

# 딥러닝 기반 단일 프레임 및 연속 시퀀스 HAR에서 해상도의 영향

박주환(2017124096), 김정현(2021124046), 권은중(2017124019)

지도교수: 나종화

## 요 약

현대 사회에서 고령자 모니터링, 보안 감시, 실내 활동 분석 등 다양한 분야에서 인간 행동 인식(HAR)의 수요가 증가하고 있으나, 기존 HAR 연구에 활용된 고해상도 RGB 및 IR 센서는 사생활 침해, 비용, 환경 제약 등의 이유로 실사용에 한계가 있다. 본 연구는 이러한 문제를 해결하기 위해 저해상도 적외선(IR) 센서를 기반으로 해상도 저하가 인식 성능에 미치는 영향을 정량적으로 분석하고, 이를 보완할 수 있는 딥러닝 모델 구조를 제안하였다. 단일 프레임 데이터에는 CNN을, 연속 시퀀스 데이터에는 CNN-LSTM을 적용하여, 네 가지 해상도 조건( $32 \times 24$ ,  $16 \times 12$ ,  $8 \times 6$ ,  $8 \times 8$ )에서 분류 정확도와 학습 안정성을 비교하였다. 그 결과, CNN 모델은 해상도 감소에 따라 정확도가 0.9973에서 0.9328로 크게 하락한 반면, CNN-LSTM 모델은 0.9845에서 0.9756으로 소폭 감소하는 데 그쳐 시계열 기반 구조가 해상도 저하에 강건한 성능을 유지함을 확인하였다. 본 연구는 자원이 제한된 환경에서도 효과적인 HAR 구현이 가능함을 실험적으로 입증하였다.

**Keywords :** Human Activity Recognition (HAR), 저해상도 적외선(LRIR) 센서, CNN, CNN-LSTM, 시계열 딥러닝 모델

## I. 서 론

현대 사회에서는 고령화, 1인가구 증가, 비대면 서비스 확대 등으로 인해 사람의 자세나 동작을 인식하는 기술의 필요성이 점점 커지고 있다. 예를 들어, 독거노인의 낙상 사고를 실시간으로 감지하거나, 병실에서 응급상황을 조기에 파악하는 데 있어 사람의 행동을 정확히 인식하는 기술은 중요한 역할을 한다. 이러한 인간 동작 인식 기술은 “Human Activity Recognition (HAR)”이라 불리며<sup>[1]</sup>, 최근에는 인공지능 기술과 결합되어 다양한 분야에서 활발히 연구되고 있다. 기존의 고해상도 RGB 카메라를 기반으로 한 영상 HAR 시스템은 우수한 인식 정확도를 제공하는 장점이 있으나, 촬영 대상의 신체 특징이 명확히 드러나 개인 식별이 가능하다는 점에서 사생활 침해의 우려가 제기된다<sup>[2]</sup>. 또한, 조도가 낮은 환경에서는 인식 성능이 저하되며, 화장실·탈의실·병실 등 사적 공간에서는 활용에 근본적인 제약이 따른다<sup>[3]</sup>.

이러한 문제를 해결하기 위한 대안으로, 저해상도 적외선(Infrared, IR) array 센서를 활용한 HAR 기술이 제안되고 있다<sup>[2]</sup>. IR 센서는 비가시광 대역을 이용해 인체의 열 신호를 감지하므로, 인물 식별이 어렵고 조

명 환경에 영향을 받지 않는다는 이점이 있다.

그러나 대부분의 기존 연구는  $32 \times 24$  또는  $16 \times 12$  수준의 중해상도 IR 데이터를 기반으로 하여<sup>[4]</sup>, 여전히 인체 윤곽이나 특정 자세가 부분적으로 식별 가능하다는 한계를 가진다. 실제로 IR array 센서는 비가시광 대역의 열 정보를 감지함으로써 프라이버시 보호에 유리하지만, 공간 해상도가  $8 \times 8$  또는  $16 \times 16$  정도일 경우에만 인물의 민감한 세부정보를 숨기고 대략적인 신체 형상만을 포착할 수 있다고 보고된 바 있다<sup>[5]</sup>. 완전한 프라이버시 보호를 위해서는 이처럼 낮은 해상도의 비식별 데이터를 사용하는 것이 필요하지만, 이로 인해 공간적 정보가 감소하면서 인식 성능이 저하되는 문제가 발생한다<sup>[6]</sup>. 본 연구는 이러한 정보 손실 문제를 극복하기 위한 방안으로,  $8 \times 6$  해상도의 극저해상도 IR 데이터를 활용하되, CNN 및 CNN-LSTM과 같은 딥러닝 기반 모델 구조를 적용하여 시공간 특징을 효과적으로 학습함으로써 인식 성능을 보완하고자 한다. 이러한 접근을 바탕으로, 극단적으로 낮은 해상도 환경에서도 실용적인 수준의 HAR 성능이 가능함을 실험을 통해 검증하고자 한다.

## II. 연구 동향

Human Activity Recognition(HAR)은 다양한 센서 기반 기술을 활용하여 사람의 움직임이나 상태를 인식하는 분야로, 특히 최근에는 사생활 보호와 인식 정확도의 균형이 중요한 연구 과제로 주목받고 있다. 이에 따라 인식별이 가능한 고해상도 RGB 영상 대신, 비식별성이 높은 저해상도 적외선(IR) 센서를 활용한 HAR 연구가 활발히 이루어지고 있다.

IR 센서를 활용한 초기 연구는 주로 공간 정보에 기반한 분류 정확도 향상에 초점을 맞추었다. Gochoo et al.<sup>[7]</sup>는  $8 \times 8$  해상도의 IR 센서 세 개를 무선 센서 네트워크로 구성하고, DCNN 모델을 적용하여 다양한 자세를 분류하였다. Zhang et al.<sup>[8]</sup>은  $32 \times 24$  IR 데이터를 활용하여 연산 효율을 고려한 경량 CNN 구조인 LowResIRNet을 제안하였으며, 정확도와 연산 속도 간의 균형을 달성하였다. 한편, Lee et al.<sup>[9]</sup>는 데이터 확보의 어려움을 해결하기 위해 CycleGAN 기반의 데이터 증강 방식을 도입하였고, Ahmed et al.<sup>[6]</sup>는 세미지도학습을 통해 라벨 부족 문제를 완화하였다.

공간 해상도가 낮아짐에 따라 발생하는 정보 손실을 보완하기 위한 접근으로 시계열 정보를 활용한 구조가 주목받고 있다. Wang et al.<sup>[10]</sup>은 CNN-LSTM 구조를 적용하여 정적인 CNN 기반 공간 특징에 시간 정보를 결합함으로써 인식 성능을 향상시켰다. Ahn et al.<sup>[11]</sup>은 연속된 동작 시퀀스를 학습하는 LSTM 기반 구조를 활용하여, 동작 흐름의 맥락을 안정적으로 파악하였다. 또한 Zhou et al.<sup>[12]</sup>은 BiLSTM을 도입해 과거와 미래 방향의 시계열 정보를 동시에 반영하였고, Chen et al.<sup>[13]</sup>은 Transfer Learning 기반 CNN 특징 추출기와 BiLSTM을 결합해 복잡한 일상 동작 분류 성능을 개선하였다.

이외에도 다양한 보조적 접근이 시도되고 있다. Transformer 기반의 동작 인식 구조<sup>[14]</sup>, Super-resolution 기법을 활용한 해상도 보완 방식<sup>[15]</sup>, 다중 IR 센서의 융합 방식<sup>[16]</sup>, 그리고 CNN 경량화 및 정규화 기법<sup>[17][18]</sup> 등이 이에 해당한다. 또한 Qiu et al.<sup>[5]</sup>은 실시간 사람 수 추정을 위한 경량화된 딥러닝 모델을 제안하여, IR 센서 기반 HAR의 실용적 적용 가능성을 높이 고자 하였다. 이러한 기법들은 IR 기반 HAR 시스템의 성능과 적용 범위를 다각도로 확장하고 있다.

그러나 대부분의 기존 연구는  $16 \times 16$  이상의 중해상도 IR 데이터를 기반으로 하며,  $8 \times 8$  이하의 극저해상도 환경에서의 HAR 성능을 체계적으로 분석한 사례는 매우 제한적이다. Nguyen et al.<sup>[19]</sup>은 해상도를  $64 \times 64$ 에서  $8 \times 8$ 까지 점진적으로 낮추며 CNN 모델의 성능 저하를

정량적으로 분석하였으나, 시계열 기반 구조와의 성능 유지력 비교까지 확장된 연구는 드물다. 시계열 정보를 결합한 CNN-LSTM 구조의 성능 향상을 다룬 연구는 다수 존재하지만, 해상도 저하에 따른 구조 간 강건성 차이를 정량적으로 비교한 사례는 여전히 부족한 실정이다.

본 연구는 이러한 한계를 극복하고자,  $8 \times 6$  해상도의 극저해상도 IR 데이터를 기반으로 CNN 및 CNN-LSTM 모델을 각각 구성하고, 해상도 저하에 따른 분류 성능 변화와 구조별 강건성 차이를 정량적으로 분석하였다. 특히 CNN-LSTM 모델이 CNN에 비해 해상도 변화에 덜 민감하게 반응함을 실험적으로 확인함으로써, 공간 정보 손실을 시계열 정보로 보완하는 접근의 타당성을 제시하였다. 이는 단순한 구조 제안을 넘어, 해상도에 따른 인식 성능의 저하 양상을 모델 구조 수준에서 체계적으로 분석했다는 점에서 기존 연구와의 뚜렷한 차별성을 가진다.

### III. 연구 절차

본 연구의 전체 절차는 [그림 1]과 같이 ① 데이터 수집 및 구성, ② 모델 설계 및 학습, ③ 성능 비교 및 분석의 세 단계로 구성된다.

첫째, 데이터 수집 및 구성 단계에서는 저해상도 IR 센서를 활용하여 공간 기반 데이터와 시계열 기반 데이터를 각각 수집하고, 다양한 해상도 조건에 맞게 실험 데이터를 구성하였다. 정적 단일 프레임 데이터(Data C)와 연속 시퀀스 데이터(Data H)를 각각 구축하였으며, 다양한 해상도 시나리오에 대응하기 위해 다운스케일링과 실제 센서 수집을 병행하였다.

둘째, 모델 설계 및 학습 단계에서는 데이터 유형에 따라 CNN과 CNN-LSTM 구조를 각각 적용하였다. Data C에는 CNN 모델을 적용하여 공간 정보 학습 성능을 평가하고, Data H에는 CNN-LSTM 구조를 활용하여 시공간 정보의 활용 효과를 분석하였다.

셋째, 성능 비교 및 분석 단계에서는 해상도와 모델 구조의 다양한 조합을 대상으로 정량적 비교를 수행하였으며, 이를 통해 극저해상도 환경에서도 강건한 인식 성능을 유지할 수 있는 HAR 모델 구조를 도출하고자 하였다. 각 단계의 세부 내용은 이후 절에서 구체적으로 기술한다.

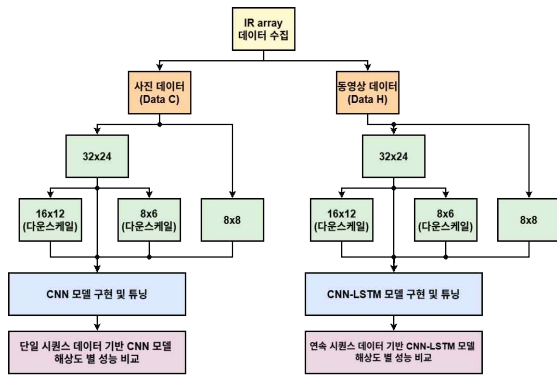


그림 1. IR 센서 기반 데이터 유형 및 모델 실험 구조

## 1. 데이터 수집

본 연구는 저해상도 환경에서의 HAR 성능을 분석하기 위해, IR 센서를 활용한 두 가지 데이터세트인 Data C와 Data H를 구축하였다. MLX90640( $32 \times 24$ )과 AMG8833( $8 \times 8$ ) 센서를 기반으로 단일 이미지와 연속 시퀀스로 구성된 IR 데이터를 수집하였으며, 해상도 조건 비교를 위해  $32 \times 24$  데이터는  $16 \times 12$  및  $8 \times 6$ 으로 다운스케일링하였다. 이에 따라 총 네 가지 해상도 조건( $32 \times 24$ ,  $16 \times 12$ ,  $8 \times 6$ ,  $8 \times 8$ )이 실험에 포함되었다.

Data C는 단일 프레임 기반의 정적 자세 인식 데이터세트로, CNN 모델의 공간 정보 처리 성능 평가를 목적으로 구축되었다([그림 2]). 피험자 3인은 통제된 실내 환경에서 다양한 자세(예: 서기, 앉기, 팔 들기 등)를 유지하며 촬영에 참여하였고, 총 8개의 동작 클래스를 반복 수행하여 약 30,000장의 IR 이미지를 수집하였다. 모든 데이터는 동일 조건에서 수집되어, 다양한 해상도에서의 분류 성능 비교가 가능하도록 설계되었다.



그림 2. Data C의 레이블별 시각화 예시

Data H는 시계열 구조의 연속 동작 인식 데이터세트

로, CNN-LSTM 모델의 시공간 정보 처리 성능 평가에 활용되었다<sup>[20]</sup>. 데이터 구축은 공개 데이터세트인 Coventry-2018의 구조를 참조하여 진행되었으며, 해당 데이터세트는  $8 \times 8$  해상도의 Grid-EYE 센서를 활용하여 다양한 단일 및 복합 동작을 고정된 위치에서 수집한 예시로, HAR 모델의 일반화 성능 평가에 널리 사용된다<sup>[8][14][21]</sup>. 본 연구는 보다 정밀한 해상도 비교를 위해  $32 \times 24$  데이터를 기반으로 직접 시계열 데이터를 수집하고, 이를  $16 \times 12$  및  $8 \times 6$ 으로 다운스케일링하였다. 여기에 별도로 수집한  $8 \times 8$  데이터(AMG8833)를 포함하여, 동일한 조건 하에 총 450개의 시계열 데이터를 확보하였다.

실험은 약  $5 \times 5\text{m}$ 의 실내 공간에서 진행되었으며, 피험자는 [표 1], [표 2]에 제시된 15개의 단일 및 이인 동작을 수행하였다. 동작 시나리오는 Coventry-2018의 분류 체계를 바탕으로 하되, 실제 센서 특성과 해상도 비교를 고려하여 수정·보완되었다.

표 1. Coventry-2018과 Data H의 단일 동작

단일		
	Coventry-2018	Data H
S1	Sit Down	앉기
S2	Stand Still	서있기
S3	Sit Down & Stand Up	앉았다 일어서기
S4	Stand Up	일어서기
S5	Left & Right Move	좌우로 움직이기
S6	Forward & Back Move	앞뒤로 움직이기
S7	Walking Diagonally 1	대각선으로 움직이기 (좌>우)
S8	Walking Diagonally 2	대각선으로 움직이기 (우>좌)

표 2. Coventry-2018과 Data H의 이인 동작

이인		
	Coventry-2018	Data H
D1	Both Sitting	두 명 가만히 앉기
D2	Sitting & Moving	한 명 앉고, 한 명 움직이기
D3	Sitting & Standing	한 명 앉고, 한 명 서있기
D4	Random Moving	두 명 아무렇게 움직이기
D5	Both Standing	두 명 서있기
D6	Standing & Moving	한 명 서있고, 한 명 움직이기
D7	Walking Cross	서로 마주보고 걷기

해상도별 시각화 결과는 해상도 저하가 단순히 공간 정보의 손실에 그치지 않고, 시계열적 패턴 보존에도 직접적인 영향을 미친다는 점을 시사한다.

Data C의 시각화 결과([그림 3])는 해상도가 낮아질수록 인체 윤곽이 흐려지고 자세 구분이 어려워지는 경

향을 보여준다.  $32 \times 24$  해상도에서는 인체 형태와 자세 차이가 비교적 명확하게 표현되지만,  $16 \times 12$ 와  $8 \times 6$ 으로 갈수록 열 분포의 세부 구조가 단순화되어 클래스 간 구분이 흐려지는 모습이 관찰된다. 특히  $8 \times 8$  해상도에서는 상지와 하지 간의 구분이 명확히 드러나지 않고, 자세 변별력이 현저히 저하되는 현상이 나타난다.

Data H의 시각화 결과(그림 4) 역시 해상도 저하에 따른 시공간 정보의 왜곡을 보여준다. 고해상도에서는 시계열적으로 나타나는 움직임의 흐름이 뚜렷하게 유지되지만, 저해상도에서는 프레임 간 온도 변화의 일관성이 약화되면서 패턴의 형태가 단순화되고, 모델이 학습 가능한 특징의 밀도가 감소한다. 이는 시공간 기반 딥러닝 모델의 인식 성능에 직접적인 영향을 미치는 요소로 작용할 수 있다.

또한  $8 \times 8$  해상도에서는 센서의 물리적 특성 차이도 중요한 변수로 나타난다. AMG8833의 경우 MLX90640과 비교해 열 감지 범위가 좁고, 온도 분포 해상도도 낮아, 동일한 행동 조건에서도 출력 영상의 밀도와 패턴이 상이하게 나타난다. 이러한 물리적·시각적 차이는 단순 해상도 차이를 넘어, 센서 자체의 응답 특성에 따른 인식 성능 차이로 이어질 수 있다.

따라서 본 연구는 이러한 해상도·센서 조건에 따른 시각적 특징과 정보 손실 양상을 기반으로, 다양한 모델 구조(CNN, CNN-LSTM)에 대한 정량적 성능 비교를 수행함으로써, 극저해상도 환경에서도 효과적으로 동작할 수 있는 HAR 모델 구조를 도출하는 것을 목표로 한다.

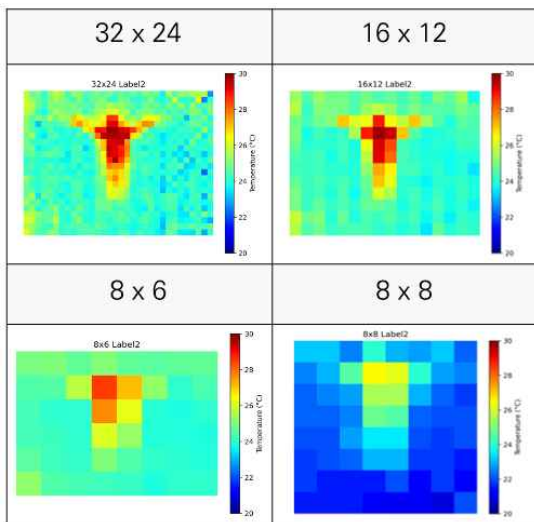


그림 3. Data C의 레이블 2 데이터의 해상도별 시각화

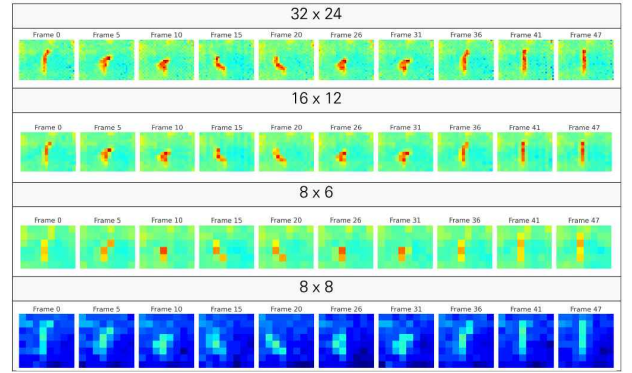


그림 4. Data H의 레이블 S3 데이터의 해상도별 시각화

## 2. 모델 학습

본 절에서는 수집된 IR 기반 데이터(Data C, Data H)에 대해 전처리, 모델 구성, 최적화 순으로 수행한 학습 절차를 기술한다. 해상도, 시퀀스 구조, 모델 구조 간의 상호작용이 분류 성능에 미치는 영향을 분석하는 것을 목표로 하였으며, 데이터 유형별 특성에 맞춰 서로 다른 정규화 방식을 적용하였다.

정적 이미지 기반의 Data C에는 CNN 모델을, 시계열 구조의 Data H에는 CNN-LSTM 하이브리드 모델을 사용하였으며, 시계열 처리 방향성에 따른 성능 차이를 분석하기 위해 LSTM과 BiLSTM 간의 비교도 함께 수행하였다. 모델 최적화는 Optuna 기반의 하이퍼파라미터 탐색 기법을 활용하여 수행하였다.

### 가. 모델별 전처리(정규화) 방식

모델 학습에 앞서, IR 센서 기반 입력 데이터에 정규화 과정을 적용하여 수치적 안정성과 학습 효율성을 확보하였다.

정적 이미지 기반의 Data C는 프레임 단위의 Min-Max 정규화를 적용하였으며<sup>[17]</sup>, 각 프레임의 픽셀 값은 해당 프레임 내 최댓값을 기준으로 0~1 범위로 변환되었다. 이는 IR 센서가 측정한 절대 온도 값보다는 상대적 열 분포 패턴이 중요하다는 점을 반영한 처리 방식이며, CNN 기반 분류 모델에서 널리 사용되는 표준 전처리 기법이다.

시계열 기반의 Data H는 샘플 간 프레임 수가 상이하므로, 모든 샘플을 40프레임으로 리샘플링하여 시퀀스 길이를 통일하였다. 이는 시계열 모델이 요구하는 고정된 입력 크기를 만족시키기 위한 사전 처리 단계로, 슬라이딩 윈도우 방식과 유사한 효과를 갖는다<sup>[22]</sup>. 정규화는 해상도별 픽셀 위치를 기준으로 전체 시퀀스를 통합하여 Z-score 방식(평균 0, 표준편차 1)으로 수

행하였다. 이 방식은 시계열 데이터의 통계적 분포 구조를 보존하면서도, 학습 과정의 수렴 안정성을 높이는 데 효과적인 것으로 알려져 있다<sup>[18]</sup>.

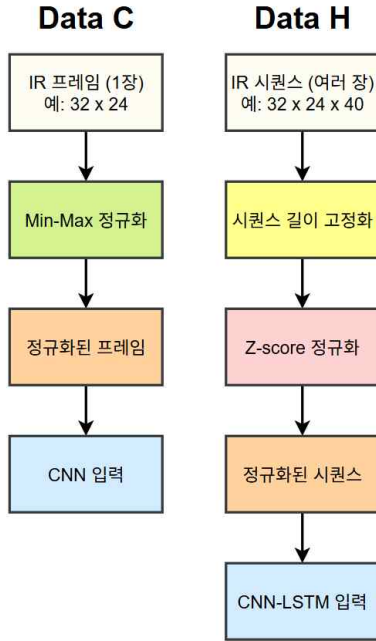


그림 5. 데이터세트별 전처리 과정 플로우차트

#### 나. 딥러닝 모델 (CNN, LSTM, CNN-LSTM)

본 연구에서는 입력 데이터의 시퀀스 구성 방식에 따라 서로 다른 딥러닝 모델 구조를 적용하였다. 정적 단일 프레임 기반의 Data C에는 Convolutional Neural Network(CNN) 모델을, 연속 시퀀스 기반의 Data H에는 CNN과 Long Short-Term Memory(LSTM)를 결합한 CNN-LSTM 하이브리드 구조를 활용하였다.

[그림 6]에 제시된 CNN 모델은 입력된 IR 이미지로부터 공간 특징을 추출하기 위한 다단계 합성곱 계층 기반의 구조를 따른다. 합성곱 계층은 각 프레임의 지역적인 열 분포 패턴을 탐지하며, 필터를 통해 저차원 입력을 점차 고차원 표현으로 확장한다. 활성화 함수로는 ReLU(Rectified Linear Unit)를 적용하여 비선형성을 도입하고, MaxPooling 계층을 통해 특징 맵의 차원을 축소하여 계산 효율성과 일반화 성능을 확보한다. 이러한 구조는 저해상도 IR 이미지에서도 인체 윤곽과 자세 구조를 안정적으로 인식할 수 있도록 설계되었으며, 정적 이미지 기반 HAR에서 뛰어난 성능을 보인다<sup>[15]</sup>.

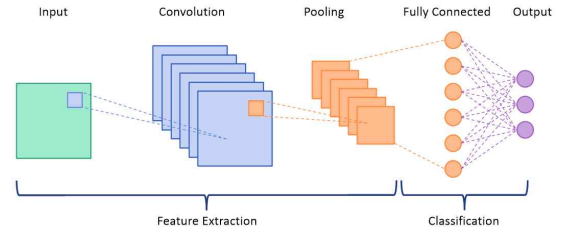


그림 6. CNN 모델 구조

[그림 7]의 CNN-LSTM 모델은 시공간 정보를 통합적으로 학습하기 위해 CNN과 LSTM을 결합한 구조로 구성된다. CNN은 각 프레임의 공간 특징을 추출하며, 이를 시간 순서대로 LSTM 계층에 전달하여 시계열적 맥락을 반영한다. LSTM은 Recurrent Neural Network(RNN) 계열의 확장 구조로, 입력 게이트, 망각 게이트, 출력 게이트로 구성된 내부 메커니즘을 통해 장기 의존성을 효과적으로 학습할 수 있다. 이러한 구조는 반복 동작, 점진적인 자세 변화 등 시간 흐름에 기반한 행동 인식에 강점을 보이며, 정적인 프레임에서 포착되지 않는 연속성과 리듬 정보를 보완한다.

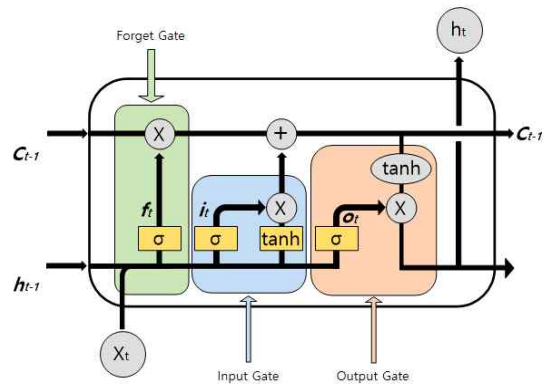


그림 7. LSTM 모델 구조

CNN-LSTM 모델은 특히 저해상도 IR 시퀀스 데이터에서도 강인한 성능을 보이며, HAR 분야에서도 그 유효성이 다수의 연구에서 검증된 바 있다<sup>[12][23][24]</sup>. 본 연구에서는 CNN-LSTM을 통해 공간 정보 손실을 시계열 정보로 보완하고, 해상도가 제한된 조건에서도 높은 인식 성능을 달성하는 것을 목표로 하였다.



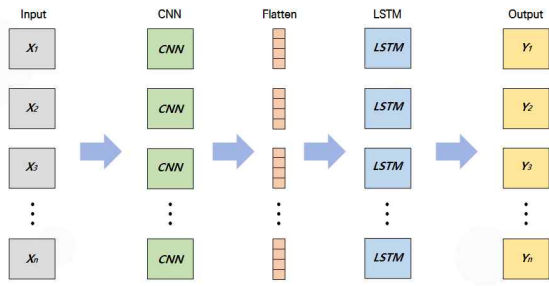


그림 8. CNN-LSTM 모델 구조

#### 다. 하이퍼파라미터 최적화 (Optuna 기반 자동 탐색 기법)

본 연구에서는 해상도별 CNN 및 CNN-LSTM 모델의 성능을 정밀하게 비교하기 위해, 자동화된 하이퍼파라미터 최적화 도구인 Optuna 프레임워크를 적용하였다. Optuna는 TPE(Tree-structured Parzen Estimator) 기반의 베이지안 최적화 알고리즘과 프루닝(Pruning) 기능을 결합하여, 성능이 낮을 것으로 예측되는 탐색 경로를 조기에 중단함으로써 계산 자원을 효율적으로 활용한다<sup>[25]</sup>.

실험에서는 필터 수, Dense 유닛 수, dropout 비율, 학습률 등 주요 하이퍼파라미터를 대상으로 해상도 조건별로 독립적인 Optuna 세션을 구성하였다. 이러한 자동 탐색 방식은 수작업 기반의 튜닝 대비 높은 성능 재현성과 공정한 모델 비교 환경을 제공한다는 점에서 본 연구의 목적에 부합한다.

Optuna의 작동 절차는 다음과 같다. 우선 사용자는 최적화할 하이퍼파라미터를 입력 변수로 정의하고, 모델 학습 및 성능 지표를 반환하는 Objective 함수를 구성한다. Optuna는 내부 샘플링 알고리즘에 따라 Trial 단위로 하이퍼파라미터 조합을 제안하며, 각 Trial 결과를 바탕으로 이후 탐색 방향을 점진적으로 개선해나간다. 이 과정에서 성능이 일정 기준 이하일 것으로 예측되는 Trial은 조기에 종료되고, 최종적으로 가장 우수한 성능을 기록한 조합이 선택된다.



그림 9. Optuna 동작방식의 플로우차트

#### 라. CNN 모델 구조

본 연구에서는 입력 해상도에 따라 최적화된 경량 CNN 모델을 개별적으로 설계하였다. 각 해상도( $32 \times$

$24, 16 \times 12, 8 \times 6, 8 \times 8$ )에 맞춰 합성곱 계층의 깊이, 필터 수, 완전연결층 구조 등을 조정하였으며, 분류 정확도를 기준으로 Optuna 기반의 하이퍼파라미터 최적화를 수행하였다.

모든 모델은 Conv2D와 MaxPooling 계층을 통해 공간 정보를 추출하고, Flatten과 Dense 계층을 거쳐 다중 클래스 분류를 수행하는 기본 구조를 따르되, 해상도 조건에 따라 계층 수와 파라미터 수를 유연하게 조정하였다. 이러한 구성은 연산 자원이 제한된 환경에서도 안정적인 분류 성능을 달성할 수 있도록 설계되었으며, 구조는 [그림 10]에 시각화하였다. 또한 [표 3]에는 입력 해상도가  $32 \times 24$ 일 때 CNN 모델의 주요 하이퍼파라미터 구성을 정리하였다.

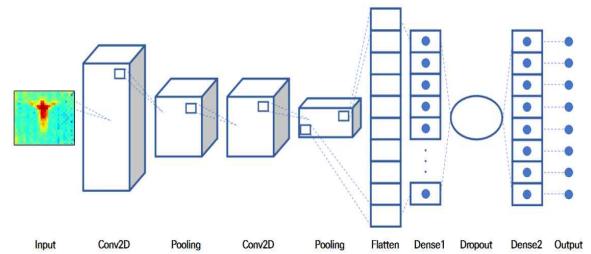


그림 10. 연구팀의 CNN 모델 구조도

표 3.  $32 \times 24$  해상도의 CNN 모델 하이퍼파라미터

하이퍼파라미터	값
Input shape	(32, 24, 1)
Conv layers	2
Filters	[16, 128]
Kernel size	(5, 5)
Activation	ReLU
Pooling	MaxPooling(2, 2)
Flatten	1920
Dense units	128
Dropout	0.018
Output layer	Softmax
Optimizer	Adam
Learning rate	0.000366
Batch size	32

#### 마. CNN-LSTM 모델 구조

해상도 저하로 인한 공간 정보 손실을 시계열 정보를 활용해 보완하고자, 본 연구에서는 [그림 10]과 같은 CNN-LSTM 하이브리드 모델을 설계하였다. 해당 모델은 입력 시퀀스의 각 프레임에 대해 TimeDistributed CNN 계층으로부터 공간 특징을 추출하고, 이후 Bidirectional LSTM 계층에서 시계열 상의 동작 패턴을 학습하는 구조를 갖는다.

양방향 LSTM은 시퀀스의 과거 및 미래 정보를 모두

반영함으로써 시계열 표현력을 강화하며, 마지막 Dense 계층에서는 이를 기반으로 동작을 분류한다. 전체 구조는 입력 해상도에 따라 유연하게 조정되었으며, 시공간 정보를 통합적으로 학습할 수 있도록 설계되었다. 입력 해상도가  $32 \times 24$ 일 때의 CNN-LSTM 모델 하이퍼파라미터 구성은 [표 4]에 정리하였다.

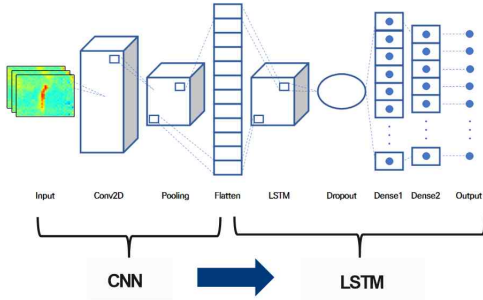


그림 11. 연구팀의 CNN-LSTM 모델 구조도

표 4.  $32 \times 24$  해상도의 CNN-LSTM 모델 하이퍼파라미터

하이퍼파라미터	값
Input shape	(40, 32, 24, 1)
Conv layers	1
Filters	64
Kernel size	(5, 5)
Activation	ReLU (Conv2D), Tanh(LSTM)
Pooling	MaxPooling (2, 2)
Flatten	12,288
LSTM units	256
LSTM dropout	0.31
Bidirectional	True
Dense units	32
Dropout(after LSTM)	0.3119
Output layer	Softmax
Optimizer	Adam
Learning rate	0.00283
Batch size	32
Epochs	50

LSTM은 입력 게이트, 망각 게이트, 출력 게이트로 구성된 내부 메커니즘을 통해 시계열 데이터의 장기 의존성을 안정적으로 학습할 수 있다. 기본적으로 단방향 구조에서는 과거 정보를 기반으로만 예측이 가능하지만, [그림 11]에 제시된 BiLSTM은 입력 시퀀스를 순방향과 역방향의 두 LSTM 셀로 병렬 처리함으로써 양방향 문맥 정보를 동시에 학습할 수 있다<sup>[26]</sup>.

이와 같은 구조는 연속 동작 인식이 요구되는 HAR

문제에서 보다 유리하게 작용할 수 있으나, 모델 파라미터 수와 연산량이 증가하여 실시간 처리 환경에서는 지연을 초래할 수 있다. 특히 저해상도 기반의 경량 HAR 시스템에서는 성능과 효율성 간의 균형이 중요한 고려사항이 된다<sup>[13]</sup>. 그럼에도, 본 연구에서는 인식 정확도 향상이 더욱 중요하다고 판단되어, 일정 수준의 처리 지연을 감수하고 BiLSTM 구조를 적용하였다.

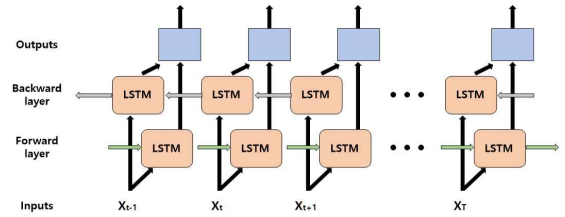


그림 12. BiLSTM 모델 구조

이에 따라 본 연구에서는 양방향 시계열 정보 활용의 실질적 효과를 검증하고자, 해상도  $32 \times 24$  조건의 시계열 IR 데이터에 대해 LSTM과 BiLSTM 구조를 동일 조건에서 비교하였다. 실험은 ablation study 방식으로 설계되었으며, CNN-LSTM 기반 모델 내의 LSTM 방향성만을 변경하여 두 구조 간 성능 차이를 정량적으로 분석하였다.

Optuna 기반의 동일 조건 하 하이퍼파라미터 최적화 결과, [그림 12]에서 확인할 수 있듯 BiLSTM 모델이 더 빠르게 안정화되었고, [그림 13]에서는 평균 정확도 0.9845로, LSTM 모델(0.8600)보다 우수한 분류 성능을 나타냈다. 이 결과는 고해상도 조건에서 양방향 시계열 학습이 HAR 모델의 일반화 성능 향상에 효과적일 수 있음을 시사한다.

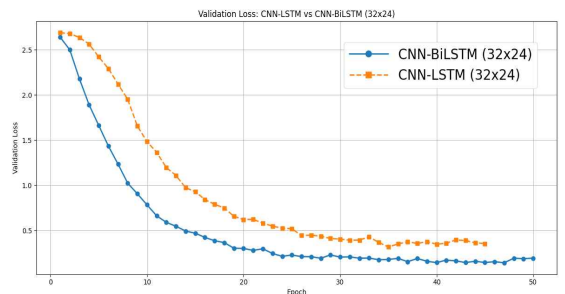


그림 13. CNN-LSTM과 CNN-BiLSTM의 Epoch 별 Validation Loss 비교

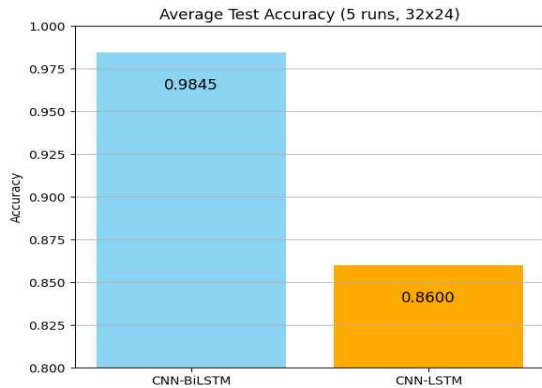


그림 14. CNN-LSTM과 CNN-BiLSTM의 정확도 비교 (5회 반복)

#### IV. 성능 평가 및 비교분석

본 장에서는 설계된 실험을 바탕으로 해상도 및 시퀀스 구조에 따른 행동 인식 성능 차이를 분석하였다. 단일 프레임 기반 CNN과 연속 시퀀스 기반 CNN-LSTM을 비교하여, 해상도 저하가 분류 정확도 및 학습 안정성에 미치는 영향을 평가하고, 반복 실험을 통해 모델의 일반화 능력과 강건성을 종합적으로 검토하였다.

##### 1. 실험 설계

다양한 해상도 조건에서 행동 인식 모델의 성능을 비교하기 위해, 본 연구는 데이터 특성과 구조에 따라 서로 다른 모델을 적용하였다. 단일 프레임 기반의 Data C에는 공간적 특징 추출에 강한 CNN 모델을 적용하고, 연속 시퀀스 기반의 Data H에는 시공간 정보를 통합적으로 학습할 수 있는 CNN-LSTM 하이브리드 모델을 구성하였다. Data C는 자체 수집한 대규모 정적 프레임 (총 30,000장)을 기반으로 해상도별 단일 학습을 수행하였으며, 무작위 요인에 따른 성능 편차는 데이터의 충분한 양으로 보완 가능하다고 판단하였다<sup>[27]</sup>. 반면, 비교적 소규모인 연속 시퀀스 데이터인 Data H는 초기화 조건이나 배치 구성에 따른 성능 변동을 고려하여, 모든 조건에서 5회 반복 학습을 수행하고 평균 정확도를 기준으로 분석하였다. 각 모델은 세 가지 해상도 조건 ( $32 \times 24$ ,  $16 \times 12$ ,  $8 \times 6$ ,  $8 \times 8$ )에 따라 서로 독립적으로 학습되었으며, 해상도 변화에 따른 모델 구조별 민감도와 강건성을 비교할 수 있도록 실험을 설계하였다. 또한 각 모델 구조에 대해 Optuna 프레임워크를 활용한 하이퍼파라미터 최적화를 수행하였으며, 필터 수, 커널 크기, Dense 유닛 수, Dropout 비율, 학습률 등을 탐색 변수로 설정하여 총 8개의 모델(해상도별 CNN 4개,

CNN-LSTM 4개)을 독립적으로 최적화하고 성능을 비교 분석하였다.

##### 2. 딥러닝 모델 최적화 전후 성능 비교

[표 5]는 해상도별로 구성된 각 모델에 대해 하이퍼파라미터 최적화 전후의 정확도 변화를 요약한 결과이다. CNN 모델은 최적화 이전에 해상도 저하에 따라 성능이 급격히 감소하는 경향을 보였으며, 특히  $8 \times 6$  해상도에서는 정확도가 0.9250에 그쳤다. 반면, 최적화 이후에는 모든 해상도 구간에서 정확도가 소폭 향상되었고,  $8 \times 6$  조건에서는 0.9328로 개선되어 저해상도에서의 학습 안정성이 일부 확보되었음을 확인할 수 있었다.

CNN-LSTM 모델 또한 전반적으로 성능 향상이 관찰되었으며, 최적화 이후 모든 해상도에서 0.97 이상의 높은 정확도를 유지하였다. 특히  $32 \times 24$  해상도에서  $0.9845(\pm 0.0054)$ ,  $8 \times 6$  해상도에서  $0.9756(\pm 0.0083)$ 의 성능을 기록하며, 최적화 과정이 시계열 기반 모델에서도 효과적으로 작용했음을 보여준다. 이로써 두 모델 모두에서 하이퍼파라미터 튜닝이 성능 향상에 기여했으며, 특히 CNN-LSTM 구조는 최적화 이후에도 해상도 변화에 강건한 성능을 지속적으로 유지하였다.

표 5. CNN과 CNN-LSTM 모델의 최적화 전, 후 해상도별 정확도

해상도	모델 종류			
	CNN		CNN-LSTM	
	최적화 전	최적화 후	최적화 전	최적화 후
32 x 24	0.9948	0.9973	$0.9022 \pm 0.0193$	$0.9845 \pm 0.0054$
16 x 12	0.9870	0.9893	$0.9355 \pm 0.0145$	$0.9822 \pm 0.0054$
8 x 6	0.9250	0.9328	$0.9422 \pm 0.0319$	$0.9756 \pm 0.0083$
8 x 8	0.9600	0.9685	$0.9222 \pm 0.0176$	$0.9756 \pm 0.0083$

##### 3. 모델 및 해상도별 정확도

###### 가. CNN 모델

Data C에 CNN 모델을 적용한 결과, 해상도에 따라 분류 정확도가 크게 달라지는 양상이 나타났다. 고해상도( $32 \times 24$ )에서는 0.9949로 매우 높은 정확도를 보였으나, 저해상도( $8 \times 6$ )에서는 0.9260으로 정확도가 크게 하락하며, 이는 CNN이 공간 정보에 민감하게 반응함을 시사한다. 특히 혼동 행렬 분석에서도 저해상도에서 유



사 동작 간의 분류 혼동이 두드러져, CNN 단독 구조의 한계와 함께 보다 강건한 모델의 필요성이 제기되었다.

표 6. Data C, 32 x 24 분류 성능평가 결과

동작	Precision	Recall	F1-score	Support
0	1.0000	1.0000	1.0000	750
1	0.9987	0.9987	0.9987	750
2	0.9973	0.9973	0.9973	750
3	0.9960	0.9960	0.9960	750
4	1.0000	1.0000	1.0000	750
5	0.9973	0.9987	0.9980	750
6	0.9920	0.9947	0.9933	750
7	0.9973	0.9933	0.9953	750
Accuracy	0.9973			6000

표 7. Data C, 8 x 6 분류 성능평가 결과

동작	Precision	Recall	F1-score	Support
0	1.0000	1.0000	1.0000	750
1	0.9290	0.9253	0.9272	750
2	0.9223	0.9493	0.9356	750
3	0.8918	0.8680	0.8797	750
4	0.9973	0.9973	0.9973	750
5	0.9520	0.9520	0.9520	750
6	0.8622	0.8840	0.8729	750
7	0.9085	0.8867	0.8976	750
Accuracy	0.9328			6000

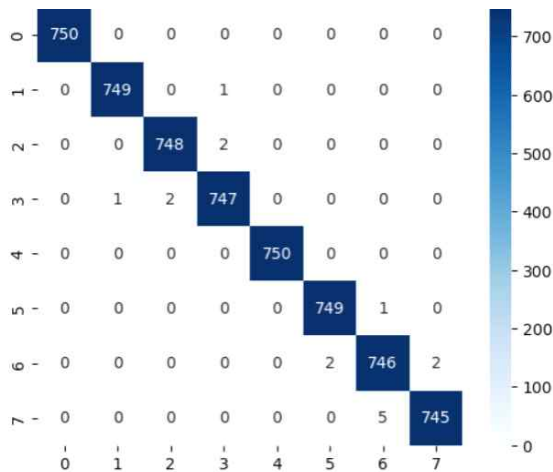


그림 15. Data C, 32 x 24 혼동행렬

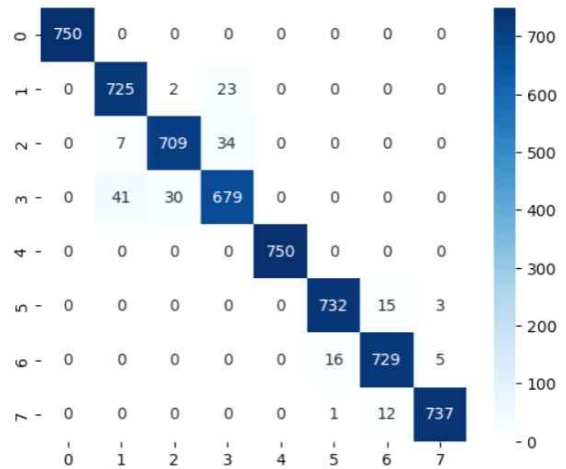


그림 16. Data C, 8 x 6 혼동행렬

## 나. CNN-LSTM 하이브리드 모델

Data H에 CNN-LSTM 모델을 적용한 결과, 해상도 저하에 따른 성능 감소는 CNN 모델에 비해 현저히 작았으며, 이는 시계열 정보를 활용한 구조의 강건성을 보여준다. 낮은 해상도에서도 높은 정확도와 낮은 편차를 유지한 것은 시간적 패턴을 학습함으로써 공간 정보 손실을 보완할 수 있었기 때문이며, 이는 CNN-LSTM 모델이 실용성과 안정성을 모두 갖춘 구조임을 시사한다<sup>[28]</sup>.

표 8. Data H, 32 x 24 분류 성능평가 결과

동작	Precision	Recall	F1-score	Support
d1	1.0000	1.0000	1.0000	6
d2	1.0000	1.0000	1.0000	6
d3	1.0000	1.0000	1.0000	6
d4	1.0000	1.0000	1.0000	6
d5	1.0000	0.8333	0.9091	6
d6	1.0000	1.0000	1.0000	6
d7	1.0000	1.0000	1.0000	6
s1	1.0000	1.0000	1.0000	6
s2	0.8571	1.0000	0.9231	6
s3	1.0000	1.0000	1.0000	6
s4	1.0000	1.0000	1.0000	6
s5	1.0000	1.0000	1.0000	6
s6	1.0000	1.0000	1.0000	6
s7	1.0000	1.0000	1.0000	6
s8	1.0000	1.0000	1.0000	6
Accuracy	0.9889			90

표 9. Data H, 8 x 6 분류 성능평가 결과

동작	Precision	Recall	F1-score	Support
d1	1.0000	0.6667	0.8000	6
d2	0.7500	1.0000	0.8571	6
d3	1.0000	1.0000	1.0000	6
d4	1.0000	1.0000	1.0000	6

d5	1.0000	0.8333	0.9091	6
d6	0.8571	1.0000	0.9231	6
d7	1.0000	1.0000	1.0000	6
s1	1.0000	1.0000	1.0000	6
s2	1.0000	1.0000	1.0000	6
s3	1.0000	1.0000	1.0000	6
s4	1.0000	1.0000	1.0000	6
s5	1.0000	1.0000	1.0000	6
s6	1.0000	1.0000	1.0000	6
s7	1.0000	1.0000	1.0000	6
s8	1.0000	1.0000	1.0000	6
Accuracy	0.9667			90

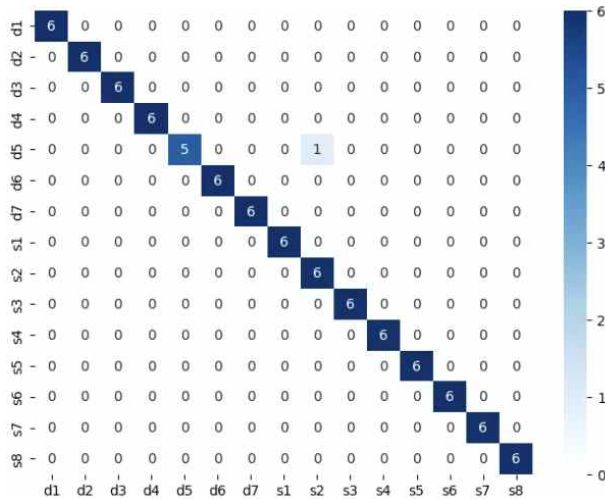


그림 17. Data H, 32 x 24 혼동행렬

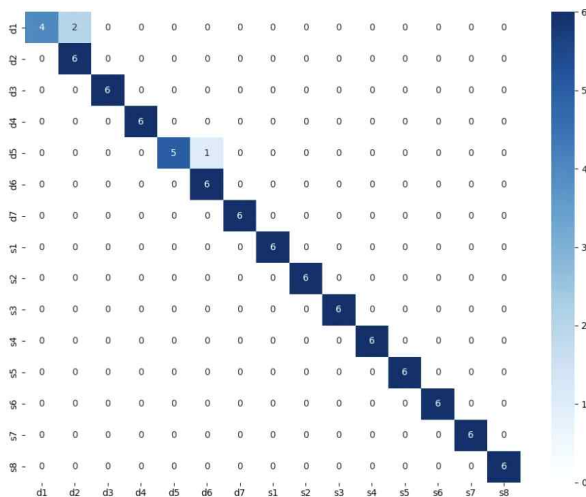


그림 18. Data H, 8 x 6 혼동행렬

#### 4. 모델별 해상도 민감도 비교

이러한 경향은 각 모델의 해상도별 정확도 수치를 통해 명확히 확인할 수 있다. CNN 모델은 32 × 24 해상도에서 0.9973의 정확도를 기록했으나, 8 × 6으로 낮아질 경우 0.9328로 감소하여 약 6.47%(0.0645)의 성능 저하가 발생하였다. 이때 오답률은 0.0027에서 0.0672로

증가하며 약 24.9배가 되었고, 이는 저해상도에서 CNN 구조의 분류 오류 가능성이 급격히 증가했음을 보여준다. 반면, CNN-LSTM 모델은 동일 조건에서 각각 0.9845( $\pm 0.0054$ )와 0.9756( $\pm 0.0083$ )의 정확도를 유지하였고, 성능 감소폭은 0.90%(0.0089)에 불과하였다. 이 경우 오답률은 0.0155에서 0.0244로 증가하며 약 1.57배에 그쳤다. 이는 CNN-LSTM 구조가 해상도 변화에 훨씬 더 강인하며, 시계열 정보를 활용한 학습 구조가 공간 정보 손실을 효과적으로 보완하고 있음을 보여주는 정량적 근거라 할 수 있다.

## V. 결 론

본 연구는 해상도 및 시계열 구성 방식에 따른 IR 센서 기반 행동 인식 성능을 정량적으로 비교하고, 저해상도 환경에서도 효과적으로 동작 가능한 CNN 및 CNN-LSTM 하이브리드 모델 구조의 가능성을 실험적으로 입증하였다. 특히, 시계열 정보를 활용한 모델이 해상도 저하에도 강건한 성능을 유지함을 확인하였으며, 이는 사생활 보호와 비용 효율성이 요구되는 환경에서 실용적인 대안이 될 수 있다. 향후 연구에서는 Transformer<sup>[29]</sup>, TCN 등 다양한 시계열 구조에 대한 비교<sup>[20]</sup>, 해상도 저하로 인한 정보 손실 보완 기법, 경량화 모델 설계<sup>[30]</sup>, 실제 응용 시나리오 기반의 평가 및 일반화 검증 등을 통해 본 연구의 한계를 보완하고 기술의 실용성을 더욱 강화할 필요가 있다.

## 참고문헌

- [1] V. R. Saxena, P. Singh, and A. Vatsa, "Human Activity Tracker and Recognition," \*Journal of Management and Service Science\*, vol. 3, no. 2, pp. 1 - 20, Nov 2023.
- [2] I. Y. Jung, "A review of privacy-preserving human and human activity recognition," \*International Journal on Smart Sensing and Intelligent Systems\*, vol. 13, no. 1, Art. 008, 2020.
- [3] S. Wang, X. Wang, B. Jiang, et al., "Human Activity Recognition using RGB-Event based Sensors: A Multi-modal Heat Conduction Model and a Benchmark Dataset," \*arXiv preprint\* arXiv:2504.05830, Apr 2025.
- [4] T. Zhang, B. Yang, and N. Gu, "Indoor Human Activity Perception Based on Infrared Thermopile Array Sensor," in \*Proc. 19th IEEE Conf. Industrial Electronics and Applications (ICIEA)\*, pp. 1 - 6,

Chengdu, China, 2024.

- [5] C. Xie, F. Daghero, Y. Chen, et al., "Efficient Deep Learning Models for Privacy-Preserving People Counting on Low-Resolution Infrared Arrays," *\*IEEE Internet of Things Journal\**, vol. 10, no. 15, pp. 13895 - 13907, Aug 2023.
- [6] M. Bouazizi, C. Ye, and T. Ohtsuki, "Low-Resolution Infrared Array Sensor for Counting and Localizing People Indoors: When Low End Technology Meets Cutting-Edge Deep Learning Techniques," *\*Information\**, vol. 13, Art. 132, pp. 1 - 32, Mar 2022.
- [7] M. Gochoo, T.-H. Tan, T. Batjargal, O. Seredin, and S.-C. Huang, "Device-Free Non-Privacy Invasive Indoor Human Posture Recognition Using Low-Resolution Infrared Sensor-Based Wireless Sensor Networks and DCNN," in *\*Proc. IEEE Int. Conf. Systems, Man, and Cybernetics (SMC)\**, pp. 2311 - 2316, Miyazaki, Japan, Oct 2018.
- [8] S. Mekruksavanich, W. Phaphan, and A. Jitpattanakul, "LowResIR-Net: Lightweight Residual Network for ADL Recognition Using Low-Resolution Infrared Sensor Data," *\*Procedia Computer Science\**, vol. 256, pp. 1358 - 1366, 2025.
- [9] K. A. Muthukumar, M. Bouazizi, and T. Ohtsuki, "An Infrared Array Sensor-Based Approach for Activity Detection, Combining Low-Cost Technology with Advanced Deep Learning Techniques," *\*Sensors\**, vol. 22, Art. 3898, pp. 1 - 24, May 2022.
- [10] G. K. Mohan, N. Gowthami, M. L. Tulasi, and M. Geethika, "Recognizing Human Activity Using Hybrid Models of CNN and LSTM in Deep Learning," *\*International Journal of Food and Nutritional Sciences\**, vol. 11, no. 12, pp. 1663 - 1674, Dec 2022.
- [11] C. Yin, J. Chen, X. Miao, H. Jiang, and D. Chen, "Device-Free Human Activity Recognition with Low-Resolution Infrared Array Sensor Using Long Short-Term Memory Neural Network," *\*Sensors\**, vol. 21, Art. 3551, 20 May 2021.
- [12] S. Mathew, A. Subramanian, S. Pooja, and B. M. S., "Human Activity Recognition Using Deep Learning Approaches: Single-Frame CNN and Convolutional LSTM," manuscript, School of Electronics Engineering, Vellore Institute of Technology, India, 2024.
- [13] N. Hassan, A. S. M. Miah, and J. Shin, "A Deep Bidirectional LSTM Model Enhanced by Transfer-Learning-Based Feature Extraction for Dynamic Human Activity Recognition," *\*Applied Sciences\**, vol. 14, Art. 603, Jan 2024.
- [14] Y. Karayaneva, S. Sharifzadeh, Y. Jing, and B. Tan, "Human Activity Recognition for AI-Enabled Healthcare Using Low-Resolution Infrared Sensor Data," *\*Sensors\**, vol. 23, Art. 478, 2 Jan 2023.
- [15] V. Sarveshwaran, I. T. Joseph, M. Maravarman, and K. P. Karthikeyan, "Investigation on Human Activity Recognition Using Deep Learning," *\*Procedia Computer Science\**, vol. 204, pp. 73 - 80, 2022.
- [16] Y. Jang, I. Jeong, M. Y. Heravi, S. Sarkar, H. Shin, and Y. Ahn, "Multi-Camera-Based Human Activity Recognition for Human - Robot Collaboration in Construction," *\*Sensors\**, vol. 23, Art. 6997, 7 Aug 2023.
- [17] W. Shi, Y. Gong, and J. Wang, "Improving CNN Performance with Min-Max Objective," in *\*Proc. 25th Int. Joint Conf. on Artificial Intelligence (IJCAI)\**, pp. 2004 - 2010, New York, USA, Jul 2016.
- [18] A. Pranolo, F. U. Setyaputri, A. K. I. Paramarta, et al., "Enhanced Multivariate Time Series Analysis Using LSTM: A Comparative Study of Min-Max and Z-Score Normalization Techniques," *\*Ilkom Jurnal Ilmiah\**, vol. 16, no. 2, pp. 210 - 220, Aug 2024.
- [19] S. P. Kannoja and G. Jaiswal, "Effects of Varying Resolution on Performance of CNN-Based Image Classification: An Experimental Study," *\*International Journal of Computer Sciences and Engineering\**, vol. 6, no. 9, pp. 451 - 456, Sep 2018.
- [20] W. C. Carlos, A. Copetti, L. Bertini, L. B. Moreira, and O. S. M. Gomes, "Human Activity Recognition: An Approach 2D CNN-LSTM to Sequential Image Representation and Processing of Inertial Sensor Data," *\*AIMS Bioengineering\**, vol. 11, no. 4, pp. 527 - 560, Nov 2024.
- [21] L. Tao, T. Volonakis, B. Tan, Y. Jing, K. Chetty, and M. Smith, "Home activity monitoring using low-resolution infrared sensor array," arXiv preprint, arXiv:1811.05416, Nov. 2018.
- [22] M. Kim and H. Kim, "A dynamic analysis data preprocessing technique for malicious code detection with TF-IDF and sliding windows," *Electronics*, vol. 13, art. 963, Mar. 2024.
- [23] A. N. Rachman, H. Mubarak, E. N. F. Dewi, and R. E. Putra, "Implementation of convolutional neural network and long short-term memory algorithms in human activity recognition based on visual processing video," *International Journal on Informatics Visualization*, vol. 7, no. 2, pp. 494-501, Jun. 2023.
- [24] D. R. Deepika, M. Lithika, V. Meenakshi, and S. Swetha, "Data securing using LSTM-CNN model," in *Proc. 2023 International Conference on Intelligent Computing and Control for Engineering and Business Systems (ICCEBS)*, Chennai, India, pp. 1-6, Oct. 2023.
- [25] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proc. 25th ACM SIGKDD*

International Conference on Knowledge Discovery and Data Mining (KDD '19), Anchorage, AK, USA, pp. 2623–2631, Aug. 2019.

[26] M. K. A. Muthukumar, M. Bouazizi, and T. Ohtsuki, “A novel hybrid deep learning model for activity detection using wide-angle low-resolution infrared array sensor,” *IEEE Access*, vol. 9, pp. 82563–82574, Jun. 2021.

[27] M. Atikuzzaman, T. R. Rahman, E. Wazed, M. P. Hossain, and M. Z. Islam, “Human activity recognition system from different poses with CNN,” in *Proc. 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI 2020)*, Dhaka, Bangladesh, pp. 1–5, Dec. 2020.

[28] D. Akinola, A. O. Oyedemi, and M. O. Ajinaja, “A deep learning-based hybrid CNN-LSTM model for human activity recognition,” *International Journal of Scientific Research in Computer Science and Engineering*, vol. 12, no. 6, pp. 66–74, Dec. 2024.

[29] L. Badarch, M. Gochoo, G. Batnasan, F. Alnajjar, and T.-H. Tan, “Ultra-low resolution infrared sensor-based wireless sensor network for privacy-preserved recognition of daily activities of living,” in *Proc. 20th IEEE International Symposium on Network Computing and Applications (NCA 2021)*, pp. 1–8, Nov. 2021.

[30] C. Yin, X. Miao, J. Chen, H. Jiang, D. Chen, Y. Tong, and S. Zheng, “Human activity recognition with low-resolution infrared array sensor using semi-supervised cross-domain neural networks for indoor environment,” *IEEE Internet of Things Journal*, vol. 10, no. 13, pp. 11761–11774, Jul. 2023.