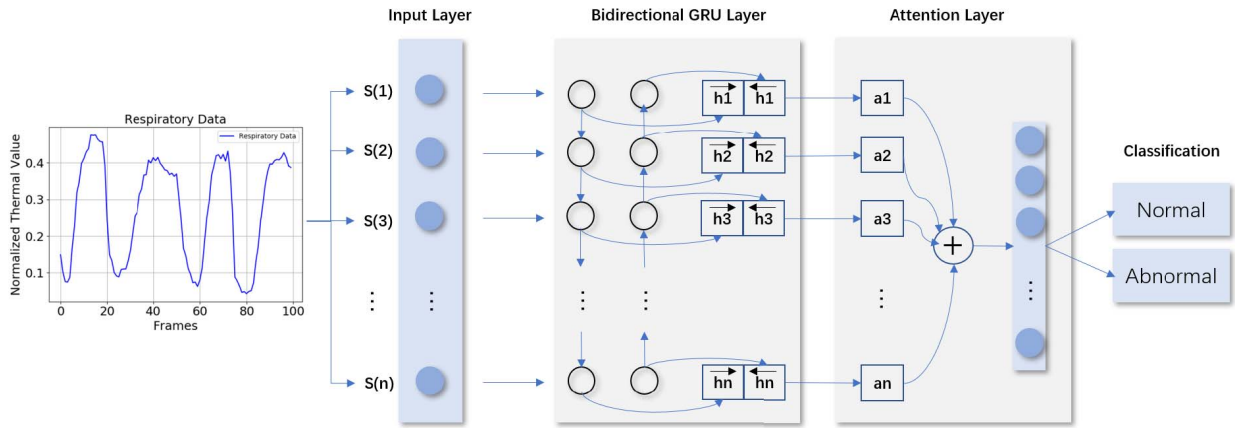

Fig. 3. The structure of the BiGRU-AT network: the network consists of four layers: the input layer, the bidirectional layer, the attention layer and an output layer. The output is a 2 dimension tensor which indicates normal or abnormal respiration condition.

calculated as shown in Eq. 2, where μ stands for the mean value of $\bar{s}(t)$.

$$\sigma_s^2(n) = \frac{\sum(\bar{s}(t) - \mu)^2}{T}(0 < t < T) \qquad (2)$$

Since respiration is a periodic data spread out from the nostril area, we can consider that the block with the largest variance is the position where the heat changes most in both frequency and value within the mask, which stands for the breath data mostly in the masked region. We adjust the corresponding block size according to the size of the masked region. For a masked region with N blocks, the final ROI is selected by:

$$ROI = \arg\max_{1 \leq n < N} \sigma_s^2(n) \qquad (3)$$

For each thermal video, we traverse all possible blocks in the mask regions of each frame and find the ROIs for each frame by the method above. The respiration data is then defined as $\bar{s}(t)(0 < t < T)$, which is the pixel intensities of ROIs in all the frames.

D. BiGRU-AT Neural Network

We apply a BiGRU-AT neural network to do the classification task on judging whether the respiration condition is

healthy or not as shown in Fig. 3. The input of the network is the respiration data obtained by our extraction method. Since the respiratory data is time series, it can be regarded as a time series classification problem. Therefore, we choose the Gate Recurrent Unit (GRU) network with the bi-direction and attention layer to work on the sequence prediction task.

Among all the deep learning structures, recurrent neural network (RNN) is a type of neural network which is specially used to process time series data samples [31]. For a time step t, the RNN model can be represented by:

$$h^{(t)} = \phi\left(Ux^{(t)} + Wh^{(t-1)} + b\right) \qquad (4)$$

$$o^{(t)} = Vh^{(t)} + c \qquad (5)$$

$$\widehat{y}^{(t)} = \sigma\left(o^{(t)}\right) \qquad (6)$$

where $x^{(t)}, h^{(t)}$ and $o^{(t)}$ stand for the current input state, hidden state and output at time step t respectively. V, W, U are parameters obtained by training procedure. b is the bias and σ and ϕ are activation functions. The final prediction is $\widehat{y}^{(t)}$.

Long-short term memory network is developed based on RNN [32]. Compared to RNN, which can only memorize and analyze short-term information, it can process relatively long-term information, and is suitable for problems with short-term delays or long time intervals. Based on LSTM,

many related structures are proposed in recent years [33]. GRU is a simplified LSTM which merges three gates of LSTM (forget, input and output) into two gates (update and reset) [34]. For tasks with a few data, GRU may be more suitable than LSTM since it includes less parameters. In our task, since the input of the neural network is only the respiration data in time sequence, the GRU network may perform a better result than LSTM network. The structure of GRU can be expressed by the following equations:

$$r_t = \sigma \left(W_r \cdot [h_{t-1}, x_t] + b_r \right) \tag{7}$$

$$z_t = \sigma \left(W_z \cdot [h_{t-1}, x_t] + b_z \right) \tag{8}$$

$$\tilde{h}_t = \tanh \left(W_{\tilde{h}} \cdot [r_t * h_{t-1}, x_t] + b_h \right) \tag{9}$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \tag{10}$$

where r_t is the reset gate that controls the amount of information being passed to the new state from the previous states. z_t stands for the update gate which determines the amount of information being forgotten and added. W_r, W_z and W_h are trained parameters that vary in the training procedure. $\tilde{h}_t$ is the candidate hidden layer which can be regarded as a summary of the above information h_{t-1} and the input information x_t at time step t. h_t is the output layer at time step t which will be sent to the next unit.

The bidirectional RNN has been widely used in natural language processing [35]. The advantage of such a network structure is that it can strengthen the correlation between the context of the sequence. As the respiratory data is a periodic sequence, we use bidirectional GRU to obtain more information from the periodic sequence. The difference between bidirectional GRU and normal GRU is that the backward sequence of data is spliced to the original forward sequence of data. In this way, the hidden layer of the original $h(t)$ is changed to:

$$h_t = [\overrightarrow{h_t}, \overleftarrow{h_t}] \tag{11}$$

where $\overrightarrow{h_t}$ is the original hidden layer and $\overleftarrow{h_t}$ is the backward sequence of $\overrightarrow{h_t}$.

During the analysis of respiratory data, the entire waveform in time sequence should be taken into consideration. For some specific breathing patterns such as asphyxia, several particular features such as sudden acceleration may occur only at a certain point in the entire process. However, if we only use the BiGRU network, these features may be weakened as the time sequence data is input step by step, which may cause a larger error in prediction. Therefore, we add an attention layer to the network, which can ensure that certain keypoint features in the breathing process can be maximized.

Attention mechanism is a choice to focus only on those important points among the total data [36]. It is often combined with neural networks like RNN. Before the RNN model summarizes the hidden states for the output, an attention layer can make an estimation of all outputs and find the most important ones. This mechanism has been widely used in many research areas. The structure of attention layer is:

$$u_t = \tanh \left(W_u h_t + b_w \right) \tag{12}$$

$$a_t = \frac{\exp \left(u_t^\top u_w \right)}{\sum_t \exp \left(u_t^\top u_w \right)} \tag{13}$$

$$s = \sum_t \alpha_t h_t \tag{14}$$

where h_t represents the BiGRU layer output at time step t, which is bidirectional. W_u and b_w are also parameters that vary in the training process. a_t performs a softmax function on u_t to get the weight of each step t. Finally, the output of the attention layer s is a combination of all the steps from BiGRU with different weights. By applying another softmax function to the output s, we get the final prediction of the classification task. The structure of the whole network is shown in Fig. 3.

III. EXPERIMENTS

A. Dataset Explanation and Experimental Settings

Our goal is to distinguish whether there is an epidemic of infectious disease such as COVID-19 according to the abnormal breathing in the respiratory system. We collect the healthy dataset from people around our authors. And the abnormal dataset was obtained from the inpatients of the respiratory disease department and cardiology department in Ruijin Hospital. Most of the patients we collected data from caught basic or chronic respiratory diseases. Only some of them have a fever, which is one of the typical respiratory symptoms of infectious diseases. Therefore, the body temperature is not taken into consideration in our current screening system.

We use a FLIR one thermal camera connected to an Android phone to work on the data collection. We collected data from 73 people. For each person, we collected two 20-second infrared and RGB videos with a sampling frequency of 10 Hz. During the data collection process, all testers were required to be about 50 cm away from the camera and face the camera directly to ensure data consistency. The testers were asked to stay still and breathe normally during the whole process, but small movements were allowed. For each infrared and RGB videos collected, we sampled in step of 100 frames with a sampling interval of 3 frames. Thus we finally obtained 1,925 healthy breathing data and 2,292 abnormal breathing data, a total of 4,217 data. Each piece of data consists of 100 frames of infrared and RGB videos in 10 seconds.

In the BiGRU-AT network, the hidden cells in the BiGRU layer and attention layers are 32 and 8 respectively. The breathing data is normalized before input into the neural network and we use cross-entropy as the loss function. During the training process, we separate the dataset into two parts. The training set includes 1,427 healthy breathing data and 1,780 abnormal breathing data. And the test set contains 498 healthy breathing data and 512 abnormal breathing data. Once this paper is accepted, we will release the dataset used in the current work for non-commercial users.

Among the whole dataset, we choose 4 typical respiratory data examples as shown in Fig. 4. Fig. 4(a) and Fig. 4(b) stand for the abnormal respiratory patterns extracted from patients.
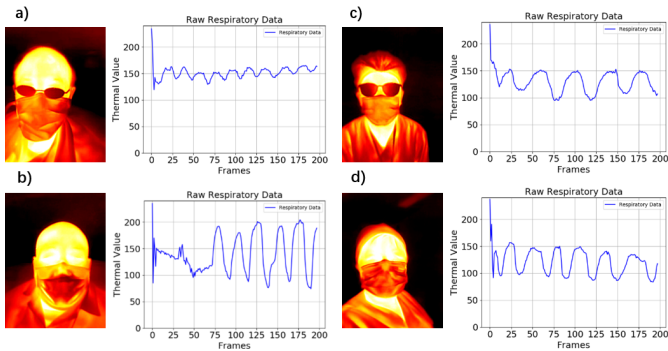
Fig. 4. Comparison of normal and abnormal respiratory data extracted by our method: a), b) are abnormal data collected from patients in the general ward of the respiratory department in Ruijin Hospital; c), d) are normal data collected from healthy volunteers.

TABLE I
EXPRIMENTAL RESULTS ON THE TEST SET

| Model	Accuracy	Sensitivity	Specificity	F1%
BiGRU-AT	83.69%	90.23%	76.31%	84.61%
GRU-AT	79.31%	90.62%	67.67%	81.62%
BiLSTM-AT	74.46%	87.50%	61.04%	77.64%
LSTM	71.98%	72.07%	74.49%	71.97%

Fig. 5. Confusion matrices of the four models. Each row is the number of real labels and each column is the number of predicted labels. The left one is the result of BiGRU-AT, the right one is the result of LSTM.

Fig. 4(c) and Fig. 4(d) represent the normal respiratory pattern called Eupnea from healthy participants. By comparison, we can find that the respiratory of normal people are in strong periodic and evenly distributed while abnormal respiratory data tend to be more irregular. Generally speaking, most abnormal breathing data from respiratory infections have faster frequency and irregular amplitude.

B. Experimental Result

The experimental results are shown in Table. I. We consider four evaluation metrics viz. Accuracy, Sensitivity, Specificity and F1. To measure the performance of our model, we compare the result of our model with three other models which are GRU-AT, BiLSTM-AT and LSTM respectively. The result of sensitivity reaches 90.23% which is far more higher than the specificity of 76.31%. This may have a positive effect on the screening of potential patients since the false negative rate is relatively low. Our method performs better than any other network in all evaluation metrics with the only exception in the sensitivity value of GRU-AT. By comparison, the experimental result demonstrates that attention mechanism is well-performed in keeping important node features in the time series of breathing data since the networks with attention layer all perform a better result than LSTM. Another point is that GRU based networks achieve better results than LSTM based networks. This may because our data set is relatively small which can't fill the demand of the LSTM based networks. GRU based networks require less data than LSTM and perform better result in our respiration condition classification task.

To figure out the detailed information about the classification of the respiratory state, we plotted the confusion matrix of the four models as demonstrated in Fig. 5. As can be seen
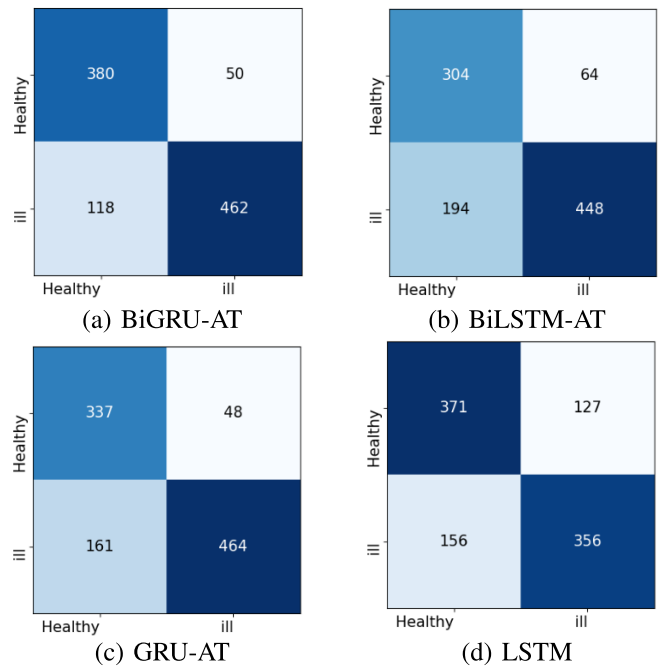
from the results, the performance improvement of BiGRU-AT compared to LSTM is mainly in the accuracy rate of the negative class. This is because many scatter-like abnormalities in the time series of abnormal breathing are better recognized by the attention mechanism. Besides, the misclassification rate of the four networks is relatively high to some extent which may be because many positive samples do not have typical respiratory infections characteristics.

C. Analysis

During the data collection process, all testers were required to be about 50 cm away from the camera and face the camera directly to ensure data consistency. However, in real-time conditions, the distance and angles of the testers towards the device cannot be so accurate. Therefore, in the analysis section, we give 3 comparisons from different aspects to prove the robustness of our algorithm and device.

1) Influence of Mask Types on Respiratory Data: To measure the robustness of our breathing data acquisition algorithm and the effectiveness of the proposed portable device, we analyze the breathing data of the same person wearing different masks. We design 3 mask-wearing scenarios that cover most situations: wearing one surgical mask (blue line); wearing one KN95 (N95) mask (red line) and wearing two surgical masks (green line). The results are shown in Fig. 6. It can be seen from the experimental results that no matter what kind of mask is worn, or even two masks, the respiratory data can be well recognized. This proves the stability of our algorithm and device. However, since different masks have different thermal insulation capabilities, the average breathing temperature may vary as the mask changes. To minimize this error, respiratory data are normalized before input into the neural network.
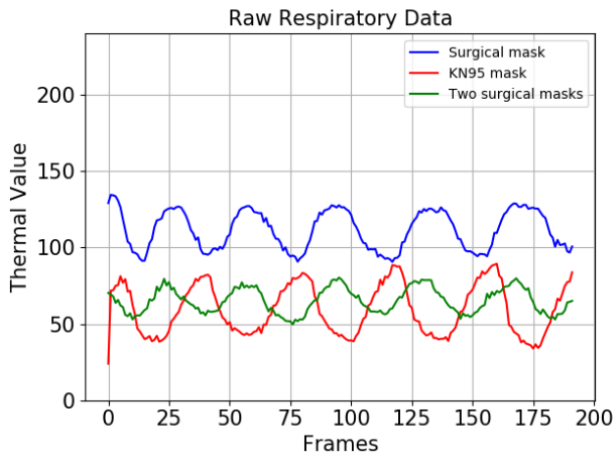
Fig. 6. The raw respiratory data obtained through the breathing data extraction algorithm with different types of masks.
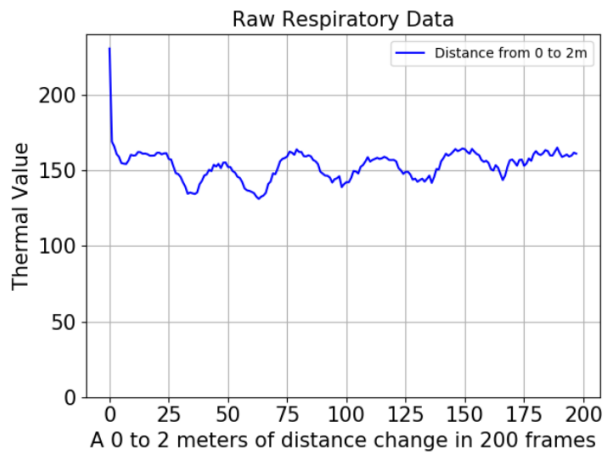

Fig. 7. The raw respiratory data which is obtained under the distance between camera and device from 0 to 2 meters. During the 200 frames' acquisition process, the distance between the tester and the device varies from 0.1 meters to 2 meters at a stable speed.
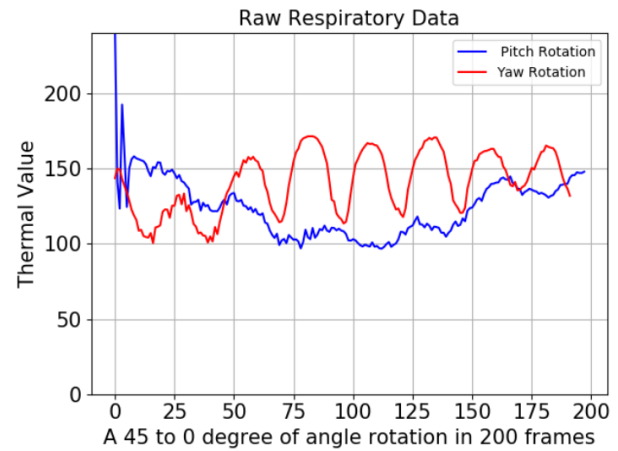

Fig. 8. The raw respiratory data obtained while the rotation angle varies from 45 degrees to 0 degrees at a stable speed. The blue line stands for the rotation in pitch axis and the red line stands for the rotation in yaw axis.

3) Limitation of Rotations Toward the Camera During Measurement: Considering that breath detection will be applied in different scenarios, we cannot assure that the testers would face the device directly with no error angle. Therefore, this experiment is designed to show the respiratory data extraction performance when faces rotate in pitch axis and yaw axis. We define the camera directly towards the face to be 0 degrees, and design an experiment in which the rotation angle gradually changed from 45 degrees to 0 degrees. We consider the transformation of two rotation axes: yaw and pitch, which respectively represent left and right turning and nodding. The results in the two cases are quite different as shown in Fig. 8. Our algorithm and device maintain good results in yaw rotation, but it is difficult to obtain precise respiratory data in pitch rotation. This means participants can turn left or turn right during the measurement but can't nod or head up a lot since this may impact the current measurement result.

IV. CONCLUSION

In this paper, we propose an abnormal breathing detection method based on a portable dual-mode camera which can record both RGB and thermal videos. In our detection method, we first accomplished an accurate and robust respiratory data detection algorithm which can precisely extract breathing data from people wearing masks. Then, a BiGRU-AT network is applied to work on the screening of respiratory infections. In validation experiments, the obtained BiGRU-AT network achieves a relatively good result with an accuracy of 83.7% on the real-world dataset. For patients caught respiratory diseases, the sensitivity value is 90.23%. It is foreseeable that among patients with COVID-19 who have more clinical respiratory symptoms, this classification method may yield better results. The result of our experiments provide a theoretical basis for scanning respiratory diseases through thermal respiratory data, which serves as an encouragement for controlled clinical trials and call for labeled breath data. During the current spread of COVID-19, our research can work as a pre-scan method for

2) Limitation of Distance to the Camera During Measurement: To verify the robustness of our algorithm and device in different scenarios, we design experiments to collect respiratory data at different distances. Considering the limitations of handheld devices, we test the collection of facial respiration data from a distance of 0 to 2 meters. During one 20 seconds' data collection process, the distance between the tester and the device varies from 0 to 2 meters at a stable speed. The result is demonstrated in Fig. 7. The signal tends to be periodic from the position of 0.1 meters, and it does not lose regularity until about 1.8 meters. At a distance of about 10 centimeters, the complete face begins to appear in the camera video. When the distance comes to 1.8 meters, our face detection algorithm begins to fail gradually due to the distance and pixel limitation. This experiment verifies that our algorithm and device can guarantee relatively accurate measurement results in the distance range of 0.1 meters to 1.8 meters. In future research, we will improve the effective distance through improvements in algorithm and device precision.

abnormal breathing in many scenarios such as communities, campuses and hospitals, contributing to distinguishing the possible cases, and then slowing down the spreading of the virus.

In future research, based on ensuring portability, we plan to use a more stable algorithm to minimize the effects caused by different masks on the measurement of breathing conditions. Besides, temperature may be taken into consideration to achieve a higher detection accuracy on respiratory infections.

REFERENCES

[1] M. A. Cretikos, R. Bellomo, K. Hillman, J. Chen, S. Finfer, and A. Flabouris, "Respiratory rate: The neglected vital sign," *Med. J. Aust.*, vol. 188, no. 11, pp. 657–659, Jun. 2008.

[2] A. D. Droitcour *et al.*, "Non-contact respiratory rate measurement validation for hospitalized patients," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Sep. 2009, pp. 4812–4815.

[3] R. Boulding, R. Stacey, R. Niven, and S. J. Fowler, "Dysfunctional breathing: A review of the literature and proposal for classification," *Eur. Respiratory Rev.*, vol. 25, no. 141, pp. 287–294, Sep. 2016.

[4] Z. Xu *et al.*, "Pathological findings of COVID-19 associated with acute respiratory distress syndrome," *Lancet Respiratory Med.*, vol. 8, no. 4, pp. 420–422, Apr. 2020.

[5] C. Sohrabi *et al.*, "World health organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19)," *Int. J. Surg.*, vol. 76, pp. 71–76, Apr. 2020.

[6] E. J. Chow *et al.*, "Symptom screening at illness onset of health care personnel with SARS-CoV-2 infection in King County, Washington," *JAMA*, vol. 323, no. 20, p. 2087, May 2020.

[7] F. Q. Al-Khalidi, R. Saatchi, D. Burke, H. Elphick, and S. Tan, "Respiration rate monitoring methods: A review," *Pediatric Pulmonol.*, vol. 46, no. 6, pp. 523–529, Jun. 2011.

[8] J. Kranjec, S. Beguš, J. Drnovšek, and G. Geršak, "Novel methods for noncontact heart rate measurement: A feasibility study," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 4, pp. 838–847, Apr. 2014.

[9] Y. Wang, M. Hu, Q. Li, X.-P. Zhang, G. Zhai, and N. Yao, "Abnormal respiratory patterns classifier may contribute to large-scale screening of people infected with COVID-19 in an accurate and unobtrusive manner," 2020, *arXiv:2002.05534*. [Online]. Available: http://arxiv.org/abs/2002.05534

[10] M.-H. Hu, G.-T. Zhai, D. Li, Y.-Z. Fan, X.-H. Chen, and X.-K. Yang, "Synergetic use of thermal and visible imaging techniques for contactless and unobtrusive breathing measurement," *J. Biomed. Opt.*, vol. 22, no. 3, Mar. 2017, Art. no. 036006.

[11] M. Hu *et al.*, "Combination of near-infrared and thermal imaging techniques for the remote and simultaneous measurements of breathing and heart rates under sleep situation," *PLoS ONE*, vol. 13, no. 1, Jan. 2018, Art. no. e0190466.

[12] C. B. Pereira, X. Yu, M. Czaplik, R. Rossaint, V. Blazek, and S. Leonhardt, "Remote monitoring of breathing dynamics using infrared thermography," *Biomed. Opt. Express*, vol. 6, no. 11, pp. 4378–4394, 2015.

[13] G. F. Lewis, R. G. Gatto, and S. W. Porges, "A novel method for extracting respiration rate and relative tidal volume from infrared thermography," *Psychophysiology*, vol. 48, no. 7, pp. 877–887, Jul. 2011.

[14] L. Chen, N. Liu, M. Hu, and G. Zhai, "RGB-thermal imaging system collaborated with marker tracking for remote breathing rate measurement," in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, Dec. 2019, pp. 1–4.

[15] S. Feng, C. Shen, N. Xia, W. Song, M. Fan, and B. J. Cowling, "Rational use of face masks in the COVID-19 pandemic," *Lancet Respiratory Med.*, vol. 8, no. 5, pp. 434–436, May 2020.

[16] C. C. Leung, T. H. Lam, and K. K. Cheng, "Mass masking in the COVID-19 epidemic: People need guidance," *Lancet*, vol. 395, no. 10228, p. 945, Mar. 2020.

[17] L. Wang and A. Wong, "COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images," 2020, *arXiv:2003.09871*. [Online]. Available: http://arxiv.org/abs/2003.09871

[18] O. Gozes *et al.*, "Rapid AI development cycle for the coronavirus (COVID-19) pandemic: Initial results for automated detection & patient monitoring using deep learning CT image analysis," 2020, *arXiv:2003.05037*. [Online]. Available: http://arxiv.org/abs/2003.05037

[19] L. Li *et al.*, "Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT," *Radiology*, Mar. 2020, Art. no. 200905.

[20] J. Chauhan, J. Rajasegaran, S. Seneviratne, A. Misra, A. Seneviratne, and Y. Lee, "Performance characterization of deep learning models for breathing-based authentication on resource-constrained devices," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 2, no. 4, pp. 1–24, Dec. 2018.

[21] B. Liu *et al.*, "Deep learning versus professional healthcare equipment: A fine-grained breathing rate monitoring model," *Mobile Inf. Syst.*, vol. 2018, pp. 1–9, Jan. 2018.

[22] Q. Zhang, X. Chen, Q. Zhan, T. Yang, and S. Xia, "Respiration-based emotion recognition with deep learning," *Comput. Ind.*, vols. 92–93, pp. 84–90, Nov. 2017.

[23] U. M. Khan, Z. Kabir, S. A. Hassan, and S. H. Ahmed, "A deep learning framework using passive WiFi sensing for respiration monitoring," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2017, pp. 1–6.

[24] Y. Cho, N. Bianchi-Berthouze, and S. J. Julier, "DeepBreath: Deep learning of breathing patterns for automatic stress recognition using low-cost thermal imaging in unconstrained settings," in *Proc. 7th Int. Conf. Affect. Comput. Intell. Interact. (ACII)*, Oct. 2017, pp. 456–463.

[25] H. E. Romero, N. Ma, G. J. Brown, A. V. Beeston, and M. Hasan, "Deep learning features for robust detection of acoustic events in sleep-disordered breathing," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 810–814.

[26] Z. Zhu, J. Fei, and I. Pavlidis, "Tracking human breath in infrared imaging," in *Proc. 5th IEEE Symp. Bioinf. Bioeng. (BIBE)*, Oct. 2005, pp. 227–231.

[27] C. B. Pereira, X. Yu, M. Czaplik, V. Blazek, B. Venema, and S. Leonhardt, "Estimation of breathing rate in thermal imaging videos: A pilot study on healthy human subjects," *J. Clin. Monit. Comput.*, vol. 31, no. 6, pp. 1241–1254, Dec. 2017.

[28] A. Ishida and K. Murakami, "Extraction of nostril regions using periodical thermal change for breath monitoring," in *Proc. Int. Workshop Adv. Image Technol. (IWAIT)*, Jan. 2018, pp. 1–5.

[29] X. Tang, D. K. Du, Z. He, and J. Liu, "Pyramidbox: A context-assisted single shot face detector," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 797–813.

[30] Y. Cho, S. J. Julier, N. Marquardt, and N. Bianchi-Berthouze, "Robust tracking of respiratory rate in high-dynamic range scenes using mobile thermal imaging," *Biomed. Opt. Express*, vol. 8, no. 10, pp. 4480–4503, 2017.

[31] J. L. Elman, "Finding structure in time," *Cognit. Sci.*, vol. 14, no. 2, pp. 179–211, Mar. 1990.

[32] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[33] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017.

[34] K. Cho *et al.*, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," 2014, *arXiv:1406.1078*. [Online]. Available: http://arxiv.org/abs/1406.1078

[35] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, *arXiv:1409.0473*. [Online]. Available: http://arxiv.org/abs/1409.0473

[36] A. Vaswani *et al.*, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.