

다중 분기 TCN 기반 영상 호흡파형 추정 모델의 성능 평가

김 근 호*

Performance Evaluation of a Multi-Branch Temporal Convolutional Network for Video-based Respiratory Waveform Estimation

Keun Ho Kim*

이 연구는 한국한의학연구원의 기본사업과 전략연구단사업의 지원을 받았음(과제 번호: KSN2512012, SSN2611011)

요 약

카메라 기반 비접촉 호흡 측정은 센서 부착 없이 지속적인 모니터링이 가능하지만, 신체 영역별 호흡 움직임이 달라 여러 영역의 정보를 효과적으로 활용하는 것이 중요하다. 본 연구에서는 시간 합성곱 신경망(TCN), 신체 영역별 특징을 독립적으로 처리하는 다중 분기 TCN(MBTCN), 유효성 기반 융합(VAF)을 적용한 MBTCN_VAF의 성능을 비교하였다. 39명의 피험자를 대상으로 피험자 단위 교차검증을 수행하고, 예측 및 기준 호흡파형 간 피어슨 상관계수를 평가하였다. 평균 상관계수는 TCN 0.744, MBTCN 0.788, MBTCN_VAF 0.795였다. MBTCN은 TCN보다 평균 상관계수가 0.044 증가하고 35명(89.7%)에서 성능이 향상되었으며 ($p<0.001$), VAF 적용 시에도 0.007 추가 증가하여 통계적으로 유의하였다($p<0.001$). 이러한 결과는 다중 분기 구조와 VAF를 통한 적응적 특징 결합이 호흡파형 추정 성능 향상에 기여함을 보여준다.

Abstract

Camera-based contactless methods enable continuous respiratory monitoring without sensor attachment; however, respiratory motion varies across body regions, requiring effective use of information from multiple regions. This study compared a temporal convolutional network (TCN), a multi-branch TCN (MBTCN), and an MBTCN with validity-aware fusion (MBTCN_VAF). Subject-level cross-validation was performed using data from 39 subjects, and performance was evaluated using the Pearson correlation coefficient between predicted and reference respiratory waveforms. Mean correlations were 0.744 for TCN, 0.788 for MBTCN, and 0.795 for MBTCN_VAF. MBTCN increased the mean correlation by 0.044 over TCN and improved performance in 35 subjects (89.7%; $p<0.001$), while VAF further increased the mean correlation by 0.007 ($p<0.001$). These results demonstrate that the multi-branch architecture and validity-aware fusion improve respiratory waveform estimation performance.

Keywords

contactless respiratory monitoring, respiratory waveform estimation, temporal convolutional network, multi-branch network, leave-one-subject-out cross-validation, deep learning

* 한국한의학연구원 디지털임상연구부
- ORCID: <https://orcid.org/0000-0001-6688-6674>

• Received: Aug. 12, 2026, Revised: Aug. 27, 2026, Accepted: Sep. 02, 2026
• Corresponding Author: Keun Ho Kim
Digital Health Research Division
Korea Institute of Oriental Medicine
Tel.: +82-42-868-9365, Email: rkim70@kiom.re.kr

1. 서 론

호흡수(RR, Respiratory Rate)와 호흡파형은 인체의 호흡기능과 생리적 상태 변화를 평가하는 중요한 생체신호이다[1][2]. 특히 고령자는 노화에 따른 생리적 예비능 감소와 만성질환의 증가로 인해 건강상태 변화에 대한 지속적인 관찰이 중요하다[2]. 폐렴, 만성 폐쇄성폐질환(COPD, Chronic Obstructive Pulmonary Disease), 심부전 등에서는 호흡수 증가나 호흡 양상의 변화가 환자의 상태 악화를 반영할 수 있으며, RR은 임상적 악화를 조기에 나타낼 수 있는 주요 활력징후 중 하나로 보고되고 있다[1][3]. 따라서 고령자의 건강관리와 질환 악화의 조기 인지를 위해 지속적인 호흡 상태 모니터링이 중요하다.

기존의 호흡 측정에는 흉부 또는 복부에 착용하는 호흡 벨트, 임피던스 기반 센서, 기류 센서 등 신체에 직접 접촉하는 방식이 주로 사용되어 왔다[4][5]. 이러한 방법은 비교적 안정적으로 호흡신호를 측정할 수 있으나 센서를 신체에 직접 부착하거나 착용해야 하므로 장시간 측정 시 불편을 유발할 수 있다[4][5]. 따라서 별도의 센서를 신체에 부착하지 않고 일상적인 환경에서 호흡 상태를 지속적으로 측정할 수 있는 비접촉 방식에 대한 관심이 증가하고 있으며[4], UWB 레이다 등을 이용한 비접촉 호흡 모니터링 시스템도 연구되고 있다[6].

카메라를 이용한 비접촉 호흡 측정은 호흡에 따라 발생하는 흉부 및 신체의 미세한 움직임을 영상으로 획득하고 이를 분석하여 호흡파형 또는 호흡수를 추정하는 방법이다[7][8]. 일반적인 RGB 카메라와 광류(Optical flow)를 이용하여 호흡 관련 움직임을 추출하는 방법도 보고되어 왔다[7][8]. 또한 RGB 카메라 영상과 CNN을 이용하여 호흡 움직임을 분석하는 등 딥러닝 기반의 비접촉 호흡 측정 연구도 수행되고 있다[9]. 최근에는 컴퓨터 비전과 딥러닝 기술을 이용한 비접촉 생체신호 추정 연구가 활발히 이루어지고 있다[10].

그러나 영상에서 나타나는 호흡 관련 움직임은 신체의 모든 영역에서 동일하게 나타나지 않는다. 기존 연구에서는 흉복부, 얼굴, 어깨 주변 및 목 등 다양한 신체 영역의 움직임을 이용하여 호흡신호를

추출해 왔다[4][8]. 각 영역에서 관찰되는 움직임의 크기와 방향, 신호대잡음비 및 비호흡성 움직임의 영향은 서로 다를 수 있다. 따라서 여러 영역에서 추출한 특징을 초기 단계에서 하나의 특징 표현으로 단순하게 결합할 경우, 각 영역에 포함된 고유한 시간적 특성이나 상보적인 호흡 정보가 충분히 활용되지 못할 가능성이 있다. 이러한 특성은 새로운 피험자에게 모델을 적용할 때 성능 저하의 원인이 될 수 있으므로, 비접촉 호흡 측정에서는 다양한 호흡 관련 특징을 효과적으로 처리하면서 피험자 간 차이에 강한 모델을 구축하는 것이 중요하다.

시간 합성곱 신경망(TCN, Temporal Convolutional Network)은 합성곱(Convolution) 연산을 기반으로 시계열 데이터의 시간적 의존성을 학습할 수 있으며, 팽창 합성곱(Dilated convolution)을 이용하여 넓은 시간 범위의 정보를 효율적으로 처리할 수 있다[11][12]. 이에 따라 TCN은 호흡과 같이 시간적 주기성을 갖는 생체신호의 파형 추정을 위한 시계열 모델로 적용할 수 있다. 그러나 단일 TCN 구조에서 특성이 서로 다른 여러 입력 신호를 함께 처리하는 경우, 각각의 입력에 포함된 고유한 시간적 특징을 독립적으로 학습하는 데 한계가 있을 수 있다.

본 연구에서는 이러한 문제를 개선하기 위하여 서로 다른 호흡 관련 특징을 각각 독립적인 시간적 분기(Temporal branch)에서 처리한 후 통합하는 다중 분기 시간 합성곱 신경망(MBTCN, Multi-Branch Temporal Convolutional Network)을 구성하였다. 여러 ROI 특징을 입력 단계에서 함께 처리하는 기본 TCN과 달리, MBTCN은 얼굴, 어깨 및 가슴에서 추출된 특징을 각각 독립적인 시간적 분기에서 처리함으로써 ROI별 고유한 시간적 특성을 유지하면서 상보적인 호흡 정보를 활용하도록 설계하였다. 또한 MBTCN에 유효성 기반 융합(VAF, Validity-Aware Fusion)을 적용한 MBTCN_VAF를 구성하여, ROI의 유효성과 각 분기의 상대적인 기여도를 반영한 적응적 가중 결합이 가능하도록 하였다. 이를 통해 다중 분기 구조의 효과와 VAF의 추가적인 효과를 단계적으로 평가하고자 하였다.

본 연구에서는 기본 TCN, MBTCN 및 MBTCN_VAF의 세 모델을 동일한 조건에서 비교하

고, 39명의 피험자를 대상으로 피험자 단위 교차검증(LOSO, Leave-One-Subject-Out)을 수행하여 새로운 피험자에 대한 호흡파형 추정 성능을 평가하였다. 이를 통해 다중 분기 구조가 피험자 독립적 호흡파형 추정 성능에 미치는 영향을 분석하고, VAF 모듈의 추가적인 기여도를 정량적으로 평가하고자 하였다.

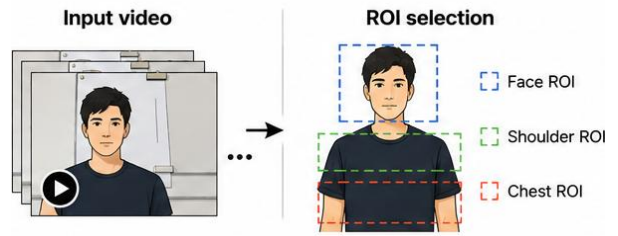
II. 연구 방법

2.1 데이터 및 호흡신호

본 연구에서는 공개 생체신호 데이터베이스인 COHFACE를 이용하였다[13]. COHFACE는 비접촉 생체신호 추정 연구에 활용되는 데이터셋으로, 그림 1(a)와 같이 40명의 피험자를 대상으로 각각 4회의 약 1분 길이 영상을 획득하여 총 160개의 RGB 영상으로 구성되어 있다. 영상은 Logitech HD Webcam C525를 이용하여 640×480 픽셀의 해상도와 초당 20 프레임으로 획득되었다. 영상 촬영과 동시에 접촉식 센서를 이용하여 혈액용적맥파(Blood volume pulse)와 호흡신호를 측정하였으며, 생리신호와 영상은 시간적으로 동기화하여 기록되었다[13].

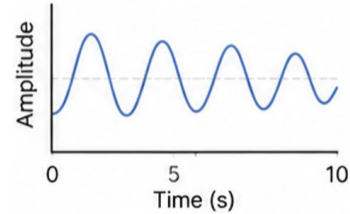
호흡신호는 Thought Technologies의 호흡 벨트(Respiration belt, SA9311M)를 이용하여 흉곽의 신장(Thoracic stretch)을 측정하는 방식으로 획득되었으며, 원 데이터의 호흡신호 샘플링 주파수(Sampling rate)는 32 Hz이다[13]. 본 연구에서는 그림 1(b)와 같이 이 호흡 벨트 신호를 영상 기반 호흡파형 추정 모델의 기준 호흡파형(Ground-truth respiratory waveform)으로 사용하였다.

모델링에 앞서 영상과 이에 대응하는 기준 호흡파형의 데이터 품질을 검토하였다. 전체 40명의 피험자 중 피험자 28에서는 영상의 흔들림과 함께 기준 호흡파형에서 안정적인 주기적 호흡 패턴을 확인하기 어려운 데이터 품질 문제가 관찰되었다. 이러한 데이터는 영상 기반 호흡파형 추정 모델의 학습 및 평가에 영향을 미칠 수 있다고 판단하여 피험자 28을 모델 학습 및 성능 평가 이전에 사전 제외하였다. 따라서 최종적으로 39명의 피험자 데이터를 모델 개발 및 평가에 사용하였으며, 이들을 대상으로 LOSO 교차검증을 수행하였다.



(a) 획득된 영상 및 관심영역

(a) Acquired image and regions of interest (ROIs)



(b) 기준 호흡파형

(b) Ground-truth respiratory waveform

그림 1. 영상 기반 호흡신호 추정을 위한 관심영역 설정 및 기준 호흡파형의 예

Fig. 1. ROI selection for video-based respiratory waveform estimation and an example of the ground-truth respiratory waveform

2.2 영상 기반 호흡 특징 추출 및 전처리

COHFACE 영상으로부터 호흡 관련 움직임을 분석하기 위해 그림 1(a)와 같이 얼굴, 어깨 및 가슴 영역을 관심영역(ROI, Region of Interest)으로 설정하였다. 각 ROI의 위치는 MediaPipe Pose[14]를 이용하여 추정하였으며, 얼굴 영역은 코와 양쪽 귀 특징점(Landmark)을 기준으로, 어깨와 가슴 영역은 양쪽 어깨 특징점의 위치와 어깨 폭을 기준으로 정의하였다. 각 ROI는 회색조(Grayscale) 영상으로 변환한 후 160×160 픽셀 크기로 정규화하였다.

연속 프레임 사이의 움직임은 Farneback 광류[15]를 이용하여 계산하였으며, 각 ROI에서 광류의 수직 방향 성분의 평균값을 호흡 움직임 특징으로 사용하였다. 얼굴, 어깨 및 가슴에서 추출된 세 신호에는 0.1 - 0.7 Hz의 2차 Butterworth 대역통과 필터(Band-pass filter)를 적용하고 평균 제거 및 표준편차 기반 정규화를 수행하였다. 최종 입력 특징은 정규화된 얼굴, 어깨 및 가슴의 수직 방향 광류 신호와 가슴 ROI의 유효성을 나타내는 이진 마스크(Binary mask)로 구성하였다. 가슴 ROI의 유효성은 추정된

ROI가 영상 내에서 유효한 영역을 형성하는지에 따라 이진값으로 정의하였으며, 유효한 경우 1, 유효하지 않은 경우 0으로 설정하였다.

모델의 입력 구간은 20 Hz를 기준으로 30초(600 샘플)로 설정하였으며, 5초(100 샘플) 간격으로 이동하면서 입력 구간을 생성하였다.

2.3 TCN 기반 호흡파형 추정 모델

기본 비교 모델로 그림 2의 TCN을 사용하였다 [11][12]. TCN의 입력은 얼굴, 어깨 및 가슴의 수직 방향 광류 신호와 가슴 유효성 마스크를 포함하는 4차원 시계열 특징으로 구성하였다.

모델은 입력 특징을 시간축 합성곱으로 처리하도록 구성하였다. 첫 번째 Conv1D 계층은 커널 크기 (Kernel size) 7을 이용하여 4개의 입력 채널을 64개의 특징 채널(Feature channel)로 변환하였다. 이후 커널 크기 7과 팽창률(Dilation rate) 1 및 2를 갖는 두 개의 Conv1D 계층을 순차적으로 적용하였으며, 각 합성곱 뒤에는 ReLU 활성화 함수를 적용하였다. 마지막으로 커널 크기 1의 Conv1D를 이용하여 64차원 시간적 특징(Temporal feature)을 단일 채널의

연속적인 호흡파형으로 변환하였다.

2.4 ROI별 다중 분기 TCN

단일 TCN에서는 서로 다른 신체 영역에서 획득한 신호가 입력 단계에서 함께 처리되므로 각 ROI가 갖는 고유한 시간적 움직임 특성이 충분히 분리되어 학습되지 않을 가능성이 있다. 이를 개선하기 위해 얼굴, 어깨 및 가슴 신호를 독립적으로 처리하는 그림 3과 같은 ROI별 MBTCN을 구성하였다.

MBTCN은 얼굴, 어깨 및 가슴 수직 방향 광류 신호에 각각 독립적인 시간적 분기를 적용하였다. 각 분기는 커널 크기 7과 팽창률 1, 2 및 4를 갖는 세 개의 Conv1D 계층으로 구성하였다. 각 계층은 32개의 특징 채널을 출력하며 ReLU 활성화 함수를 적용하였다.

세 분기에서 생성된 특징맵(Feature map)을 채널 방향으로 연결(Concatenation)하여 96채널 특징을 형성하였다. 이후 1×1 Conv를 이용하여 96개 채널을 32개 채널로 축소하고, 커널 크기 3의 Conv1D와 최종 1×1 Conv를 순차적으로 적용하여 단일 호흡파형을 출력하였다.

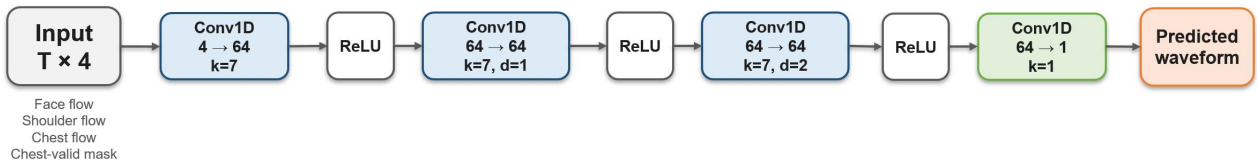


그림 2. TCN 기반 호흡파형 추정 모델
Fig. 2. TCN-based respiratory waveform estimation

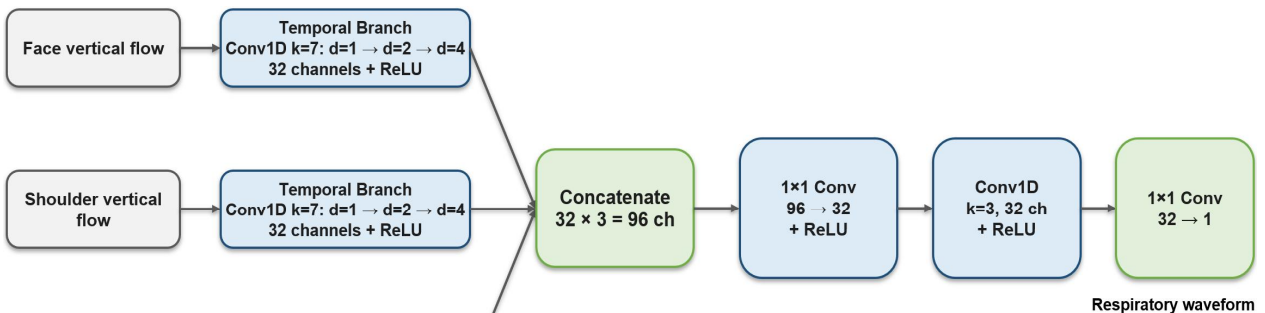


그림 3. 다중 분기 TCN 기반 호흡파형 추정 모델
Fig. 3. ROI-specific multi-branch TCN

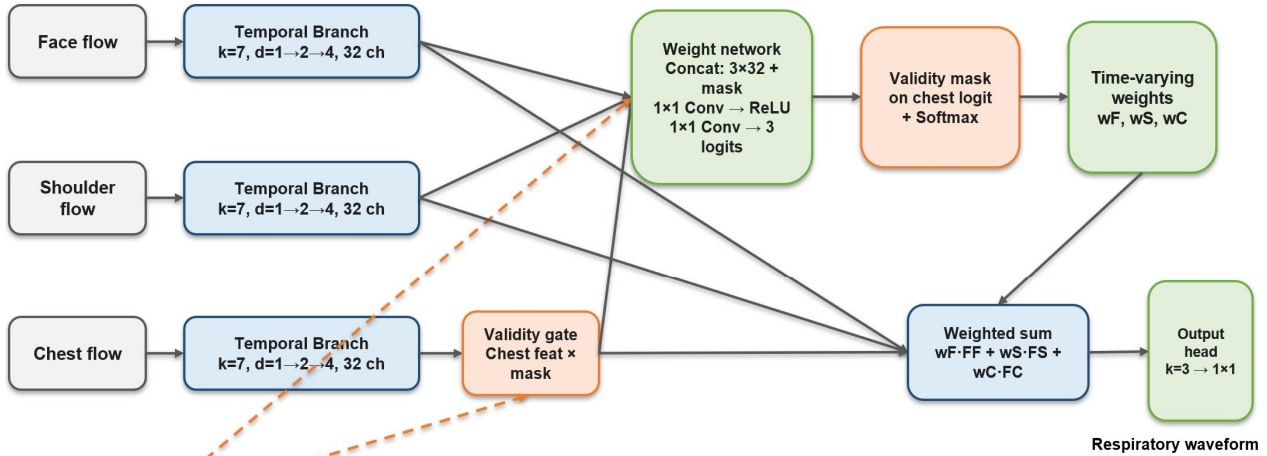


그림 4. VAF를 적용한 다중 분기 TCN 기반 호흡파형 추정 모델
Fig. 4. Multi-branch TCN with validity-aware fusion

이 모델에서는 데이터에 포함된 가슴 유효성 마스크를 사용하지 않았으며, 각 ROI의 유효성을 고려한 적응적 가중치도 적용하지 않았다. 따라서 기본 TCN과의 비교를 통해 ROI별 독립적인 시간적 특징 추출을 수행하는 다중 분기 구조 자체의 효과를 평가하였다.

2.5 VAF를 적용한 다중 분기 TCN

MBTCN에서 추출된 ROI별 시간적 특징을 데이터의 유효성에 따라 적응적으로 통합하기 위해 VAF를 적용한 그림 4와 같은 MBTCN_VAF 모델을 구성하였다.

얼굴, 어깨 및 가슴 신호는 MBTCN과 동일하게 각각 독립적인 시간적 분기를 통과하며, 각 분기는 팽창률 1, 2 및 4의 Conv1D 계층을 이용하여 32채널의 시간적 특징을 생성한다.

VAF에서는 가슴 유효성 마스크를 명시적으로 이용하였다. 가슴 ROI가 유효하지 않은 경우 가슴 특징에 유효성 마스크를 곱하여 해당 특징의 영향을 제거하였다. 이후 세 ROI 특징과 가슴 유효성 마스크를 결합하여 가중치 신경망(Weight network)에 입력하였다. 가중치 신경망은 1×1 Conv, ReLU 및 세 개의 가중치 로짓(Weight logit)을 출력하는 1×1 Conv로 구성하였다.

각 시점에서 얼굴, 어깨 및 가슴에 대한 가중치는 softmax를 통해 계산하였다. 특히 가슴 유효성이 0인 경우 가슴 분기의 로짓을 마스킹(Masking)하여 softmax 기반 특징 융합(Feature fusion)에서 해당 분기가 제외되도록 하였다. 최종 융합 특징(Fused feature)은 식 (1)과 같이 세 ROI 특징의 가중합으로 구성하였다.

$$F_{fused}(t) = w_F(t)F_F(t) + w_S(t)F_S(t) + w_C(t)F_C(t) \quad (1)$$

여기서 $F_F(t)$, $F_S(t)$, $F_C(t)$ 는 각각 시점 t 에서 얼굴, 어깨 및 가슴 분기로부터 추출된 시간적 특징을 의미하며, $w_F(t)$, $w_S(t)$, $w_C(t)$ 는 각 특징에 대응하여 가중치 신경망과 softmax를 통해 산출된 가중치를 의미한다. 각 시점에서 세 가중치의 합은 1이 되며, $F_{fused}(t)$ 는 이들 ROI별 시간적 특징의 가중합으로 생성된 최종 융합 특징을 나타낸다.

이후 커널 크기 3의 Conv1D와 1×1 Conv로 구성된 출력부(Output head)를 이용하여 최종 호흡파형을 추정하였다.

따라서 세 모델은 TCN에서 ROI별 다중 분기 구조를 적용한 MBTCN과 VAF를 적용한 MBTCN_VAF로 단계적으로 확장하였으며, 이를 통해 다중 분기 구조의 효과와 VAF의 추가적인 효과

를 각각 평가하였다. 그림 2-4에 제시된 TCN, MBTCN 및 MBTCN_VAF의 구조적 차이를 보다 명확하게 비교하기 위해 각 모델의 주요 특징, 장점 및 제한점을 표 1에 정리하였다.

표 1. TCN, MBTCN 및 MBTCN_VAF 모델의 주요 특징, 장점 및 제한점

Table 1. Key characteristics, advantages, and limitations of the TCN, MBTCN, and MBTCN_VAF models

Model	Category	Structural characteristics and comparison
TCN	Key feature	Processes features from multiple ROIs within a single temporal network
	Advantage	Simple architecture that simultaneously processes features from multiple ROIs
	Limitation	Limited ability to independently learn ROI-specific temporal characteristics
MBTCN	Key feature	Processes features from the face, shoulder, and chest independently through separate temporal branches and subsequently combines them
	Advantage	Enables independent learning of ROI-specific temporal characteristics and complementary information
	Limitation	Does not apply adaptive weighting based on ROI validity
MBTCN_VAF	Key feature	Applies VAF to MBTCN to adaptively weight and fuse ROI features
	Advantage	Accounts for ROI validity and dynamically adjusts the contribution of each ROI
	Limitation	More complex than MBTCN, and the effectiveness of VAF may depend on variations in input signal quality and validity

2.6 모델 학습 및 평가

피험자 28은 사전 데이터 품질 검토 결과 분석에서 제외하였으며, 나머지 39명의 피험자를 대상으로 피험자 단위 교차검증을 이용한 피험자 독립적 평가(Subject-independent evaluation)를 수행하였다. 각 LOSO 교차검증에서는 한 명의 피험자를 시험 피험자(Test subject)로 설정하고, 나머지 피험자 가운데 4명을 검증 피험자(Validation subject)로 무작위 선택하였다. 이 과정은 난수 시드(Random seed) 2026과

교차검증 번호를 이용하여 재현 가능하도록 하였다.

학습에는 배치 크기(Batch size) 6, Adam 최적화 알고리즘(Adam optimizer), 학습률(Learning rate) 1×10^{-3} 및 평균제곱오차(MSE, Mean Squared Error) 손실함수를 사용하였다. 각 LOSO 교차검증에서 최대 30 에포크(Epoch) 동안 학습하였으며, 검증 데이터의 호흡파형 상관계수가 연속 5 에포크 동안 향상되지 않는 경우 학습을 조기 종료하였다. 시험 데이터(Test set)가 아닌 검증 데이터(Validation set)에서 예측 호흡파형과 기준 호흡파형 간 피어슨 상관계수(Pearson correlation coefficient)가 가장 높은 에포크의 모델을 최적 모델로 선택하였다.

선택된 최적 모델을 복원한 후 해당 LOSO 교차검증의 시험 피험자에 대해서는 최종 성능을 한 번만 평가하였다. 따라서 시험 피험자의 결과는 에포크 또는 모델 선택에 사용되지 않도록 구성하였다.

이러한 절차를 39명의 모든 피험자에 대해 반복함으로써 모델 선택 과정에서 시험 피험자의 정보가 사용되지 않도록 하였다.

2.7 성능 지표 및 통계 분석

세 모델의 호흡파형 추정 성능은 예측 호흡파형과 기준 호흡파형 간의 피어슨 상관계수(r)를 이용하여 평가하였다. 두 파형 사이의 시간적 지연을 고려하기 위해 각 입력 구간에서 ± 2 초 범위의 시간 지연(Temporal lag)에 대해 상관계수를 계산하고, 최대 상관계수를 해당 입력 구간의 성능으로 사용하였다. 각 LOSO 교차검증에서는 시험 피험자에 포함된 입력 구간들의 상관계수를 평균하여 피험자별 최종 상관계수를 산출하였으며, 39명 피험자의 평균과 표준편차를 모델별 전체 성능으로 제시하였다.

동일한 피험자에 대해 TCN, MBTCN 및 MBTCN_VAF의 세 모델이 반복적으로 평가되었으므로, 모델 간 전체 성능 차이는 Friedman 검정을 이용하여 분석하였다. 유의한 전체 차이가 확인된 경우 Wilcoxon 부호순위 검정(Wilcoxon signed-rank test)을 이용하여 모델 간 쌍별 비교를 수행하였으며, 다중 비교에 대해서는 Holm 보정을 적용하였다. 추가적으로 TCN에서 MBTCN으로, 그리고 MBTCN에

서 MBTCN_VAF로 모델을 확장하였을 때 피험자별 상관계수가 증가한 비율을 산출하여 각 구조의 추가에 따른 성능 변화의 일관성을 보조적으로 평가하였다. 통계적 유의수준은 $p < 0.05$ 로 설정하였으며, 모델 간 쌍별 비교에서는 Holm 보정된 유의확률(p-value)을 기준으로 통계적 유의성을 판단하였다.

III. 결 과

3.1 모델별 호흡파형 추정 성능

피험자 28을 제외한 39명의 피험자를 대상으로 피험자 단위 교차검증을 수행하였다. 각 LOSO 교차검증에서는 검증 데이터의 호흡파형 상관계수를 기준으로 최적 모델을 선택한 후, 해당 교차검증의 시험 피험자에 대해 최종 성능을 한 번만 평가하였다.

그 결과, 표 2에서 보듯이 기본 TCN 모델의 평균 피어슨 상관계수는 0.744 ± 0.110 이었으며, ROI 별 시간적 특징을 독립적으로 처리하는 MBTCN에서는 0.788 ± 0.102 로 증가하였다. 여기에 VAF를 적용한 MBTCN_VAF의 평균 상관계수는 0.795 ± 0.102 로 가장 높은 값을 나타냈다.

표 2. 모델별 호흡파형 추정 성능 및 최적 에포크
Table 2. Respiratory waveform estimation performance and optimal training epochs for each model

Item	TCN	MBTCN	MBTCN_VAF
Pearson r, mean $\pm$ SD	0.744 ± 0.110	0.788 ± 0.102	0.795 ± 0.102
Median	0.776	0.809	0.823
Range	0.325 - 0.865	0.425 - 0.905	0.413 - 0.904
Mean optimal epoch	13.69	12.77	15.33
Max. optimal epoch	27	20	28

각 LOSO 교차검증에서 검증 데이터의 상관계수를 기준으로 선택된 최적 에포크의 평균은 TCN, MBTCN 및 MBTCN_VAF에서 각각 13.69, 12.77 및 15.33이었으며, 최대 최적 에포크는 각각 27, 20 및 28이었다.

TCN에서 MBTCN으로 구조를 확장하였을 때 평균 상관계수는 약 0.044 증가하였다. 또한 MBTCN에 VAF를 추가하였을 때 평균 상관계수는 약 0.007 추가 증가하였다.

3.2 다중 분기 구조에 따른 피험자별 성능 변화

다중 분기 구조의 효과를 보다 구체적으로 확인하기 위해 동일한 피험자에서 TCN과 MBTCN의 상관계수를 비교하였다. 그 결과, 39명 중 35명의 피험자에서 MBTCN의 상관계수가 TCN보다 증가하였다.

TCN에서 MBTCN으로 변경하였을 때 피험자별 상관계수의 평균 증가량은 +0.0441이었으며, 전체 피험자의 약 89.7%(35/39)에서 성능 향상이 관찰되었다.

3.3 VAF 적용에 따른 피험자별 성능 변화

VAF의 추가적인 효과를 확인하기 위해 MBTCN과 MBTCN_VAF의 피험자별 성능을 비교하였다. MBTCN의 평균 상관계수는 0.788이었으며, VAF 적용 후 0.795로 증가하여 평균적으로 +0.0071의 추가적인 향상을 보였다.

피험자별로 살펴보면 MBTCN에서 MBTCN_VAF로 변경하였을 때 39명 중 29명(74.4%)에서 상관계수가 증가하였다. 이는 VAF 적용에 따른 성능 향상이 모든 피험자에서 나타난 것은 아니지만 다수의 피험자에서 추가적인 성능 향상이 관찰되었음을 보여준다.

3.4 모델 간 통계적 성능 비교

세 모델에서 얻어진 39명의 피험자별 상관계수에 대해 Friedman 검정을 수행한 결과, 식 (2)와 같이 모델 간 성능에는 통계적으로 유의한 차이가 나타났다.

$$\chi^2(2) = 50.67, p = 9.95 \times 10^{-12} \quad (2)$$

여기서 $\chi^2(2)$ 는 자유도 2에서의 Friedman 검정 통계량을 의미하며, p는 해당 검정 통계량에 대한 유

의확률을 나타낸다. 식 (2)의 유의확률이 유의수준 0.05보다 작으므로 세 모델의 호흡파형 추정 성능이 동일하다는 귀무가설을 기각할 수 있으며, 이는 세 모델 중 적어도 하나의 성능이 다른 모델과 통계적으로 유의한 차이가 있음을 의미한다.

이에 따라 구체적인 모델 간 차이를 확인하기 위해 Wilcoxon 부호순위 검정을 이용하여 모델 간 쌍별 비교를 수행하였다. TCN과 MBTCN 간에는 통계적으로 유의한 차이가 나타났으며($p=3.13 \times 10^{-10}$), TCN과 MBTCN_VAF 간에도 유의한 차이가 확인되었다($p=2.18 \times 10^{-11}$). 또한, MBTCN과 MBTCN_VAF 간에도 통계적으로 유의한 차이가 나타났다($p=8.91 \times 10^{-4}$).

따라서 통계분석 결과에서도 단일 TCN에서 MBTCN으로의 다중 분기 구조 확장이 호흡파형 추정 성능을 유의하게 향상시켰으며, MBTCN에 VAF를 적용한 경우에도 추가적인 성능 향상이 나타났다.

IV. 고 찰

본 연구에서는 영상으로부터 비접촉 방식으로 호흡파형을 추정하기 위해 TCN 기반 모델을 구성하고, ROI별 시간적 특징을 독립적으로 학습하는 다중 분기 구조와 VAF의 효과를 비교하였다. 피험자 28을 사전에 제외한 COHFACE 39명의 데이터를 대상으로 피험자 단위 LOSO 평가를 수행한 결과, MBTCN은 39명 중 35명(89.7%)에서 기본 TCN보다 높은 상관계수를 나타냈으며, 평균 상관계수도 0.744에서 0.788로 증가하였다. 또한, VAF를 적용한 MBTCN_VAF의 평균 상관계수는 0.795로 증가하였으며, MBTCN과의 차이는 통계적으로 유의하였다. 이러한 결과는 얼굴, 어깨 및 가슴 영역에서 나타나는 호흡 관련 움직임을 하나의 입력으로 일괄 처리하기보다 각 ROI의 시간적 특징을 독립적으로 학습하는 것이 호흡파형 추정에 효과적이며, ROI의 유효성을 고려한 VAF의 적용이 추가적인 성능 향상에 기여할 수 있음을 보여준다.

특히 호흡에 따른 움직임은 신체 영역에 따라 서로 다른 특성을 나타낼 수 있다. 기존의 카메라 기반 비접촉 호흡 측정 연구에서는 흉부 및 흉복부

영역에서 나타나는 호흡 관련 움직임을 영상으로 측정하여 호흡 패턴과 호흡수를 추정할 수 있음이 보고되었다[7][8]. 또한 호흡에 따른 신체 움직임은 측정 영역에 따라 그 크기와 형태가 다르게 관찰될 수 있으므로, 단일 영역의 정보만을 이용하기보다 여러 신체 영역에서 얻어진 정보를 상호 보완적으로 활용할 필요가 있다. 본 연구의 MBTCN은 얼굴, 어깨 및 가슴 ROI에서 추출된 특징을 각각 독립적인 시간적 분기에서 처리한 후 통합함으로써, ROI별 시간적 특성을 유지하면서 상보적인 호흡 정보를 활용하도록 설계되었다. 이러한 구조가 단일 TCN 대비 호흡파형 추정 성능 향상에 기여한 것으로 판단된다.

VAF를 적용한 MBTCN_VAF는 가슴 ROI의 유효성을 고려하고 각 분기의 기여도를 시간에 따라 적응적으로 조절하도록 설계하였다. VAF 적용 후 평균 상관계수는 MBTCN의 0.788에서 0.795로 증가하였으며, 두 모델 간 차이는 통계적으로 유의하였다($p<0.001$). 다만 VAF 적용에 따른 평균 상관계수 증가량(+0.0071)은 TCN에서 MBTCN으로 확장했을 때의 증가량(+0.0441)보다 작았다. 이러한 결과는 VAF가 ROI별 특징의 적응적 결합을 통해 추가적인 성능 향상에 기여할 수 있음을 보여준다. 향후에는 신체 움직임, 부분적인 ROI 소실, 조명 변화 등 다양한 영상 품질 저하 조건에서 VAF의 효과를 추가적으로 평가할 필요가 있다.

본 연구에는 몇 가지 제한점이 있다. 첫째, COHFACE는 40명의 비교적 제한된 대상자로 구성되어 있으며 본 연구에서는 데이터 품질 문제로 피험자 28을 사전 제외하여 39명을 분석하였다. 둘째, 단일 데이터셋을 이용하였으므로 다른 촬영환경과 대상자에 대한 일반화 성능을 추가로 검증할 필요가 있다. 셋째, 본 연구에서는 예측 파형과 기준 호흡파형 간 상관계수를 주요 성능지표로 사용하였으므로, 향후 연구에서는 호흡수 오차와 파형 형태 오차 등 다양한 평가 지표를 함께 적용할 필요가 있다. 또한 실제 고령자를 대상으로 한 데이터가 아니므로 고령자 원격 호흡 모니터링으로의 적용 가능성을 확인하기 위해서는 고령자를 포함한 독립적인 데이터에서 추가적인 검증이 필요하다.

V. 결 론

본 연구에서는 비접촉 영상 기반 호흡과형 추정을 위해 TCN, MBTCN 및 MBTCN_VAF의 성능을 비교하였다. 피험자 단위 LOSO 교차검증 평가 결과, 평균 피어슨 상관계수는 TCN에서 0.744 ± 0.110 , MBTCN에서 0.788 ± 0.102 , MBTCN_VAF에서 0.795 ± 0.102 로 나타났다. MBTCN은 39명 중 35명(89.7%)에서 TCN보다 높은 성능을 보였으며, TCN 대비 평균 상관계수가 약 0.044 증가하여 통계적으로 유의한 성능 향상을 나타냈다($p < 0.001$). 또한 VAF를 적용한 MBTCN_VAF는 MBTCN보다 평균 상관계수가 약 0.007 증가하였으며, 두 모델 간 차이도 통계적으로 유의하였다($p < 0.001$). 이러한 결과는 영상 기반 호흡과형 추정에서 ROI별 시간적 특징을 독립적으로 학습하고 통합하는 다중 분기 구조가 호흡과형 추정 성능 향상에 기여하며, VAF를 통한 ROI별 특징의 적응적 결합도 추가적인 성능 향상에 기여할 수 있음을 보여준다. 향후 다양한 촬영환경과 고령자를 포함한 독립적인 데이터에서의 검증을 통해 비접촉 원격 호흡 모니터링으로의 활용 가능성을 평가할 필요가 있다.

감사의 글

본 연구의 일부는 스위스 마르티니(Martigny)의 Iidiap Research Institute에서 제공한 COHFACE 데이터셋을 활용하여 수행되었다.

References

- [1] S. A. S. Aglen, H. F. Simonsen, T. E. Sjøset, and I. Jammer, "Respiratory rate as a predictor of clinical deterioration and mortality: A scoping review", *Acta Anaesthesiologica Scandinavica*, Vol. 69, No. 8, Art. no. e70113, Sep. 2025. <https://doi.org/10.1111/aas.70113>.
- [2] J. G. Chester and J. L. Rudolph, "Vital signs in older patients: Age-related changes", *Journal of the American Medical Directors Association*, Vol. 12, No. 5, pp. 337-343, Jun. 2011. <https://doi.org/10.1016/j.jamda.2010.04.009>.
- [3] K. Mochizuki, R. Shintani, K. Mori, T. Sato, O. Sakaguchi, K. Takeshige, K. Nitta, and H. Imamura, "Importance of respiratory rate for the prediction of clinical deterioration after emergency department discharge: A single-center, case-control study", *Acute Medicine & Surgery*, Vol. 4, No. 2, pp. 172-178, Apr. 2017. <https://doi.org/10.1002/ams2.252>.
- [4] F. Q. AL-Khalidi, R. Saatchi, D. Burke, H. Elphick, and S. Tan, "Respiration rate monitoring methods: A review", *Pediatric Pulmonology*, Vol. 46, No. 6, pp. 523-529, Jun. 2011. <https://doi.org/10.1002/ppul.21416>.
- [5] T. Hussain, S. Ullah, R. Fernández-García, and I. Gil, "Wearable sensors for respiration monitoring: A review", *Sensors*, Vol. 23, No. 17, Art. no. 7518, Aug. 2023. <https://doi.org/10.3390/s23177518>.
- [6] H.-J. Kim and S.-H. Kim, "Intelligent Non-contact Respiratory Monitoring System using CNN", *Journal of Korean Institute of Information Technology*, Vol. 22, No. 4, pp. 9-15, Apr. 2024. <https://doi.org/10.14801/jkiit.2024.22.4.9>.
- [7] C. Romano, E. Schena, S. Silvestri, and C. Massaroni, "Non-contact respiratory monitoring using an RGB camera for real-world applications", *Sensors*, Vol. 21, No. 15, Art. no. 5126, Jul. 2021. <https://doi.org/10.3390/s21155126>.
- [8] C. Massaroni, D. Lo Presti, D. Formica, S. Silvestri, and E. Schena, "Non-contact monitoring of breathing pattern and respiratory rate via RGB signal measurement", *Sensors*, Vol. 19, No. 12, Art. no. 2758, Jun. 2019. <https://doi.org/10.3390/s19122758>.
- [9] H. Hwang, K. Lee, and E. C. Lee, "A real-time remote respiration measurement method with improved robustness based on a CNN model," *Applied Sciences*, Vol. 12, No. 22, Art. no. 11603, Nov. 2022. <https://doi.org/10.3390/app1222758>.

122211603.

- [10] D. Kolosov, V. Kelefouras, P. Kourtessis, and I. Mporas, "Contactless camera-based heart rate and respiratory rate monitoring using AI on hardware", *Sensors*, Vol. 23, No. 9, Art. no. 4550, May 2023. <https://doi.org/10.3390/s23094550>.
- [11] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling", *arXiv preprint arXiv:1803.01271*, Mar. 2018. <https://doi.org/10.48550/arXiv.1803.01271>.
- [12] C. Lea, M. D. Flynn, R. Vidal, A. Reiter, and G. D. Hager, "Temporal convolutional networks for action segmentation and detection", *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp. 1003-1012, Jul. 2017. <https://doi.org/10.1109/CVPR.2017.113>.
- [13] G. Heusch, A. Anjos, and S. Marcel, "A reproducible study on remote heart rate measurement", *arXiv preprint arXiv:1709.00962*, Sep. 2017. <https://doi.org/10.48550/arXiv.1709.00962>.
- [14] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, "BlazePose: On-device real-time body pose tracking", *arXiv preprint arXiv:2006.10204*, Jun. 2020. <https://doi.org/10.48550/arXiv.2006.10204>.
- [15] G. Farnebäck, "Two-frame motion estimation based on polynomial expansion", *Proc. Scandinavian Conference on Image Analysis (SCIA)*, Halmstad, Sweden, pp. 363-370, Jun.-Jul. 2003. https://doi.org/10.1007/3-540-45103-X_50.

저자소개

김 근 호 (Keun Ho Kim)



1994년 2월 : 한국과학기술원

전기및전자공학과(공학사)

2001년 8월 : 한국과학기술원

전기및전자공학과(공학박사)

2001년 9월 ~ 2002년 10월 :

미네소타주립대학교 CMRR

박사후연구원

2002년 11월 ~ 2007년 5월 : 삼성종합기술원 전문연구원

2007년 5월 ~ 현재 : 한국한의학연구원 책임연구원

관심분야 : 영상처리, AI 알고리즘, 한방 의료기기, 생체
신호처리