



DNN based reliability evaluation for telemedicine data

Dong Ah Shin¹ · Jiwoon Kim² · Seong-Wook Choi^{2,3} · Jung Chan Lee⁴

Received: 19 April 2022 / Revised: 18 September 2022 / Accepted: 24 September 2022
© Korean Society of Medical and Biological Engineering 2022

Abstract

Telemedicine data are measured directly by untrained patients, which may cause problems in data reliability. Many deep learning-based studies have been conducted to improve the quality of measurement data. However, they could not provide an accurate basis for judgment. Therefore, this study proposed a deep neural network filter-based reliability evaluation system that could present an accurate basis for judgment and verified its reliability by evaluating photoplethysmography signal and change in data quality according to judgment criteria through clinical trials. In the results, the deviation of 3% or more when the oxygen saturation was judged as normal according to each criterion was 0.3% and 0.82% for criteria 1 and 2, respectively, which was very low compared to the abnormal judgment (3.86%). The deviation of diastolic blood pressure (≥ 10 mmHg) according to criterion 3 was reduced by about 4% in the normal judgment compared to the abnormal. In addition, when multiple judgment conditions were satisfied, abnormal data were better discriminated than when only one criterion was satisfied. Therefore, the basis for judging abnormal data can be presented with the system proposed in this study, and the quality of telemedicine data can be improved according to the judgment result.

Keywords Telemedicine · Photoplethysmography · Deep neural network · Reliability evaluation

1 Introduction

Telemedicine allows both patients and doctors to receive convenient care as health care providers can care for many people without being physically there. Because of its value, telemedicine has grown gradually. However, as the demand for telemedicine is rapidly increasing due to social distancing caused by the recent coronavirus (COVID-19) pandemic, the amount of remotely transmitted medical data is

increasing significantly [1]. Among various telemedicine data, a photoplethysmography (PPG) signal is the most used indicator for telemedicine because it can provide important information such as heart rate (HR) and oxygen saturation (SpO₂) [2, 3]. PPG signal might have different measurement performances depending on body movement, measurement environment, and measurement equipment. However, since telemedicine data are data measured by a patient who has not received professional training, the quality of PPG signal might be lower than that measured by a medical professional. Estimated HR and SpO₂ from low-quality PPG signals may differ from normal measurements, which may cause false alarms. This is directly related to the reliability of measurement data [4, 5]. In addition, since there is a limit for medical staff to review a vast amount of telemedicine data, additional manpower for data analysis is required, which may increase medical expenses [6, 7]. Therefore, various studies based on deep learning have been performed to improve the data quality of these PPG signals.

1.1 Related works and objectives

Most previous studies presented results of detecting noise signals with very high accuracy (> 91%) using various deep

✉ Seong-Wook Choi
swchoe@kangwon.ac.kr

✉ Jung Chan Lee
ljch@snu.ac.kr

¹ Institute of Medical and Biological Engineering, Medical Research Center, Seoul National University, Seoul 03080, Republic of Korea

² Interdisciplinary Program in Biohealth-Machinery Convergence Engineering, Kangwon National University, Chuncheon-si 24341, Republic of Korea

³ Program of Mechanical and Biomedical Engineering, College of Engineering, Chuncheon-si 24341, Republic of Korea

⁴ Department of Biomedical Engineering, Seoul National University College of Medicine and Seoul National University Hospital, Seoul 03080, Republic of Korea

learning techniques [8–12]. Roy et al., Lim et al., and Goh et al. [8–11] have performed studies on the improvement of PPG signal quality based on waveform. These studies classified good and bad PPG signals through waveform segments and proposed a technique for improving signal quality by reconstructing the morphology of a signal mixed with noise with reference to a clean signal. They found that PPG signal quality could be improved by about 35% or more through the template matched method. However, this method might damage various information such as blood vessel characteristics and irregular heart rate (arrhythmias) of actual patients that could be extracted from morphological characteristics of signals. In addition, in deep learning-based signal quality determination, it might be difficult to improve signal quality because there is no accurate basis for judging abnormal data. Prasad et al. [12] have proposed a method based on a feature extracted from PPG signal and determined the signal quality through seven feature sets including extracted kurtosis and entropy in time and frequency domains. In their study, it was possible to identify partially clean signals and noisy signals with a high accuracy (> 97%). However, characteristics of the feature are not intuitive. Also, whether the quality of data was improved by detecting noise signal was not provided in that study.

Therefore, in this study, a simple deep neural network (DNN) based reliability evaluation system for extracting intuitive features was proposed and the reliability of the PPG signal was evaluated through clinical trials. In addition, it was verified whether the quality of telemedicine data could be improved during re-measurement by proposing the basis for judging signal anomalies through the proposed system and providing feedback to the patient based on this.

To evaluate the reliability of the signal, the DNN filter-based system proposed in previous studies was applied [13–16]. The DNN filter technique can extract valid indicators (six singular points) from the PPG signal by applying a neural network in the form of a digital filter. The recognition score (RS), which means the reliability of the waveform, can be obtained from the DNN filter. The explainable HR could be extracted with a high accuracy. This DNN filter technique could present an indicator (RS) that can be explained. Therefore, in this study, the DNN filter technique was applied to our evaluation system. Remeasurement conditions were also defined to evaluate signal quality improvement.

2 Materials and methods

2.1 DNN algorithm

Raw PPG signal was transformed into input data through the following pre-processing process. After downsampling to 62.5 Hz, the most recent 74 data (window length=74, 1.18 s) were

converted into the first derivative signal (VPG). Raw PPG and VPG signals were standardized through the min–max normalization method. By specifying the target region (20 data, 0.32 s) within a window of the normalized signal, 20 data were acquired for PPG and VPG signals, respectively. Outside that region, 12 data at the specific points were extracted for each signal (PPG, VPG). Also, in the same target region, information on high peaks (20 data) and low peaks (20 data) of raw PPG signals and differential peaks (20 data) were extracted through peak detection. Thus, 124 input datasets were configured for each window.

$$\begin{aligned}
 RS = & RS_S[S \text{ point index}] + RS_O[O \text{ point index}] \\
 & + RS_W[W \text{ point index}] \\
 & \left(20 - \sum_{k=w \text{ point index}-30}^{w \text{ point index}+30} RS_{ES[k]} \right) \\
 & + \left(20 - \sum_{k=w \text{ point index}-30}^{w \text{ point index}+30} RS_{EE[k]} \right)
 \end{aligned} \quad (1)$$

The DNN filter consists of six deep learning filters. It plays a role in finding six singular points (S, O, W, Z, ES, EE) that are common in PPG waveforms (Fig. 1a). S, O, W, and Z are features extracted from the normal section of the waveform. S is the systolic peak. O is the pulse onset. W is the maximum slope before systolic. Z is the maximum slope before diastolic. ES and EE are abnormal sections of the waveform. ES is Error start and EE is Error End. Each DNN filter consists of one input layer, two hidden layers, and one output layer as shown in Fig. 1b [13–16].

The output of each DNN filter represents the location information of the singular point as 0 and 1 (output=21). The window of input signal strides by 1 to generate 20 accumulated values on the target region of output for the location information. This value is the recognition score of the singular point ($RS_{singularpoint}$) per beat. Therefore, each filter outputs the $RS_{singularpoint}$ values for 5 singular points with a maximum score of 20. The total RS for one heartbeat can be calculated with Eq. (1). Point Z represents the second peak of the PPG waveform generated by the reflection of the pulse from the peripheral vessels, not by the heartbeat, which may decrease or disappear depending on the patient's physical condition. Therefore, it was not related to the reliability of measured data and was excluded from Eq. 1 for RS calculation.

RS represents an index for judging the reliability of the signal. It has the value between 0 and 100. If $RS \geq 80$ is determined as a reliable region, and $RS < 80$ is determined as a noisy region. This value determines the first normal/abnormal judgement of the signal [13–16].

HR is determined based on the W value among singularities, and is calculated using the location information of the region with $RS \geq 80$ among the W position values (Eq. (2)). Calculated HR is used for determining criteria.

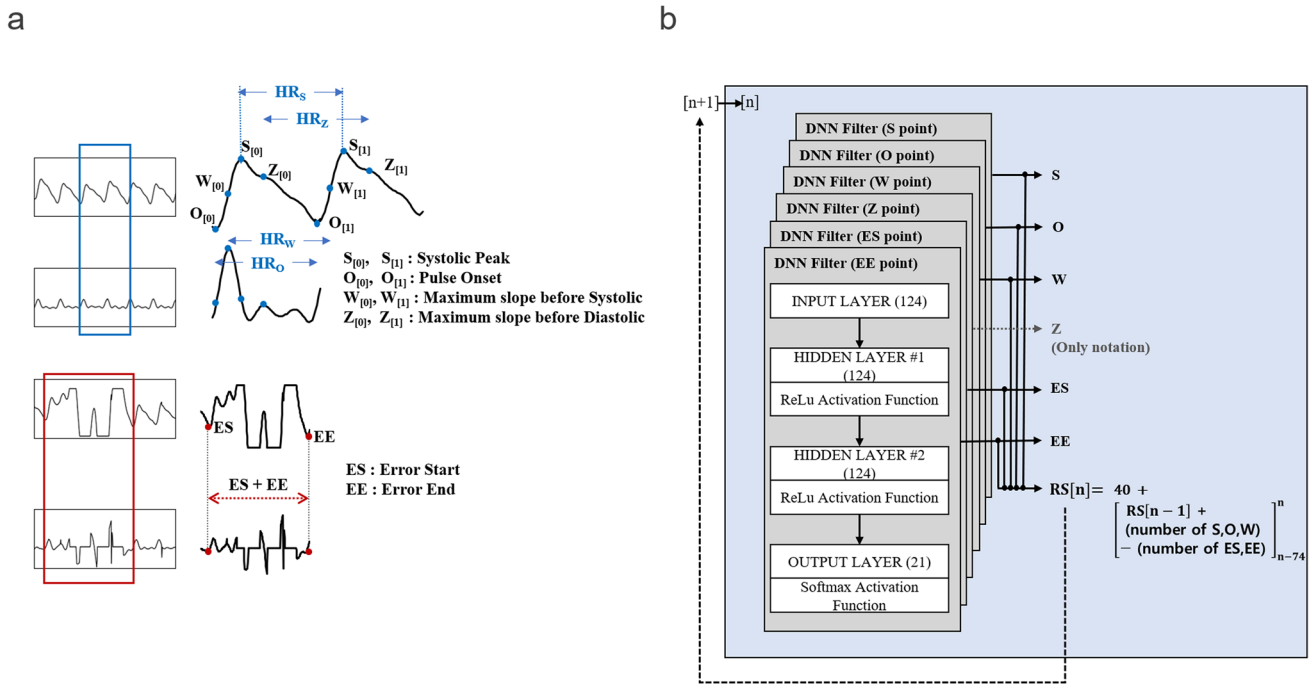


Fig. 1 **a** Deep neural network (DNN) filter-based heart rate measurement algorithm finds each S, O, W, Z, ES, and EE singularity in photoplethysmography (PPG) and the first derivative of the PPG (i.e., VPG) waveforms to measure HR. **b** Configuration for each DNN filter

$$HR = \frac{60}{Average(Time_{[present\ W\ point]} - Time_{[previous\ W\ point]})}$$

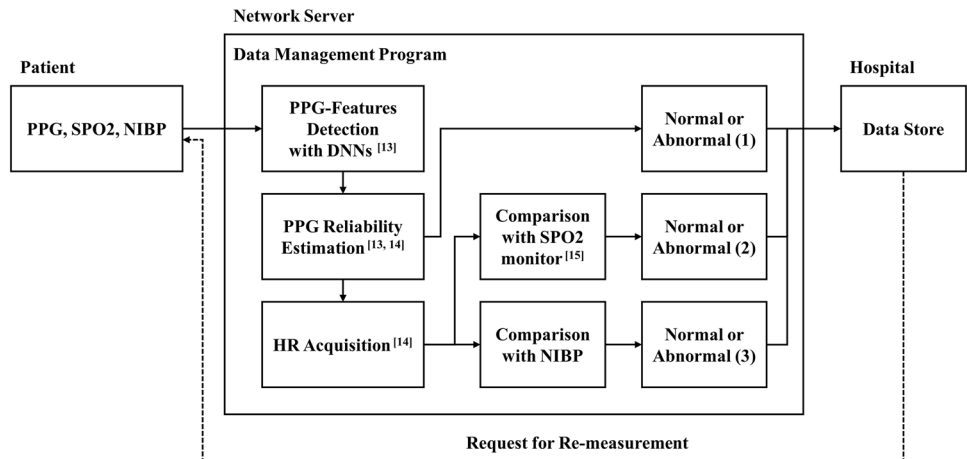
(when $RS_{[previous]}$ and succeeding $RS_{[present]} \geq 80$)

(2)

2.2 Reliability evaluation system

Figure 2 show a block diagram of a DNN filter-based reliability evaluation system. First, PPG signal is evaluated through a DNN filter to determine normal/abnormal.

Fig. 2 Block diagram showing a reliability evaluation system. It includes the heart rate measurement algorithm based on the deep neural network filter developed in previous studies



Then, HR estimated from PPG signal is compared with the measured one from SpO2 and non-invasive blood pressure (NIBP) devices (Bionics Co., BPM-190, Korea), respectively. Abnormal/normal is determined according to the remeasurement criteria. Judgment data are transmitted to the system network and stored together with judgment results and evidence.

2.3 Criteria of remeasurements

The first re-measurement criterion (Re-m1) is determined by RS. It is based on when waveforms with RS of 80 or higher

are less than 90% of the total. Motion artifact in the PPG waveform is determined based on this criterion. The HR of the normal waveform is estimated. HR value estimated through the first judgment is used as a comparison value between criteria 2 and 3.

The second re-measurement criterion (Re-m2) is when the difference between HR measured by the reference device and that estimated by the DNN filter is 10 bpm or more. The reference instrument uses a SpO2 measuring device.

The third re-measurement criterion (Re-m3) uses the NIBP device as the reference. The judgment standard is the same as that of Re-m2.

The judgment result according to each criterion is stored in the network. If re-measurement is required according to the judgment result, the system sends a message to the patient that re-measurement is necessary.

2.4 Experimental protocol and participants

Reliability evaluation was conducted for outpatients 19 years of age or older who were diagnosed with hypertension or diabetes at a primary medical institution. Patients who needed treatment at a secondary or tertiary medical institution due to complications were excluded. Only those who could store PPG data were included among the recruited patients ($N = 128$).

The number of participants for each disease consisted of 76 patients with hypertension, 26 patients with diabetes, and 26 patients diagnosed with both diseases. When classified by age, there were 17 patients over 70 years old, 37 patients over 60 years old, 48 patients over 50 years old, and 26 patients under 50 years old, showing a high distribution of patients over 50 years of age.

All participants in this study were previously trained on how to use a patient monitoring device (Bionics Co., BPM-190, Korea) in a primary care institution to measure SpO2,

NIBP, and HR on their own. They were instructed to self-measure at home for three months. NIBP measurement was performed immediately after SpO2 measurement according to the study design. Measured biometric data was stored in the developed server.

The clinical trial protocol of this study was approved by the Institutional Review Board of Kangwon National University Hospital (KNUH-2020-06-008-008). Subjects voluntarily participated in this study after receiving the explanation of this study. They provided written informed consent for all matters necessary for the experiment. All methods were carried out following relevant guidelines and regulations.

3 Results

3.1 Deciding the basis for criteria based on RS

To establish the basis for the criteria for re-measurement, change in the error rate of HR estimated from the DNN filter and HR of the reference equipment (SpO2, NIBP) was confirmed (Fig. 3). For change in error rate, the ratio of beats with RS of 80 or higher determined from the DNN filter was compared between 0 and 100% at 10% intervals.

Figure 3a presents error distribution of HR with an SpO2 device, showing that the result has an error of less than 5 bpm. In addition, the error of 10 bpm or more increased as the ratio of beats with $RS > 80$ was lowered.

Figure 3b shows error distribution of an SpO2 device with a NIBP device. Although both devices are reference equipment, the error tended to increase by more than 10 bpm as the ratio of $RS > 80$ decreased. NIBP was measured immediately after SpO2 was measured for 30 s.

From this result, if there is a difference of 10 bpm or more, it could mean that more motion artifacts are included

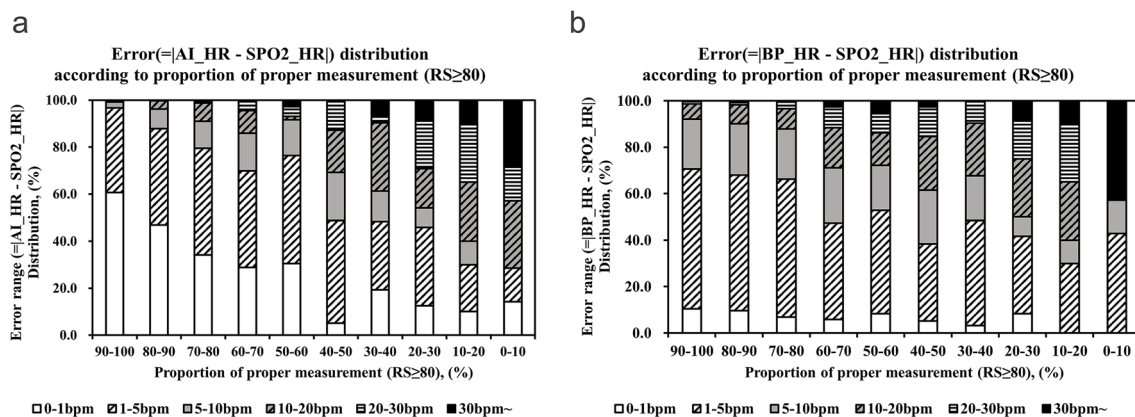


Fig. 3 Distribution of error rates in the heart rate measurements with decreasing proportions of $RS > 80$. **a** Comparison result between deep neural network filter and oxygen saturation (SpO2) device. **b** Comparison between non-invasive blood pressure (NIBP) and SpO2 device

in the measurement. Therefore, re-measurement criteria 2 and 3 were suggested based on a difference of 10 bpm for each reference equipment.

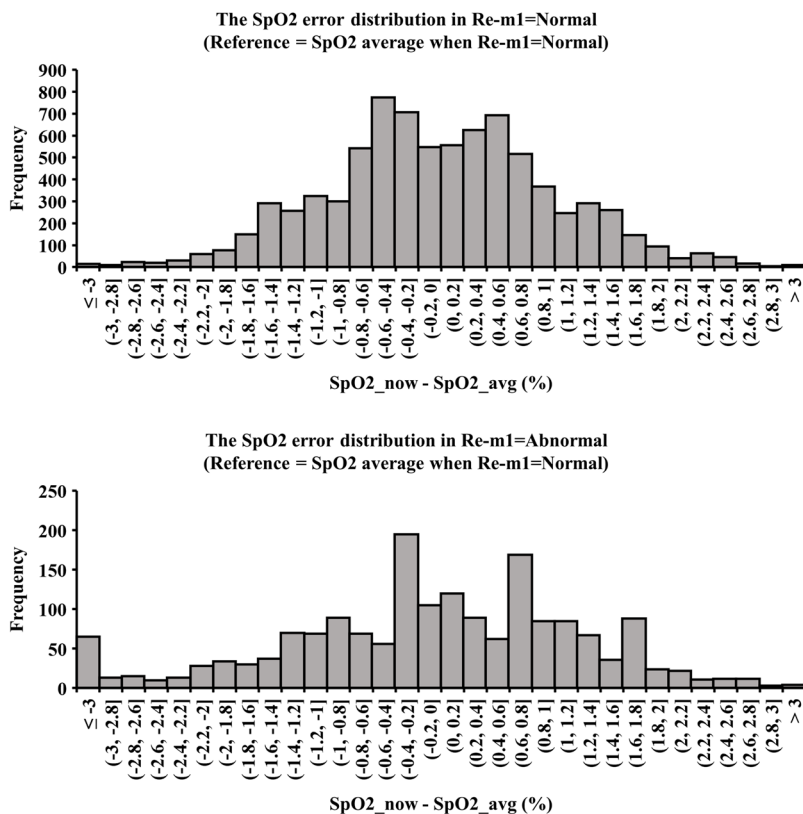
3.2 Quality assessments for re-measurement criteria

To verify the qualitative difference in telemedicine data by decision criteria, the distribution of changes in SpO2 and diastolic blood pressure (DBP) when the decision was made to re-measurement or normal according to each re-measurement criterion is shown.

Figures 4 and 5 show amounts of change of SpO2 for Re-m1 and Re-m2, respectively. Both graphs are histograms showing differences between the measured SpO2 value and the average SpO2 value on the same day. The upper graph is the result of a normal judgment and the lower graph is the result of a re-measurement judgment. From these histograms, it can be seen that when a re-measurement judgment occurs, SpO2 changes are larger than when it is normal. Comparing cases where there was a difference of 3% or more between normally measured SpO2 value and average SpO2 value according to each re-measurement criterion, normal

measurements in Re-m1 had about 0.3% of total measurements (24 out of 8,102 cases). This showed that the number of cases with a difference of 3% or more from the average SpO2 value when a re-measurement decision was made was much lower than about 3.86% (69 out of 1,787 cases). Similarly, in the Re-m2 result, the normal measurement was about 0.82% of the total number of measurements (76 out of 9,666 cases), which was very low. However, in re-measurement judgment, the number of cases with a difference of 3% or more from the average SpO2 value was about 5.38% out of a total of 233 cases.

Figure 6 shows changes in DBP with respect to Re-m3. Similar to Re-m1 and Re-m2 results, it also showed that the DBP deviation occurred more diversely when the re-measurement decision occurred. When a difference of 10 mmHg or more occurred between the normally measured DBP and the average DBP, compared with the case of normal and the re-measurement, the normal measurement was about 11.92% (1,078 out of 9,043 cases) and the decision to re-measure was about 15.13% (128 out of a total of 846 cases). DBP deviations of 20 mmHg or more occurred in 92 out of 9046 normal and 27 out of 846 abnormal.



SpO2 _{now} - SpO2 _{avg} = x (%)	Normal (%)	Abnormal (%)
$ x \geq 3$	0.29	3.86

Fig. 4 Oxygen saturation (SpO2) deviation according to normal and abnormal judgment results for Re-m1. The upper part shows result of normal judgments and the lower part shows result of abnormal judgments.

The table shows percentage of data with a deviation of 3% or more depending on the judgment

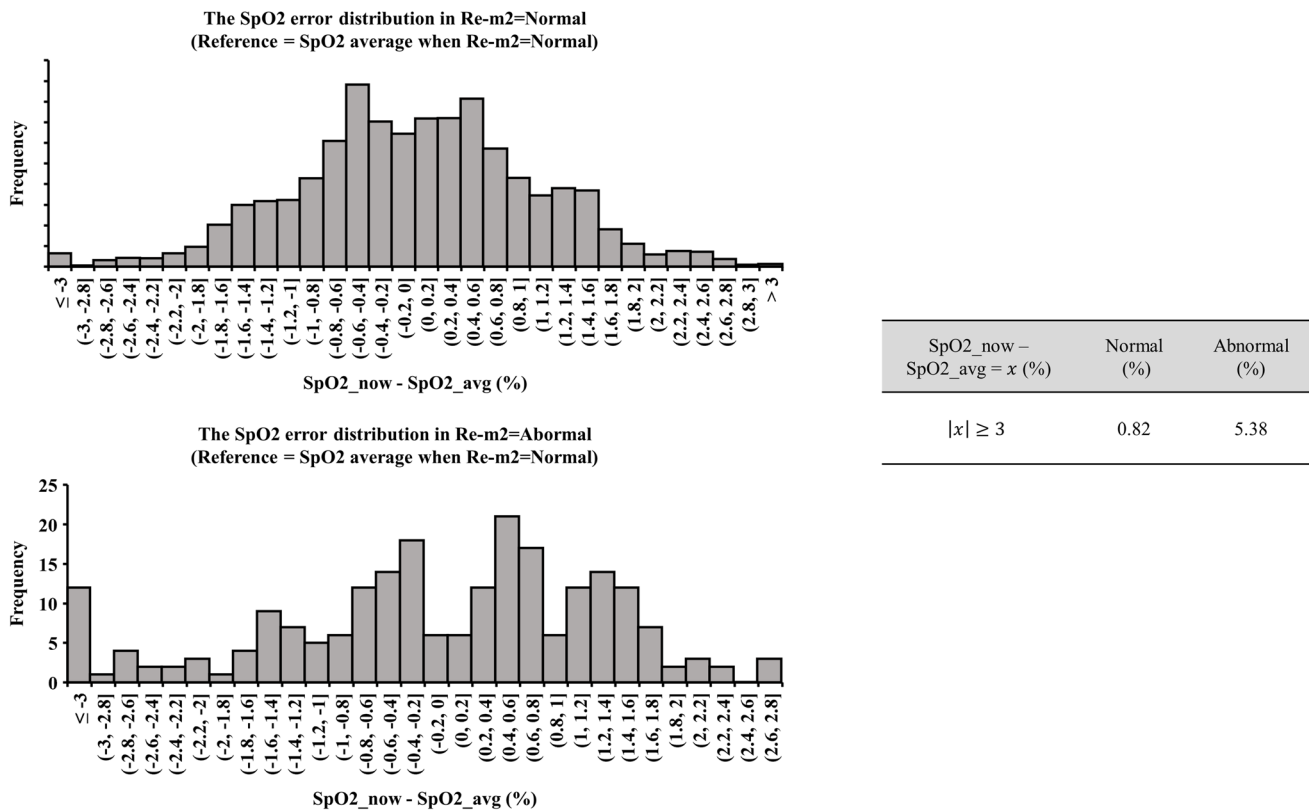


Fig. 5 Oxygen saturation (SpO2) deviation according to the normal and abnormal judgment results for Re-m2. The upper part shows result of normal judgments and the lower part shows result of abnormal judgments.

The table shows percentage of data with a deviation of 3% or more depending on the judgment

In addition, the decision relationship with each criterion and the number of abnormal data included in the data determined to be normal are represented as a Venn diagram (Fig. 7). One circle indicates the number of the data judged as normal and the number of abnormal data (number in parentheses) for each criterion. Results showed that when a normal judgment was made based on only one criterion, the abnormal data occurred in about 1% of cases (3 out of 310). However, abnormal data of about 0.2% (16 out of 9528) were included in normal judgment according to two or more criteria. Therefore, the proportion of abnormal data is lower when data are judged by more than one criterion.

4 Discussion

4.1 Reliability improvement of telemedicine data

The DNN filter-based system that can evaluate the reliability of PPG signals collected from telemedicine was proposed. This system used HR-based decisions to provide a basis for determining signal quality. In addition, through clinical

trials, reliability evaluation of telemedicine data and changes in data quality according to the criteria were verified.

HR is one of the biometric data that can be easily extracted from PPG signals [17]. However, the measurement error may vary greatly depending on the reliability of the data. Therefore, the re-measurement criterion was determined by comparing HR measured by each reference device and HR estimated by the neural network system proposed in this study. As shown in Fig. 3, the more noise is included (reducing the ratio of RS > 80), the greater the HR measurement error. It showed that an error rate of 10 bpm or more occurred despite comparison between reference devices. This may indicate that HR measurement can be a useful indicator for judging the reliability of each device.

In addition, telemedicine data that could not be directly measured or supervised by actual medical staff was used. Therefore, the reliability of the measured data was low due to abnormal measurement, operating noise, and/or measurement device errors (blood pressure greater than 500 mmHg, SpO2 measurement value less than 60, etc.) during measurement. However, this study showed that almost all abnormal data could be discriminated using the proposed system to provide evidence for judgment and that the more the criteria

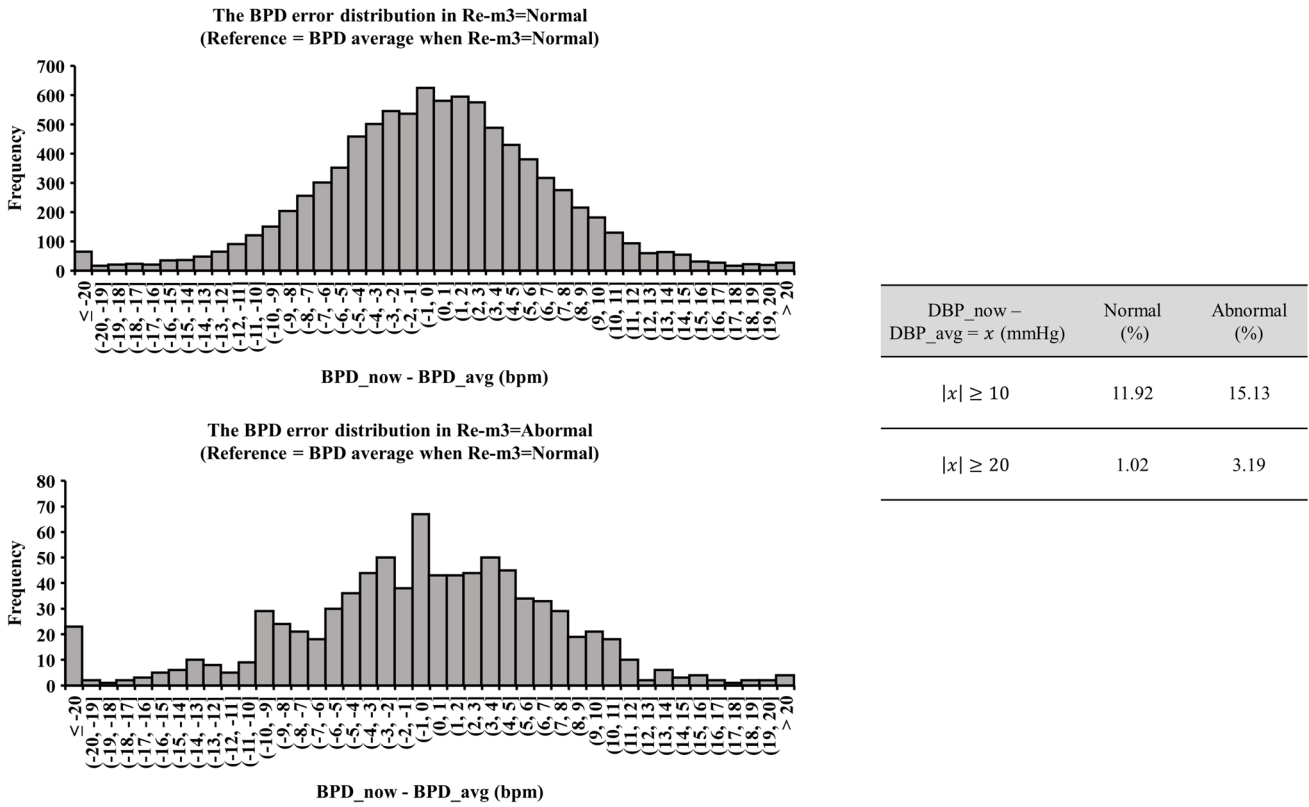
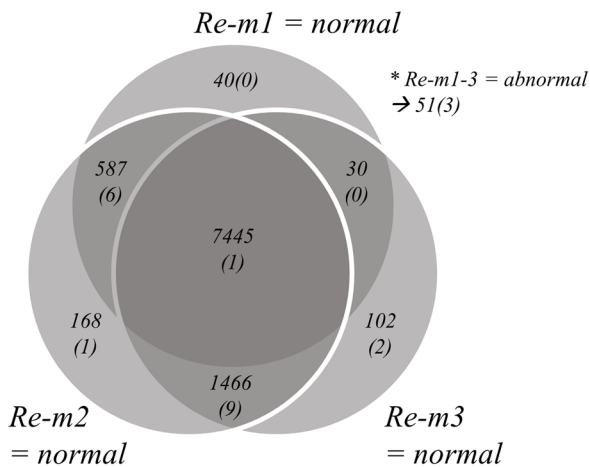


Fig. 6 Diastolic blood pressure (DBP) deviation according to normal and abnormal judgment results for Re-m3. The upper part shows result of normal judgments and the lower part shows result of abnormal

judgments. The table shows percentage of data with a deviation of 10 mmHg or more depending on the judgment



Re-m1	Re-m2	Re-m3	n	abnormal data
normal	normal	normal	7445	1
remeasurement	normal	normal	1466	9
normal	remeasurement	normal	30	0
normal	normal	remeasurement	587	6
remeasurement	remeasurement	normal	102	2
normal	remeasurement	remeasurement	40	0
remeasurement	normal	remeasurement	168	1
remeasurement	remeasurement	remeasurement	51	3

Fig. 7 Venn diagram showing the number of data determined to be normal according to each criterion. The number in parentheses indicates the number of abnormal data among the data determined to be

normal. The table shows the number of data according to the Venn diagram relationship and the number of abnormal data included

were satisfied, the higher the ratio of abnormal data could be discriminated (Fig. 7). Since this can provide a basis for determining whether the signal is abnormal, the reliability of the signal determined to be abnormal or normal can be

increased. Also, when comparing the quality of SpO2 and DBP according to each criterion, in the data determined as normal, the deviation was very small compared with the data determined as abnormal. This means that the quality of the

collected telemedicine data can be improved through the system proposed in this study (Figs. 4 and 5).

4.2 Limitation of this study

As this study was the first pilot project, re-measurement was not required more than once to reduce patient discomfort. Therefore, it was not possible to confirm the change in the re-measurement rate or data deviation according to an increase in patient's measurement skill in the long term. Patients were asked to remeasure 1 out of 5 measurements on average. However, some patients had a high remeasurement frequency. Since this can be regarded as a difference in individual measurement proficiency, it is expected that measurement proficiency can be improved by repeatedly performing re-measurement through a re-measurement or measurement error determination message over a long period. In addition, since the reliability determination system is built based on the change of a patient's waveform, it is difficult to clearly distinguish the cause of the error caused by the patient or the equipment through the reliability determination result. However, among participants, there was no case where the patient's condition or prescription was changed between before or after the measurement. There was no case where the patient recorded an abnormal situation in the questionnaire transmitted along with the data. Therefore, the fact that the deviation of the data determined to be normal was significantly reduced compared to that of abnormal data in the absence of factors that might change the measured data of the patient could justify the need for additional measurement for abnormal data.

4.3 Future works

For home-care patients with infectious diseases such as COVID-19, it is necessary to collect long-term telemedicine data rather than a one-time measurement. Therefore, in future studies, an evaluation of the stability and reliability of long-term telemedicine data is required for this system. In addition, when a patient uses a device for a long period of time, a large deviation from previously measured values might occur depending on the patient's condition. Therefore, since data deviation can reflect the patient's condition as well as the reliability of the signal, it is necessary to study data deviation according to patient's condition and proficiency in using the device.

5 Conclusion

In this study, the DNN filter-based reliability evaluation system that can evaluate the reliability of telemedicine data was proposed. Also, the criteria to provide the basis for judging abnormal data were presented. Through this system, it is

possible to determine whether the data are normal or abnormal based on an accurate judgment basis and to improve the data quality judged to be normal.

Author contributions SDA wrote this manuscript. KJ share first authorship equally and analyzed the patients' data with the new DNN algorithm. CSW conceived the original idea. LJC share corresponding authorship equally and analyzed the patients' data. All authors approved the final version of the manuscript and agree to be held accountable for the contents therein.

Funding Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2020R1A4A1019475). Korea Medical Device Development Fund grant funded by the Korea government (the Ministry of Science and ICT, the Ministry of Trade, Industry and Energy, the Ministry of Health & Welfare, the Ministry of Food and Drug Safety) (Project Number: KMDF_PR_20200901_0159-2021-02).

Declaration

Conflict of interest All authors have no conflicts of interest relevant to this study to disclose.

References

- Mann DM, Chen J, Chunara R, Testa PA, Nov O. COVID-19 transforms health care through telemedicine: evidence from the field. *J Am Med Inform Assoc.* 2020;27:1132–5.
- Antali F, Kulin D, Lucz KI, Szabó B, Szűcs L, Kulin S, et al. Multimodal assessment of the pulse rate variability analysis module of a photoplethysmography-based telemedicine system. *Sensors.* 2021;21:5544.
- Yoon G, Lee JY, Jeon KJ, Park KK, Kim HS. Development of a compact home health monitor for telemedicine. *Telemed J E Health.* 2005;11:660–7.
- Hasan S, Padman R. Analyzing the effect of data quality on the accuracy of clinical decision support systems: a computer simulation approach. *AMIA Annu Symp Proc.* 2006;324:1.
- Larburu N, Bults R, van Sinderen M, Hermens H. Quality-of-data management for telemedicine systems. *Procedia Comput Sci.* 2015;63:451–8.
- Awotunde JB, Adeniyi AE, Ogundokun RO, Ajamu GJ, Adebayo PO. MIoT-based big data analytics architecture, opportunities and challenges for enhanced telemedicine systems. *Stud Fuzzin Soft Comput.* 2021;410:199–220.
- Rai D, Thakkar HK. Cognitive big data analysis for E-health and telemedicine using metaheuristic algorithms. In: Big C, editor. Mishra S, Tripathy H, Mallick P, Sangaiah A, Chae G-S. Elsevier: Data Intelligence with a Metaheuristic Approach; 2022. p. 239–58.
- Roy MS, Gupta R, Chandra JK, Sharma KD, Talukdar A. Improving photoplethysmographic measurements under motion artifacts using artificial neural network for personal healthcare. *EEE Trans Instrum Meas.* 2018;67:2820–9.
- Lim PK, Ng S-C, Lovell NH, Yu YP, Tan MP, McCombie D, et al. Adaptive template matching of photoplethysmogram pulses to detect motion artefact. *Physiol Meas.* 2018;39: 105005.
- Roy MS, Roy B, Gupta R, Sharma KD. On-device reliability assessment and prediction of missing photoplethysmographic

- data using deep neural networks. *IEEE Trans Biomed Circuits Syst.* 2020;14:1323–32.
11. Goh C-H, Tan LK, Lovell NH, Ng S-C, Tan MP, Lim E. Robust PPG motion artifact detection using a 1-D convolution neural network. *Comput Methods Programs Biomed.* 2020;196: 105596.
 12. Prasun P, Mukhopadhyay S, Gupta R. Real-time multi-class signal quality assessment of photoplethysmography using machine learning technique. *Meas Sci Technol.* 2021;33: 015701.
 13. Kim JW, Park SM, Choi SW. Real-time photoplethysmographic heart rate measurement using deep neural network filters. *ETRI J.* 2021;43:881–90.
 14. Kim J, Park S-M, Choi S-W. Reliability of the photoplethysmographic analysis using deep neural network (dnn) algorithm. *J Biomed Eng Res.* 2021;42:31–8.
 15. Kim J, Park S, Choi S. Reliability assessment of telemedicine data by analyzing photoplethysmography with deep neural network technology. *Trans Korean Soc Mech Eng B.* 2021;1:261–9.
 16. Kim JW, Choi S-W. Normalization of photoplethysmography using deep neural networks for individual and group comparison. *Sci Rep.* 2022;12:1–10.
 17. Castaneda D, Esparza A, Ghamari M, Soltanpur C, Nazeran H. A review on wearable photoplethysmography sensors and their potential future applications in health care. *Int J Biosens Bioelectron.* 2018;4:195.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.