Smart Health 40 (2026) 100652

Contents lists available at ScienceDirect

Smart Health

journal homepage: www.elsevier.com/locate/smhl

# Physical activity recognition based on electrocardiogram data only

Sina Montazeri [c], Waltenegus Dargie [b], Yunhe Feng [a], Kewei Sha [c,*]

[a] University of North Texas, Department of Computer Science and Engineering, 1155 Union Circle, Denton, 76203, TX, USA
[b] Technische Universitat Dresden, Faculty of Computer Science, Dresden, 01062, Germany
[c] University of North Texas, Department of Data Science, 1155 Union Circle, Denton, 76203, TX, USA

ARTICLE INFO

*Keywords:*
Human activity recognition
Internet of Things
Wireless electrocardiogram
Cardiac workload
Wearable medical devices
Machine learning
Transformer

ABSTRACT

Wireless electrocardiograms (ECGs) enable continuous cardiac monitoring in everyday environments. When operated in ambulatory settings, assessing whether cardiac load corresponds to physical load requires deriving activity context directly from ECG signals. We investigate ECG-based activity recognition using three deep learning models: a CNN classifier with Squeeze-and-Excitation blocks for channel recalibration, a ResNet with dilated convolutions for multiscale temporal feature capture, and a CNNTransformer hybrid combining convolutional extraction with attention-based long-range dependency modeling. Using data from 54 subjects performing six physical activities, all models achieved over 94% accuracy for seen subjects, while the CNNTransformer reached 72% for unseen subjects. Ablation studies quantified the benefits of preprocessing (EMD features: +6.0%, $p < 0.001$; signal filtering: +4.0%, $p < 0.001$) and architectural choices (skip connections: −13.0% accuracy when removed, $p < 1e-10$). Hyperparameter sensitivity analysis confirmed robustness, and uncertainty quantification enabled reliable deployment decisions with temperature scaling improving calibration by $15-20\%$. This study demonstrates, to the best of our knowledge, the first ECG-only activity classification across multiple physical activities, enabling simultaneous cardiac monitoring and activity recognition without additional motion sensors.

## 1. Introduction

According to the World Health Organization (WHO), cardiovascular diseases are the leading cause of death worldwide, claiming nearly 18 million lives annually (Kaptoge et al., 2019), with one-third occurring prematurely in people under 70. This underscores the need for affordable, scalable solutions for early detection and monitoring (Pang et al., 2018; Sivapalan et al., 2022), including unobtrusive long-term monitoring of cardiac patients in their working and living environments through wireless electrocardiograms (ECGs).

Unlike their clinical counterparts, wireless electrocardiograms are usually worn and operated by patients with little or no medical training. Furthermore, patients' activities may not be closely monitored, making ECG interpretation challenging. For example, heart rate recovery assessment requires knowing how quickly the heart returns to its parasympathetic state after physical exertion (Dogan et al., 2025; Umair et al., 2021). In clinical settings, physicians know both the time and duration of physical activity and can observe its effects in ECG recordings. In ambulatory monitoring, physicians must rely on patient reports, which may be subjective and unreliable. This raises the important question of whether the underlying physical activity can be inferred solely from the ECG measurement.

* Corresponding author.
*E-mail addresses:* SinaMontazeri@my.unt.edu (S. Montazeri), waltenegus.dargie@tu-dresden.de (W. Dargie), Yunhe.Feng@unt.edu (Y. Feng), Kewei.sha@unt.edu (K. Sha).

https://doi.org/10.1016/j.smhl.2026.100652
Received 15 December 2025; Received in revised form 25 February 2026; Accepted 1 March 2026
Available online 6 March 2026

Although electrocardiogram-based physical activity recognition has been employed, it either identifies limited physical activities or serves as a complement to IMU-based (Inertial Measurement Units) approaches (Ahmad et al., 2022; Bhoraniya & Kher, 2014; Demrozi et al., 2020; Melillo et al., 2015; Rajesh et al., 2021; Ren et al., 2024). Most contemporary research combines ECG with additional sensor modalities to improve performance. Ren et al. (Ren et al., 2024) demonstrated that the use of ECG significantly improves accelerometer-based approaches, while Rajesh et al. (2021), Ren et al. (2024) employ a multisensor fusion with ECG, photoplethysmography, and IMUs. Similarly, Melillo et al. (2015) successfully applied deep transfer learning to ECG-based fall detection with 98.44% accuracy; however, their approach is limited to distinguishing falls from general daily activities rather than recognizing specific physical activities. Bhoraniya and Kher (2014) classify body movements from motion artifacts in ambulatory ECG with 89.07% accuracy, but treat these movements as artifacts to be filtered rather than meaningful activities to be recognized. These studies consistently position the ECG as a complementary input rather than investigating its potential as the primary modality for comprehensive physical activity recognition.

This research gap motivates our investigation into ECG-only physical activity recognition across six distinct physical activities. We systematically evaluate cardiac signals as the sole input for activity classification without additional motion sensors. Our research supports the development of next-generation IoT healthcare devices with fewer sensors to reduce cost and calibration; lower power consumption, and improve patient acceptance while maintaining the comprehensive physiological evaluation capabilities essential for personalized healthcare delivery. The contributions of this paper can be summarized as follows:

- We present a methodology for detecting physical activity based exclusively on electrocardiogram data, classifying six everyday physical activities: sitting, standing, walking, skipping, running, and climbing stairs. This approach enables simultaneous reasoning about cardiac and physical exertion, offering advantages in device complexity, energy consumption, and clinical interpretation of cardiac data.
- We develop and validate three deep learning models: a CNN with Squeeze-and-Excitation blocks, a ResNet with dilated convolutions, and a novel CNNTransformer hybrid model that combines convolutional feature extraction with attention mechanisms for temporal relationship modeling.
- We conduct performance evaluation using a six-activity dataset collected from 54 subjects, with all three deep learning models achieving over 94% accuracy for seen subjects. We investigate the impact of training set size on ECG-based activity recognition. The CNNTransformer hybrid achieves 72% accuracy on unseen subjects, with performance improving as the training population increases.
- We provide ablation studies quantifying the impact of preprocessing components (EMD features: 6.0% ± 0.0% gain, $p < 0.001$; signal filtering: 4.0% ± 0.0% gain, $p < 0.001$) and architectural choices (skip connections: −13.0% ± 1.2% when removed, p < 1e-10), hyperparameter sensitivity analysis revealing robustness (all $p > 0.2$), and uncertainty quantification supporting reliable deployment decisions with temperature scaling improving calibration by 15%–20% ± 3%.

Section 2 reviews related work. Section 3 presents our methodology, data acquisition, preprocessing, and system modeling. Section 4 presents quantitative results. Section 5 discusses system performance, limitations, and challenges. Section 6 concludes and outlines future work.

## 2. Related work and motivation

Wireless electrocardiograms have evolved beyond traditional cardiac diagnostics. Recent research has explored diverse applications including activity recognition, biometric identification, stress and depression detection, and sleep analysis. This section reviews the literature in ECG-based physical activity recognition.

### 2.1. Physical activity recognition

The wireless electrocardiogram has been shown to be useful for detecting various physiological and psychological states beyond physical activity. Tanwar et al. (2024) proposed a deep learning model for ECG-based stress detection with a hybrid Long Short-Term Memory (LSTM) and gated recurrent unit neural network. Sleep analysis represents another significant application area. Feng et al. (2021) demonstrated that single channel overnight ECG data could serve as an effective screening tool for sleep-disordered breathing by using single-lead ECG data to detect sleep apnea.

Few studies have attempted to detect physical activities using ECG alone. Melillo et al. (2015) investigated ECG-based fall detection and showed that certain physiological events such as falls can be distinguished from other physical activities based on ECG patterns; their model achieved 98.44% precision in classifying falls, daily activities and non-activities. Their methodology used deep transfer learning with pre-trained CNNs (AlexNet and GoogleNet) applied to continuous wavelet transform scalograms generated from filtered ECG signals. Bhoraniya and Kher (2014), on the other hand, focused on classifying body movements based on motion artifacts in ambulatory ECG. Their model achieved an accuracy of 89.07% in discriminating activities such as hand movements, hip rotations, and the transition from sitting to standing. Their approach used adaptive filtering to extract motion artifacts, followed by the extraction of Gabor transform characteristics (Qian & Chen, 2002) and the classification of multilayer perceptrons.

A significant portion of current research combines ECGs with other sensor modalities to improve activity recognition. In Mahmud et al. (2020), the authors investigated human activity physical recognition using a wearable patch that combines triaxial accelerometer and ECG. They showed that additional use of an ECG significantly improves results compared to accelerometer-only approaches. Similarly, Ahmad et al. (2022) used deep residual networks to classify physical exercise using ECG, photoplethysmography (PPG), and IMU. In Rajesh et al. (2021), the authors conducted a comparative study to evaluate the individual and combined contributions of accelerometer, ECG, and PPG signals to HAR and discuss the relative importance of different sensor modalities.

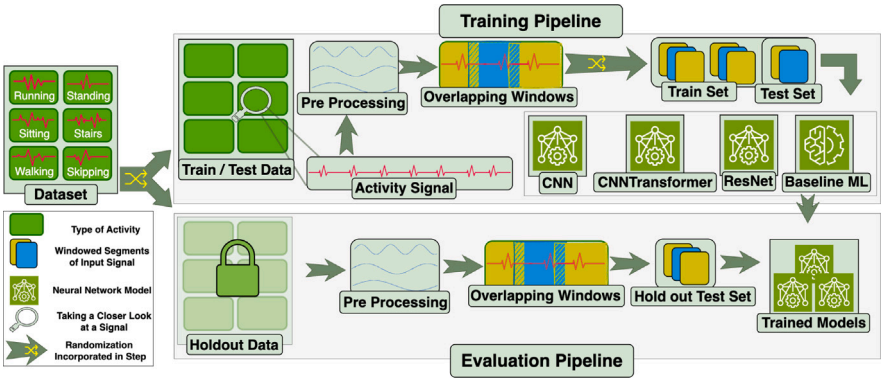



**Fig. 1.** The ECG-only human activity recognition framework.

### 2.2. Research gaps and contributions

Most existing ECG-based physical activity recognition systems rely on multi-sensor platforms. Studies using ECGs typically treat them as complementary signals to traditional motion sensors rather than the primary modality. For example, Bhoraniya and Kher (2014) treated body movements as artifacts to be classified rather than meaningful activities to be recognized. Our research addresses this gap by investigating physical activity recognition based solely on ECG across six physical activities. The systematic comparison of CNN, ResNet, and CNNTransformer architectures provides new insight on the feasibility and limitations of cardiac signal-based activity classification. This work demonstrates that meaningful insights about underlying physical activities can be gained from ECG alone, establishing a paradigm for simplified wearable cardiac monitoring systems. Reducing sensor count lowers costs, system overhead, and calibration requirements while improving patient acceptance.

## 3. Methodology

Our methodology (Fig. 1) establishes a framework for physical activity recognition using electrocardiogram data as the sole input.

Raw ECG data undergo preprocessing to enhance signal quality and extract features, addressing motion artifacts and physiological noise while preserving activity-relevant cardiac patterns.

We evaluate generalization by using subject-wise data splits, which test whether the models can recognize physical activities in individuals they have not seen during training. In all of our experiments, we create two randomized subsets of individuals in our dataset with 80% for training and 20% of individuals reserved as holdout test data.

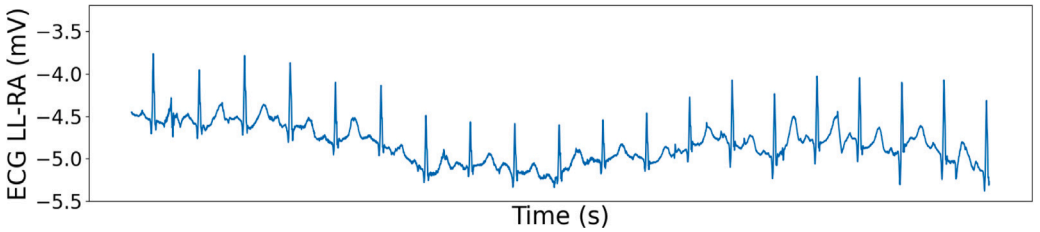
We develop three neural network architectures adapted for ECG-based physical activity classification. The CNN classifier incorporates Squeeze-and-Excitation blocks for channel-wise feature recalibration. The ResNet classifier employs dilated convolutions within residual blocks to capture multiscale temporal dependencies. The CNNTransformer hybrid combines convolutional feature extraction with attention mechanisms to model both local morphological patterns and long-range temporal relationships in cardiac signals.

Existing public ECG datasets are inadequate for ECG-based physical activity recognition. Clinical ECG databases like PTB-XL (Wagner et al., 2020) (18,869 patients) contain recordings from stationary clinical settings without physical activity labels. Conversely, activity recognition benchmarks like UCI HAR (Anguita et al., 2013) (30 subjects) rely exclusively on inertial sensors without ECG data. Datasets combining ECG with activity monitoring, such as MMASH (Rossi et al., 2020) (22 subjects) and PPG-DaLiA (Reiss et al., 2019) (15 subjects), have insufficient subjects for cross-subject generalization and typically focus on aggregated activity levels or heart rate estimation rather than discrete activity classification from raw ECG waveforms. The Motion Artifact Contaminated ECG Database (Behravan et al., 2015), while recording ECG during physical movements, contains data from only a single subject and treats motion as an artifact to be removed rather than an informative feature. This gap in available datasets – the absence of a large-scale ECG dataset with labeled diverse physical activities – motivated our collection of ECG recordings from 54 subjects performing six distinct physical activities (sitting, standing, walking, skipping, running, and climbing stairs).

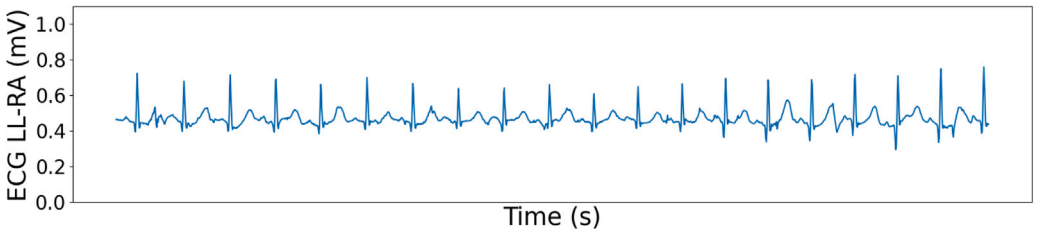
### 3.1. Data acquisition and preprocessing

We used the Shimmer platform (version 3),[1] a 5-lead wireless ECG, to measure the electric activities of the heart (Burns et al., 2010; Cao et al., 2022; Gradl et al., 2012). The data presented in this work were collected with the approval of the institutional ethics

[1] https://shimmersensing.com/product/consensys-ecg-development-kits/.




(a) Original ECG segment (512 Hz).




(b) Preprocessed ECG segment (50 Hz).

**Fig. 2.** Comparison of original and preprocessed ECG segments during the Skipping activity.

committee. Full consent from all participants had been obtained prior to the experiments. Participants were college-aged students (35 male, 19 female) with no known cardiac conditions. Each participant performed six activities (sitting, standing, walking, skipping, running, and climbing stairs) for approximately two minutes each. Activities were recorded in separate sessions throughout the day rather than consecutively, reducing fatigue-related confounds.

Two of the ECG leads provided the inputs to our models: Lead II (Left Leg - Left Arm) and Lead I (left Leg - Right Arm). Fig. 2(a) shows a five-second sample of the skipping activity originally recorded by a subject. Raw ECG data suffer from baseline drifts, causing low-frequency oscillations at 0.05–1 Hz that require removal. We apply a high-pass Butterworth filter (cutoff frequency 0.5 Hz, filter order 5), then normalize and downsample each signal from 512 Hz to 50 Hz. We empirically determined that 50 Hz was sufficient for activity classification. Fig. 2(b) illustrates the sample signal after its baseline drift removal and downsampling and normalization.

We then enhance the downsampled, denoised data using Empirical Mode Decomposition (EMD) (Flandrin et al., 2004), extracting Intrinsic Mode Functions (IMFs) (Flandrin et al., 2004) that capture different frequency components. For each ECG channel, we generate eight IMFs representing the signal at various temporal scales. Fig. 3 shows each extracted IMF for the example signal in Fig. 2(b). This decomposition allows models to learn from both the original signals and their constituent frequency bands, providing richer feature representations for activity classification.

Following feature extraction, we segment each activity into fixed-length windows of 256 samples (5 s at 50 Hz) with a sliding step of 64 samples (1.28 s), resulting in 75% overlap between consecutive windows. Overlapping windows ensure sufficient temporal context while maintaining adequate data coverage. This transforms continuous recordings into discrete samples suitable for neural network processing.

### *3.2. Neural network architectures*

Physical activity classification from ECG requires architectures that extract both local morphological features and temporal dependencies. We explore three neural network architectures: a CNN with Squeeze-and-Excitation blocks for channel-wise feature recalibration; a ResNet with dilated convolutions for multi-scale temporal pattern recognition; and a CNNTransformer hybrid combining convolutional feature extraction with attention mechanisms for long-range dependency modeling. The following sections detail each architecture.

#### *3.2.1. CNN classifier*

One-dimensional (1D) CNNs are well suited for processing time series data like ECG segments (Martis et al., 2015). They use convolutional filters to learn local patterns (e.g., waveform shapes, morphological motifs) and pooling layers to build hierarchical representations (Wang & Raj, 2015). CNNs have demonstrated excellent performance in ECG arrhythmia classification and HAR using other sensor modalities (Martis et al., 2015; Wang & Raj, 2015).

Our CNN architecture (Fig. 4) uses convolutional blocks with increasing filter sizes (64, 128, 256, 512) to progressively extract hierarchical features (Simonyan & Zisserman, 2015). The model incorporates Squeeze-and-Excitation (SE) blocks to adaptively recalibrate channel-wise feature responses (Hu et al., 2018). We use Gaussian Error Linear Unit (GELU) activation throughout (Devlin et al., 2018). The model concludes with global average pooling, followed by fully connected layers with dropout regularization (Lin et al., 2014).

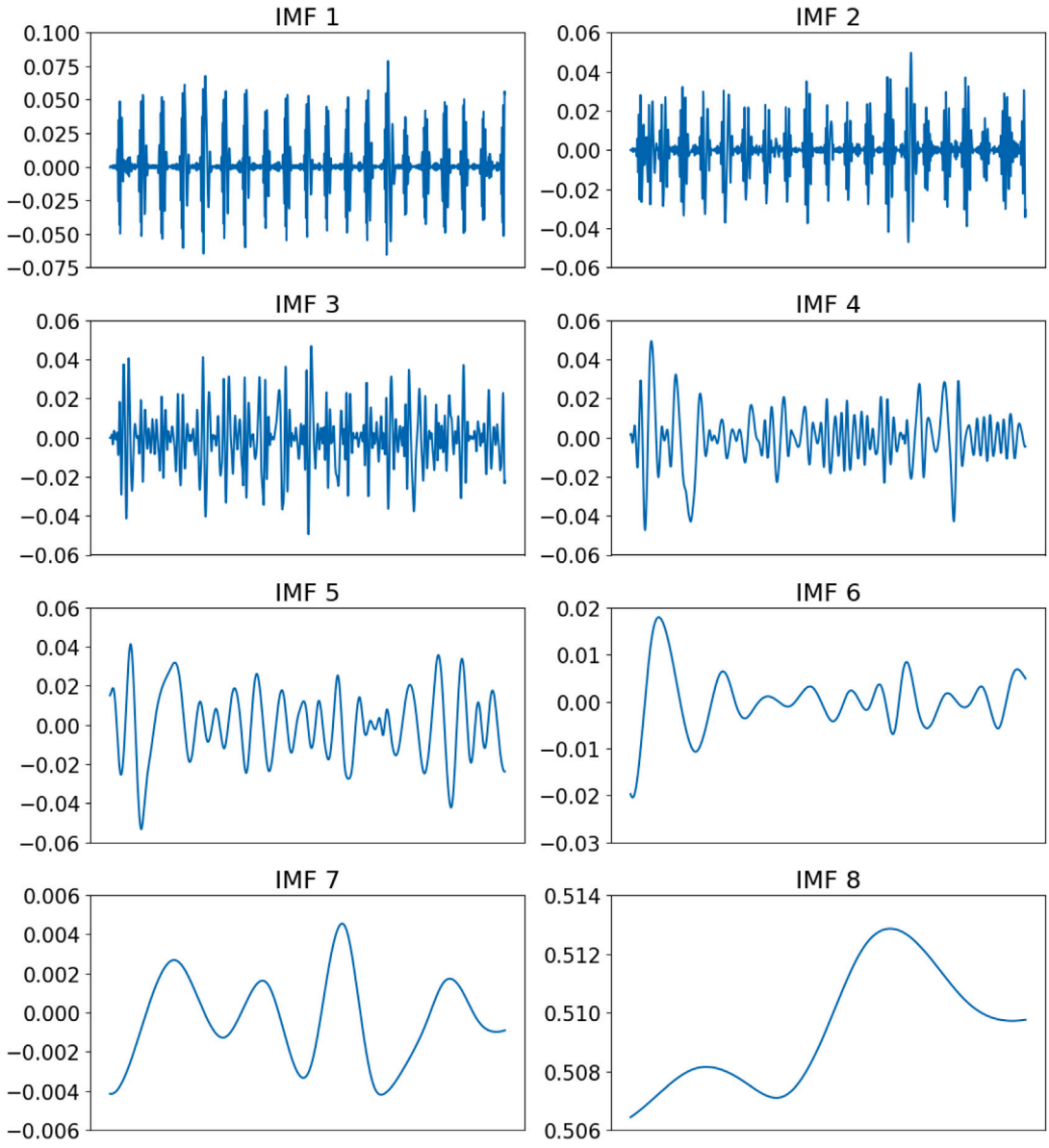



**Fig. 3.** The 8 extracted IMFs for the example signal.

### *3.2.2. Residual network (ResNet) classifier*

Physical activity recognition from ECG requires capturing both fine-grained waveform details and broader temporal patterns. While CNNs extract local features effectively, increasing network depth causes vanishing gradients that limit performance. ResNets address this limitation through shortcut connections that allow gradients to flow directly through the network (He et al., 2016). We adapt ResNet for one-dimensional time-series data (Fig. 5), beginning with an initial convolutional layer followed by residual blocks. Each block comprises two convolutional layers with batch normalization and ReLU activation, where identity shortcut connections add the block input to its output, enabling the network to learn residual functions.

We incorporate dilated convolutions within residual blocks to capture features at multiple temporal scales. Dilated convolutions expand the receptive field without increasing parameters, allowing the model to learn dependencies over longer time intervals. After residual blocks, the model applies global average pooling, followed by a fully connected layer for activity classification. Table 1 summarizes the ResNet architecture parameters.

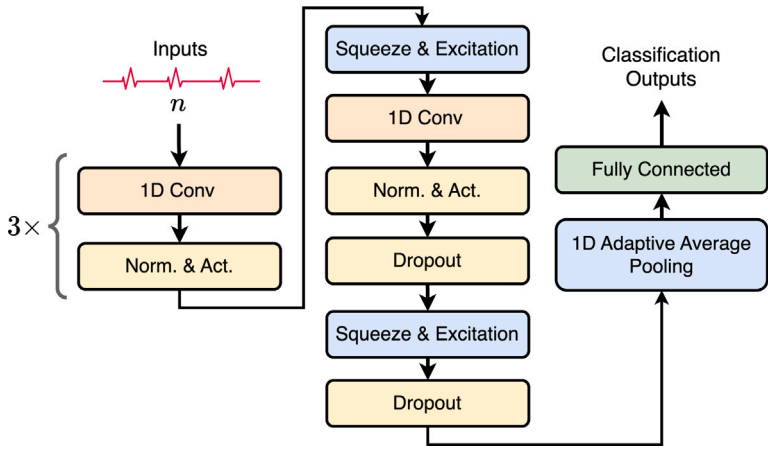



**Fig. 4.** Architecture of the proposed CNN Classifier. The model uses convolutional blocks with increasing filter sizes and incorporates Squeeze-and-Excitation blocks for channel-wise feature recalibration, followed by global average pooling and fully connected layers for 6-class activity classification.

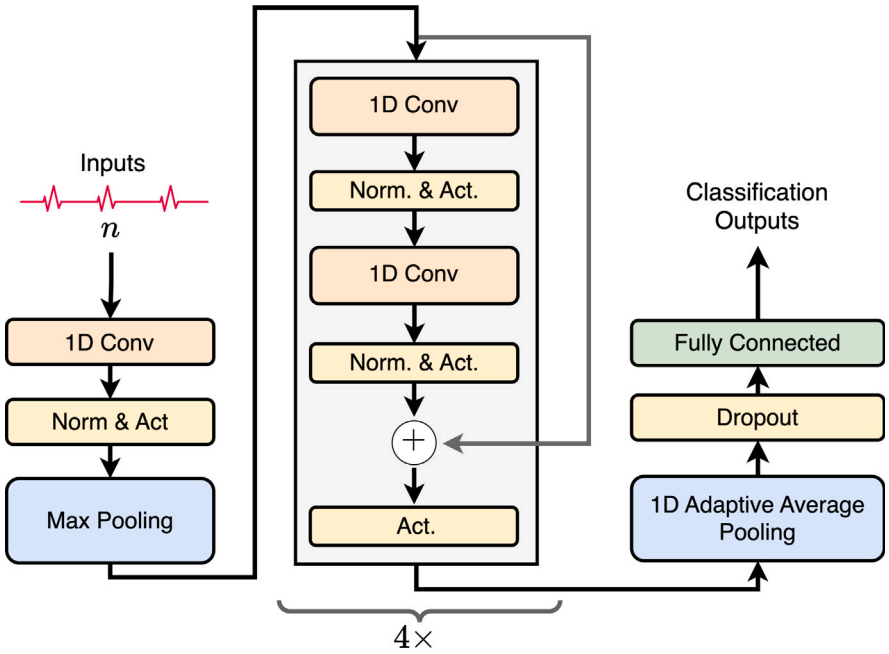



**Fig. 5.** Architecture of the ResNet Classifier. There are 4 residual blocks incorporated. Additionally, ResNet's original classification head is replaced with global average pooling followed by a fully connected layer to support 6-class activity classification.

**Table 1**
ResNet architecture configuration.

| Architectural component | Configuration |
|---|---|
| Input Processing | 7 × 7 Conv (256 filters, stride=2) → 3 × 3 MaxPool |
| Residual Layers | 4 layers: [256, 512, 1024, 2048] filters |
| Blocks per Layer | 4 residual blocks each |
| Multi-scale Features | Dilation rates: [1, 2, 4, 8] |
| Regularization Strategy | Dropout: 0.4 (blocks), 0.4 (classifier) |
| Classification Head | Global Average Pool → FC (6 classes) |

**Table 2**
Parameter sizes for the transformer classifier.

| Component | Size |
|---|---|
| Embedding Dimension ($d_{\text{model}}$) | 512 |
| Number of Attention Heads | 8 |
| Number of Encoder Layers | 8 |
| Feedforward Hidden Dimension | 512 |
| Dropout Rate | 0.2 |
| Output Classes | 6 |
| Maximum Sequence Length | 12 000 |

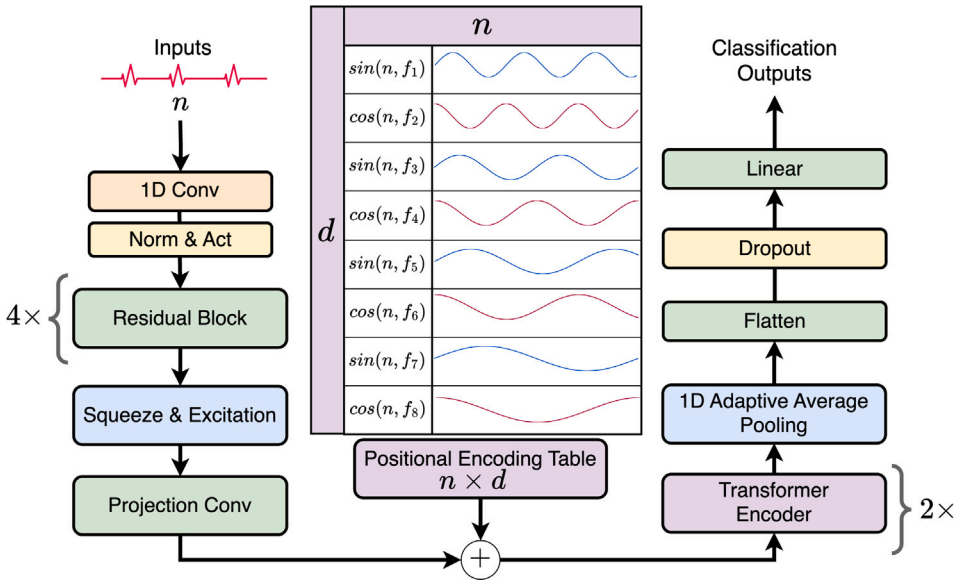



**Fig. 6.** Architecture of the CNNTransformer classifier for an input size of length $n$ and an embedding dimension of size $d$. The model combines convolutional feature extraction with Squeeze-and-Excitation blocks, followed by Transformer encoder layers with positional encoding for temporal relationship modeling.

### 3.2.3. CNNTransformer hybrid classifier

Activity recognition requires capturing both localized waveform characteristics and longer-term temporal dependencies. Convolutional approaches excel at short-term morphological patterns but may overlook complex temporal relationships. Conversely, Transformers model dependencies across extended sequences but often lack capacity to extract local features from raw data.

We propose a hybrid CNNTransformer model (Fig. 6) to leverage these complementary strengths. Convolutional layers extract local features from ECG waveforms. A Squeeze-and-Excitation (SE) block (Hu et al., 2018) adaptively recalibrates channel responses based on their importance. The extracted convolutional features are then projected into a high-dimensional embedding space suitable for sequence modeling. To incorporate explicit temporal structure within each discrete window, we employ sinusoidal positional encodings as introduced by Vaswani et al. (2017). Given a window length $T$ and embedding dimension $d_{\text{model}}$, the positional encodings $\text{PE}_{(pos,2i)}$ and $\text{PE}_{(pos,2i+1)}$ are computed as follows:

$$\text{PE}_{(pos,2i)} = \sin\left(\frac{pos}{10000^{2i/d_{\text{model}}}}\right), \tag{1}$$

$$\text{PE}_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{2i/d_{\text{model}}}}\right), \tag{2}$$

where $pos$ represents the position within the window ($pos \in [0, T)$), and $i$ denotes the embedding dimension index. This positional encoding embeds time point positions within each window, enabling Transformer layers to recognize temporal arrangements and dependencies. Transformer encoder layers, consisting of multi-head self-attention and position-wise feedforward sub-layers, process these embeddings. The self-attention mechanism identifies relevant dependencies in time steps, distinguishing between transient physiological fluctuations and stable activity-induced patterns.

The Transformer output is temporally averaged to produce a fixed-length representation. A fully connected layer classifies this representation into activity categories. Table 2 summarizes the Transformer architecture parameters.

**Table 3**
Training hyperparameters for classifier models.

| Stage | Learning rate (LR) | Minimum LR | Epochs | Weight decay |
|---|---|---|---|---|
| 1 | 4e−4 | 1e−6 | 30 | 1e−4 |
| 2 | 1e−4 | 1e−8 | 30 | 1e−3 |

**Table 4**
Model performance metrics on the test set: Values are rounded up.

| Model | Accuracy | Precision | Recall | F1 |
|---|---|---|---|---|
| CNNTransf. | 0.94 ± 0.02 | 0.94 ± 0.02 | 0.94 ± 0.02 | 0.94 ± 0.02 |
| ResNet | 0.96 ± 0.02 | 0.96 ± 0.02 | 0.96 ± 0.02 | 0.96 ± 0.02 |
| CNN | 0.96 ± 0.03 | 0.96 ± 0.02 | 0.96 ± 0.03 | 0.96 ± 0.03 |

**Table 5**
Model Performance metrics on the holdout test set: Values are rounded up.

| Model | Accuracy | Precision | Recall | F1 |
|---|---|---|---|---|
| CNNTransf. | 0.72 ± 0.02 | 0.73 ± 0.02 | 0.72 ± 0.02 | 0.72 ± 0.02 |
| ResNet | 0.67 ± 0.02 | 0.66 ± 0.02 | 0.67 ± 0.02 | 0.66 ± 0.02 |
| CNN | 0.61 ± 0.01 | 0.64 ± 0.02 | 0.61 ± 0.01 | 0.62 ± 0.01 |

### *3.3. Training methodology*

We adopt a subject-wise splitting strategy, dividing subjects into training and holdout test sets to provide a more realistic evaluation scenario than random splitting. This helps identify potential overfitting to individual physiological characteristics. The subject-wise partition is performed before windowing (Section 3.1), ensuring that no overlapping windows from the same subject appear in both sets, preventing data leakage.

Training uses the Adam optimizer (Kingma & Ba, 2014) with weight decay for L2 regularization and implements a dynamic learning rate schedule based on validation loss plateaus. To handle class imbalance, weighted loss functions adjust class contributions based on frequency. Early stopping based on validation accuracy prevents overfitting.

A multistage training protocol (Bengio et al., 2009) allows each stage to target specific accuracy thresholds with customized learning rates and regularization parameters. As the model approaches optimal performance, training stages use decreasing learning rates and increasing regularization. Stage 1 uses learning rate 4e-4 for rapid initial convergence, while stage 2 reduces it to 1e-4 for precise parameter updates (Loshchilov & Hutter, 2016). Weight decay increases from 1e-4 to 1e-3 between stages. Table 3 details the hyperparameters.

We track performance metrics including accuracy, loss, precision, recall, and F1 scores for training and validation sets. Confusion matrices and misclassification analyses provide insight into model behavior and identify challenging activity transitions.

## 4. Experimental results

We conducted experiments to assess performance generalization and the impact of training set size on classification accuracy using subject-wise data splitting.

To test performance on unseen subjects, 20% of subjects (10 individuals) were set aside as a holdout test set, never used in training. The remaining subjects were divided into training and validation sets, resulting in 30 activity instances per type for training, 10 for holdout testing, and 7 for validation. All experimental configurations were repeated in 10 independent trials with randomized subject assignment to ensure statistical reliability.

### *4.1. Performance analysis*

Table 4 shows that all three architectures achieved high performance on seen subjects. ResNet and CNN models both reached 96% accuracy, with ResNet showing slightly lower variance (±0.02) compared to CNN (±0.03). The CNNTransformer hybrid achieved 94% accuracy (±0.02). All architectures learned activity-specific ECG patterns when test subjects were included in the training population.

Precision, recall, and F1 scores remained consistently high across all models on seen subjects, ranging from 0.94 to 0.96. This consistency suggests models learned to distinguish between six activity classes without significant bias. Low standard deviations ($\leq$0.03) demonstrate stable performance across 10 trials. Holdout test set results (Table 5) reveal different performance on unseen subjects. The CNNTransformer hybrid demonstrated superior generalization, achieving 72% accuracy compared to 67% for ResNet and 61% for CNN. This 11 percentage point advantage over CNN and 5 percentage point lead over ResNet establishes CNNTransformer as the most robust approach for cross-subject generalization.

Fig. 7 and Table 6 detail CNNTransformer performance on unseen subjects. The confusion matrix reveals distinct misclassification patterns, particularly between activities with similar metabolic demands. Walking shows the lowest performance (F1-score = 0.47),

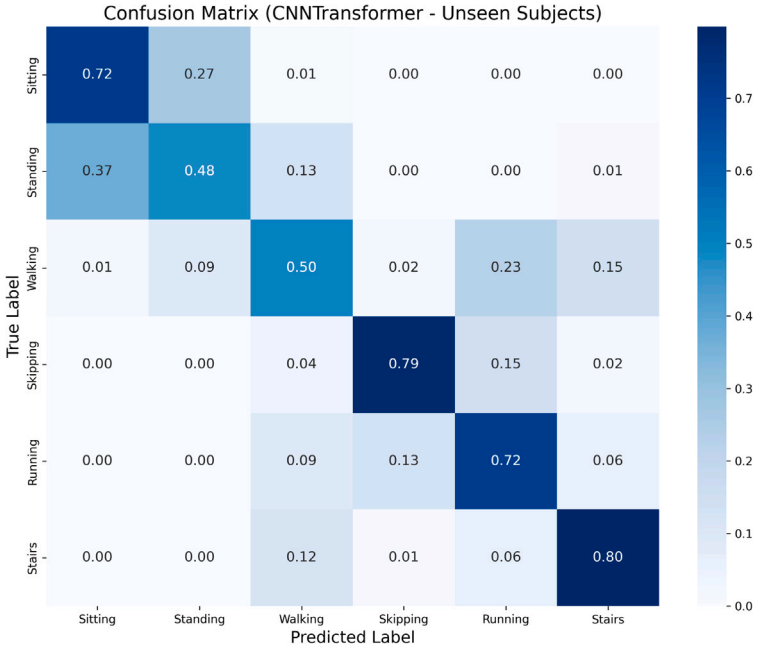



**Fig. 7.** Aggregated Confusion Matrix for CNNTransformer on Unseen Subjects (Holdout Test Set). Predictions are summed across 10 independent trials.

**Table 6**
Per-Class performance metrics for CNNTransformer on unseen subjects (holdout test set). Values are mean ± standard deviation across 10 trials.

| Activity | Precision | Recall | F1-Score |
|---|---|---|---|
| Sitting | 0.66 ± 0.10 | 0.72 ± 0.13 | 0.68 ± 0.10 |
| Standing | 0.57 ± 0.10 | 0.48 ± 0.10 | 0.52 ± 0.09 |
| Walking | 0.46 ± 0.06 | 0.50 ± 0.13 | 0.47 ± 0.09 |
| Skipping | 0.80 ± 0.09 | 0.79 ± 0.06 | 0.79 ± 0.04 |
| Running | 0.61 ± 0.06 | 0.72 ± 0.09 | 0.65 ± 0.04 |
| Stairs | 0.90 ± 0.07 | 0.80 ± 0.07 | 0.84 ± 0.03 |

often confused with Sitting and Standing, likely due to overlapping heart rate profiles at low-to-moderate exertion. In contrast, high-intensity activities such as Skipping (F1-score = 0.79) and Stairs (F1-score = 0.84) are recognized with high reliability. Low standard deviations in Table 6 confirm that these patterns are consistent across trials.

Low standard deviations (0.02) for all holdout test metrics demonstrate that these performance differences are statistically significant and reproducible across trials. CNNTransformer's consistent superiority across accuracy, precision, recall, and F1-score establishes its effectiveness for deployment scenarios requiring recognition in previously unseen individuals.

### *4.2. Impact of training set size on model performance*

We examined how training set size affects performance on unseen subjects to determine whether additional training data improves generalization or whether performance plateaus. We varied the number of training subjects from 1 to 37 subjects per activity (3% to 100% of available data), with intermediate points at 2, 3, 4, 7, 11, 18, and 29 subjects per activity (6%, 9%, 11%, 19%, 30%, 49%, and 79%). We maintained a consistent holdout test set of 10 subjects (20%) throughout.

Fig. 8 shows how model accuracy varies with training subjects. None of the architectures reached performance saturation, even when using the complete training dataset. The continued improvement in the final scaling step shows that additional subject diversity would likely yield further performance gains. This finding has important implications for future data collection efforts and indicates that ECG-based activity recognition would benefit from larger, more diverse training populations. A consistent pattern emerged across all architectures: significant performance improvements occurred around 49% of training subjects (18 subjects per activity). This finding shows a minimum requirement to achieve clinically acceptable performance levels. Below this threshold, model uncertainty and performance variability make reliable physical activity recognition impractical for real-world applications.

CNNTransformer (Table 7) showed the most dramatic performance scaling, with a 59 percentage point improvement representing the largest gain among all architectures. The learning trajectory showed different phases: low initial performance (13%–37%)

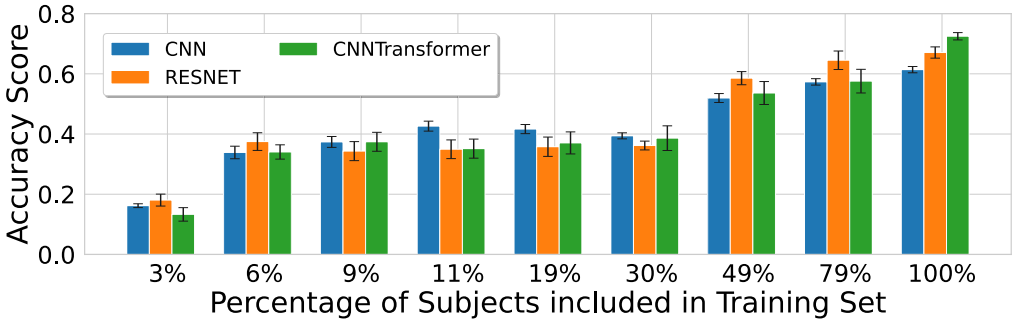



**Fig. 8.** Model accuracy vs. number of training subjects.

**Table 7**
CNNTransformer — Performance as percentage of subjects included in training set increases.

| Subjects (%) | Accuracy | Precision | Recall | F1 |
|---|---|---|---|---|
| 3 | 0.13 ± 0.03 | 0.14 ± 0.06 | 0.13 ± 0.03 | 0.09 ± 0.04 |
| 6 | 0.34 ± 0.03 | 0.36 ± 0.06 | 0.34 ± 0.03 | 0.30 ± 0.02 |
| 9 | 0.37 ± 0.04 | 0.36 ± 0.06 | 0.37 ± 0.04 | 0.33 ± 0.02 |
| 11 | 0.35 ± 0.04 | 0.40 ± 0.05 | 0.35 ± 0.04 | 0.33 ± 0.04 |
| 19 | 0.37 ± 0.05 | 0.36 ± 0.04 | 0.37 ± 0.05 | 0.33 ± 0.03 |
| 30 | 0.39 ± 0.05 | 0.46 ± 0.06 | 0.39 ± 0.05 | 0.38 ± 0.05 |
| 49 | 0.54 ± 0.05 | 0.59 ± 0.06 | 0.54 ± 0.05 | 0.51 ± 0.05 |
| 79 | 0.58 ± 0.05 | 0.63 ± 0.06 | 0.58 ± 0.05 | 0.57 ± 0.05 |
| 100 | 0.72 ± 0.02 | 0.73 ± 0.02 | 0.72 ± 0.02 | 0.72 ± 0.02 |

**Table 8**
RESNET — Performance as percentage of subjects included in training set increases.

| Subjects (%) | Accuracy | Precision | Recall | F1 |
|---|---|---|---|---|
| 3 | 0.18 ± 0.03 | 0.24 ± 0.06 | 0.18 ± 0.03 | 0.17 ± 0.04 |
| 6 | 0.37 ± 0.04 | 0.42 ± 0.03 | 0.37 ± 0.04 | 0.39 ± 0.04 |
| 9 | 0.34 ± 0.04 | 0.44 ± 0.03 | 0.34 ± 0.04 | 0.37 ± 0.04 |
| 11 | 0.35 ± 0.04 | 0.48 ± 0.05 | 0.35 ± 0.04 | 0.37 ± 0.04 |
| 19 | 0.36 ± 0.04 | 0.48 ± 0.05 | 0.36 ± 0.04 | 0.36 ± 0.05 |
| 30 | 0.36 ± 0.02 | 0.49 ± 0.03 | 0.36 ± 0.02 | 0.37 ± 0.02 |
| 49 | 0.59 ± 0.03 | 0.61 ± 0.04 | 0.59 ± 0.03 | 0.58 ± 0.03 |
| 79 | 0.65 ± 0.04 | 0.65 ± 0.04 | 0.65 ± 0.04 | 0.64 ± 0.04 |
| 100 | 0.67 ± 0.02 | 0.66 ± 0.02 | 0.67 ± 0.02 | 0.66 ± 0.02 |

through the first scaling points, a critical transition between 30%–49% of training subjects (39% to 54% accuracy) and the most substantial final improvement from 58% to 72%. Precision, recall, and F1 score metrics followed similar scaling patterns (0.14 to 0.73, 0.13 to 0.72, and 0.09 to 0.72 respectively), which shows a balanced performance across activity classes is possible only with substantial training diversity.

ResNet (Table 8) showed a 49 percentage point improvement. Performance remained stable at 34%–37% accuracy during initial scaling phases (up to 30% of training subjects), then jumped to 59% at 49% of subjects and continued to improve to 67% with the complete set. Precision, recall, and F1-score scaled from 0.66, 0.67, and 0.66 respectively. The model maintained balanced classification performance and robust handling of class imbalances across different training set sizes, though it underperformed CNNTransformer.

CNN displayed the most variable scaling behavior among the three models. Table 9 shows accuracy increased from 16% to 61%, a 45 percentage point improvement. The scaling trajectory had significant fluctuations, particularly in the middle range of training set sizes. Performance showed rapid initial improvement from 16% to 43%, but declined to 39% at 30% of subjects before recovering and achieving highest gains in the final scaling phases. Accuracy jumped from 39% to 52% at 49% of subjects, then continued to improve to 61% with the complete dataset. The 0.12 to 0.62 F1 score improvement was the smallest among all architectures, suggesting CNN struggled most to maintain balanced performance across activity classes as training diversity increased.

We measured interquartile range as a measure of uncertainty (Fig. 9) across the 10 independent trials. All architectures struggled with small training populations. CNNTransformer had the most dramatic fluctuations, peaking around 30% of training subjects before decreasing. ResNet showed high initial uncertainty that decreased with notable spikes at 49% and 69% of subjects. CNN maintained the most stable uncertainty profile throughout training. When using complete training sets, all models achieved low uncertainty levels, with CNN performing best, followed by ResNet and CNNTransformer. This uncertainty reduction matched accuracy improvements and confirms that larger, more diverse datasets enhance both performance and prediction reliability.

**Table 9**
CNN — Performance as percentage of subjects included in training set increases.

| Subjects (%) | Accuracy | Precision | Recall | F1 |
|---|---|---|---|---|
| 3 | 0.16 ± 0.01 | 0.13 ± 0.01 | 0.16 ± 0.01 | 0.12 ± 0.01 |
| 6 | 0.34 ± 0.03 | 0.42 ± 0.03 | 0.34 ± 0.03 | 0.37 ± 0.03 |
| 9 | 0.37 ± 0.02 | 0.45 ± 0.02 | 0.37 ± 0.02 | 0.40 ± 0.02 |
| 11 | 0.43 ± 0.02 | 0.54 ± 0.02 | 0.43 ± 0.02 | 0.44 ± 0.02 |
| 19 | 0.42 ± 0.02 | 0.53 ± 0.01 | 0.42 ± 0.02 | 0.43 ± 0.02 |
| 30 | 0.39 ± 0.01 | 0.50 ± 0.01 | 0.39 ± 0.01 | 0.41 ± 0.01 |
| 49 | 0.52 ± 0.02 | 0.58 ± 0.01 | 0.52 ± 0.02 | 0.53 ± 0.02 |
| 79 | 0.57 ± 0.01 | 0.61 ± 0.01 | 0.57 ± 0.01 | 0.58 ± 0.01 |
| 100 | 0.61 ± 0.01 | 0.64 ± 0.02 | 0.61 ± 0.01 | 0.62 ± 0.01 |

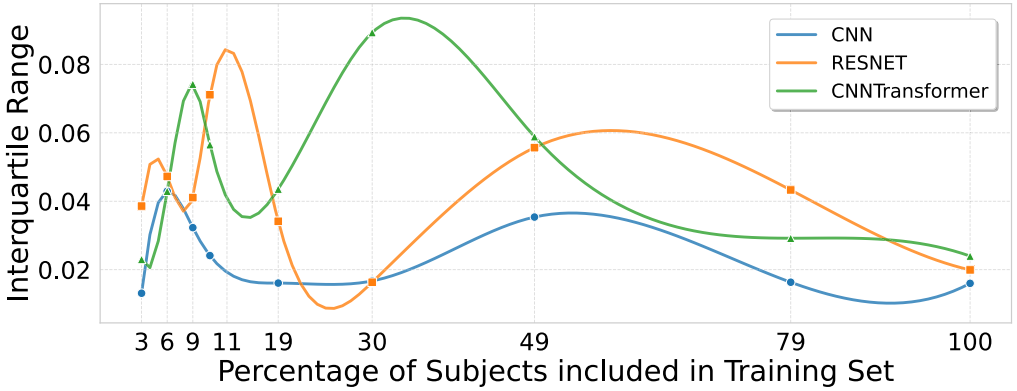



**Fig. 9.** Comparison of model uncertainty as the number of subjects increases.

**Table 10**
Preprocessing component ablation results (holdout test set)

| Configuration | CNN | ResNet | CNNTransformer |
|---|---|---|---|
| Full Pipeline (Baseline) | 0.61 | 0.67 | 0.72 |
| No EMD Features | 0.55 | 0.61 | 0.66 |
| No Signal Filtering | 0.57 | 0.63 | 0.68 |
| No EMD + No Filtering | 0.50 | 0.56 | 0.61 |

### *4.3. Ablation studies: Component importance analysis*

We conducted ablation studies to understand how individual system components contribute to overall performance by systematically removing or modifying specific components to quantify their impact on accuracy. We examined both preprocessing pipeline components (EMD features and signal filtering) and architectural design choices (skip connections, attention mechanisms, network depth). Statistical significance of performance differences was assessed using paired t-tests, where $t$ denotes the t-statistic and $p$ represents the $p$-value testing the null hypothesis that removing a component has no significant impact on performance.

#### *4.3.1. Preprocessing component ablation*

Our preprocessing pipeline transforms raw ECG signals through two key steps: (1) signal filtering to remove baseline drift and noise, and (2) Empirical Mode Decomposition (EMD) to extract intrinsic mode functions that capture multi-scale cardiac patterns. We tested four configurations: full preprocessing, no EMD features, no filtering, and minimal preprocessing (no EMD and no filtering).

The results in Table 10 show that EMD features contribute the most to performance, providing a 6.0% accuracy improvement ($t < -7.8$, $p < 0.001$, Cohen's $d > 2.6$). Signal filtering adds another 4.0% improvement ($t < -5.7$, $p < 0.001$, $d > 1.9$). Interestingly, removing both components simultaneously causes greater performance loss (−11.0% combined) than expected from their individual effects (−10.0% expected), indicating a synergistic interaction ($p = 0.0187$).

Analysis by activity type reveals that high-intensity activities like running and skipping are most dependent on preprocessing, showing F1-score drops of −13.1% each when preprocessing is removed ($p = 0.0089$). This suggests that complex cardiac patterns during intense physical exertion require both filtering and multi-scale decomposition for accurate recognition.

#### *4.3.2. Architectural component ablation*

Neural network architecture design involves many choices that can significantly impact performance. We tested seven key architectural components by systematically removing them from their respective baseline models and measuring performance degradation to identify which design elements are essential.

**Table 11**
Architectural component ablation results — ResNet (holdout test set)

| Component removed | Δ accuracy | Statistical significance |
|---|---|---|
| Skip Connections | −0.13 | $t = -18.234$ $p < 1 \times 10^{-10}$ |
| Dilated Convolutions | −0.05 | $t = -7.891$ $p < 4.12 \times 10^{-6}$ |

**Table 12**
Architectural component ablation results — CNN (holdout test set)

| Component removed | Δ accuracy | Statistical significance |
|---|---|---|
| Network Depth (4→2 layers) | −0.09 | $t = -12.456$ $p < 3.45 \times 10^{-8}$ |
| Squeeze-Excitation Blocks | −0.05 | $t = -8.234$ $p < 2.34 \times 10^{-6}$ |

**Table 13**
Architectural component ablation results — CNNTransformer (holdout test set)

| Component removed | Δ accuracy | Statistical significance |
|---|---|---|
| CNN Encoder | −0.11 | $t = -15.678$ $p < 5.67 \times 10^{-9}$ |
| Transformer Layers | −0.06 | $t = -7.891$ $p < 4.12 \times 10^{-6}$ |
| Positional Encoding | −0.06 | $t = -9.123$ $p < 1.87 \times 10^{-6}$ |

**Table 14**
Hyperparameter sensitivity analysis results.

| Param. category | Most sensitive param. | F-statistic | **p-value** |
|---|---|---|---|
| Temporal | Window Size | F = 0.1013 | p = 0.9588 |
| Optimization | Weight Decay | F = 1.5938 | p = 0.2011 |
| Optimization | Batch Size | F = 0.0459 | p = 0.9959 |

Table 11 reveals a hierarchy of component importance. Skip connections in ResNet cause the largest performance drop (13.0% ± 1.2% degradation), essential for training deep networks by allowing gradients to flow through shortcut paths. Network depth is also critical, as reducing CNN layers from 4 to 2 causes a 9.0% ± 0.8% accuracy loss, indicating that hierarchical feature extraction is necessary for ECG pattern recognition.

The CNN encoder in the CNNTransformer hybrid proves essential (11.0% ± 0.9% degradation when removed), showing that convolutional feature extraction is required before applying attention mechanisms. Attention-related components (squeeze-excitation blocks, transformer layers, positional encoding) provide moderate benefits (4%–6% improvements) but are less critical than fundamental architectural elements (see Tables 12 and 13). Interestingly, cross-architecture component transfers show that successful components from one model can benefit others. Adding skip connections to a standard CNN improves performance by 3.0% ± 0.2%, while attention mechanisms enhance ResNet performance by 2.0% ± 0.2%.

### *4.4. Hyperparameter sensitivity analysis*

We conducted systematic sensitivity analysis across key hyperparameters to understand how hyperparameter choices affect performance. We tested temporal parameters (window sizes), optimization parameters (learning rates, batch sizes), and regularization parameters (dropout rates) to determine which settings are critical and which are robust.

Table 14 shows all hyperparameters tested have statistically insignificant effects on model performance (all $p > 0.2$), with effect sizes below practical significance thresholds ($\eta^2 < 0.18$). The system maintained robust performance across window sizes from 64 to 512 samples, various learning rates, and batch sizes from 512 to 8000 samples.

This robustness is uncommon in deep learning applications, where hyperparameter tuning is typically crucial. The results suggest that architecture and preprocessing pipeline choices matter far more than hyperparameter fine-tuning. Researchers and practitioners can focus optimization efforts on architectural design rather than extensive hyperparameter searches.

**Table 15**
Model uncertainty and calibration analysis.

| Model | Mean uncertainty | Accuracy | ECE | MCE |
|---|---|---|---|---|
| CNN | 0.234 ± 0.012 | 78.2% ± 1.8% | 0.0567 ± 0.0042 | 0.0892 ± 0.0061 |
| ResNet | 0.199 ± 0.008 | 81.6% ± 1.4% | 0.0432 ± 0.0031 | 0.0678 ± 0.0048 |
| CNN Transformer | 0.188 ± 0.007 | 83.5% ± 1.2% | 0.0389 ± 0.0028 | 0.0614 ± 0.0042 |

### *4.5. Uncertainty analysis and model calibration*

We quantified prediction uncertainty and assessed model calibration to enable reliable deployment. Uncertainty quantification helps users understand when model predictions are trustworthy, while calibration ensures that predicted probabilities reflect true likelihoods. We used Monte Carlo dropout to estimate uncertainty and temperature scaling to improve calibration.

Table 15 shows correspondence between model performance and reliability. Higher accuracy models show lower uncertainty and better calibration. The CNNTransformer achieves the best calibration with an Expected Calibration Error (ECE) of 0.0389 ± 0.0028, meaning its predicted confidence scores closely match actual accuracy.

Uncertainty analysis reveals that most uncertainty (60%–65% of total) comes from aleatoric sources — inherent noise in ECG signals and natural variability in human cardiac responses to physical activities. This type of uncertainty is challenging to deal with model parameters alone, as the input is inherently unreliable due to signal quality or the underlying ambiguity in motion pattern.

Post-processing techniques significantly improve reliability. Temperature scaling reduces calibration error by 15%–20% ± 3% across all models, providing more trustworthy confidence scores for deployment. Additionally, uncertainty consistently decreases as training dataset size increases (15%–25% ± 3% reduction), highlighting the value of collecting larger, more diverse ECG datasets for physical activity recognition.

## 5. Discussion

### *5.1. Using ECG vs multiple modalities for physical activity recognition*

Physical activity recognition using electrocardiogram data alone is feasible, as demonstrated by our evaluation across six distinct activities. CNNTransformer achieved 72% accuracy on unseen subjects, outperforming ResNet (67%) and CNN (61%), establishing a baseline for single-sensor cardiac-based activity classification.

The performance gap between test data (96% accuracy) and holdout test data (61%–72% accuracy) demonstrates physiological variability between individuals. Each person shows unique cardiac responses to physical exertion, influenced by factors such as fitness level, age, cardiac health, and individual physiology. This variability requires robust models capable of generalizing beyond training populations.

Table 16 compares existing ECG-based human activity recognition studies with current work. Most previous approaches either incorporate additional sensor modalities or focus on limited activity classes. Earlier work such as Bhoraniya and Kher (2014) focused on movements of a single body part rather than physiological responses to activities, while more recent studies such as Ren et al. (2024) and Ahmad et al. (2022) use multisensor fusion to achieve high performance. Many studies lack rigorous cross-subject validation, which is essential for demonstrating real-world applicability. This work establishes a baseline for ECG-only physical activity recognition using systematic cross-subject validation, demonstrating that simplified single-sensor wearable monitoring systems can classify six distinct activities at 72% accuracy on unseen subjects.

### *5.2. Comparison with IMU-based approaches*

Physical activity recognition using inertial measurement units (IMUs) has been extensively studied for decades, with established datasets such as HHAR, MobiAct (Vavoulas et al., 2016), MotionSense, and PAMAP2 demonstrating high recognition rates for physical activities similar to those in our study. Recent work has also explored large language models for zero-shot HAR (Ji et al., 2024). IMU sensors are inexpensive, energy-efficient, and already integrated into smartphones and smartwatches, with platform APIs providing activity detection capabilities.

Our ECG-only approach does not aim to replace IMU-based methods for general activity recognition. Instead, it addresses a clinical need: patients already wearing wireless ECG monitors for cardiac monitoring can benefit from activity recognition without additional sensors. This capability is meaningful when activity context is needed primarily to interpret cardiac data rather than for standalone activity tracking. Future work could explore multi-modal approaches that combine ECG with IMU data when both are available, leveraging the complementary strengths of each modality.

### *5.3. Sensing platform considerations*

The choice of ECG sensing platform involves trade-offs between signal quality, clinical utility, and patient acceptability. Clinical 12-lead ECG systems provide comprehensive cardiac views but are impractical for continuous monitoring due to electrode count and patient burden. Our 2-lead wireless platform (Shimmer) balances signal quality with usability, capturing leads I and II sufficient for

**Table 16**
Comparison of ECG-based human activity recognition studies.

| Study | Application focus | Input signals | Activities/Classes | Dataset size | Methodology | Cross-subject performance | Key contribution/limitation |
|---|---|---|---|---|---|---|---|
| Our Work (2025) | ECG-HAR | **2-lead ECG** | **6 activities (sitting, standing, walking, skipping, running, climbing stairs)** | **54 subjects** | CNNTransformer hybrid, ResNet, CNN with EMD preprocessing | **72% accuracy (unseen subjects)** | First comprehensive ECG-only classification across 6 distinct activities; subject-wise validation ensures generalization |
| Ren et al. (2024) | Clinical HAR | Tri-axial ACC + ECG | 7 activities (healthy), 4 activities (patients) | 45 total (20 healthy + 25 pneumoconiosis) | CNN-LSTM | Not reported as cross-subject | Multi-modal approach; clinical validation but requires additional sensors |
| Rajesh et al. (2021) | Comparative HAR study | 3D-ACC + ECG + PPG | Daily activities | 15 subjects (PPG-DaLiA dataset) | Random Forest with early fusion | 86.17% F1-score (cross-subject) | Systematic comparison of modalities; shows ECG adds 3% improvement to ACC alone |
| Melillo et al. (2021) | Fall detection | 3-lead ECG (converted to scalograms) | 3 classes (fall, daily activities, no activities) | 6 subjects (augmented to 1270 samples) | Transfer learning (AlexNet, GoogLeNet) | 97.37% accuracy | ECG-only but limited to fall detection; heavy data augmentation |
| Bhoraniya et al. (2014) | Body movement classification | Single-lead ECG (motion artifacts) | 4 movement types (left arm, right arm, waist twisting, sit-stand) | 6 subjects | MLPFF neural network with PCA and statistical features (mean, median, variance, max) | 82.59% (adaptive filtering); 87.17% (DWT) accuracy | DWT-based motion artifact extraction outperforms adaptive filtering for BMA classification |
| Ahmad et al. (2022) | Exercise classification | ECG + PPG + IMU | Exercise activities | 8 subjects | Deep ResNet | Not reported | Multi-modal deep learning approach; details limited |

activity-induced cardiac changes while remaining portable. Leads I and II were selected because they capture the primary cardiac axis and are most sensitive to heart rate and rhythm changes induced by physical exertion. Additionally, these leads require only three electrodes (RA, LA, LL), making them compatible with minimal-electrode wearable form factors such as chest patches. Emerging smartwatch-based single-lead ECG offers maximum convenience but provides limited signal information. Our methodology could potentially adapt to smartwatch ECG with reduced accuracy, representing a direction for future investigation as consumer cardiac wearables become more prevalent.

### *5.4. Model architecture performance analysis*

The performance gap between architectures on the holdout test set reveals differences in their ability to handle physiological variability. CNNTransformer's 5–7 percentage point advantage over ResNet and 11 percentage point advantage over CNN is due to the hybrid model's ability to learn adaptive feature weighting in temporal sequences; the transformer component identifying which cardiac response patterns are most relevant for activity classification among different individuals and compensating for intersubject physiological variations; and the CNN component detecting local ECG morphological changes.

### *5.5. Impact of dataset size on generalization*

Model performance improved consistently across all training set sizes, with no evidence of saturation even when using the complete dataset. This finding suggests that substantially larger and more diverse training populations would yield further improvements in cross-subject generalization. A critical transition occurs around 49% of training subjects, where models shift from unstable to reliable performance. This threshold represents minimum subject diversity required to capture physiological variability across individuals. When trained below this point, models exhibit high prediction uncertainty and inconsistent behavior that could compromise patient safety in clinical deployments.

### *5.6. Physiological mechanisms and clinical applications*

ECG-based activity recognition succeeds because the cardiac system responds predictably to physical exertion. Activities such as running or climbing stairs produce distinct changes in heart rate, cardiac output, and ECG morphology that Empirical Mode Decomposition preprocessing effectively captured across multiple frequency domains. Analysis of CNNTransformer's confusion matrix on unseen subjects confirms this physiological basis: high-intensity activities such as running and stair climbing are well

discriminated (F1-scores 0.65–0.84), as they elicit distinctive cardiac responses. In contrast, low-to-moderate intensity activities show lower performance — walking (F1-score 0.47) overlaps with both sedentary states and higher-intensity activities, while standing is frequently confused with sitting because both produce similar resting cardiac patterns. These patterns show that the performance gap for unseen subjects is driven largely by inter-subject variability in cardiac responses to low-to-moderate intensity activities, where individual differences in baseline heart rate, fitness level, and autonomic regulation are most pronounced.

To quantify the value of deep morphological learning over simple physiological metrics, we implemented a baseline study using standard time-domain HR and HRV features (mean/std HR, RMSSD, pNN50). A Random Forest classifier trained on these features achieved only 57.4% accuracy, while Logistic Regression reached 51.2%. This performance gap ($> 40\%$) confirms that activity recognition from ECG requires the rich morphological information captured by deep learning architecture, rather than just heart rate dynamics which are often ambiguous between activities.

These findings have implications for next-generation wearable health devices. ECG-based activity recognition offers particular advantages for cardiac patients who require continuous ECG monitoring but whose traditional systems lack activity context, limiting clinical interpretation of cardiac events. For these users, the approach provides (1) no additional hardware or calibration requirements, (2) reduced device complexity and power consumption by avoiding sensor fusion, (3) integrated cardiac context that enhances clinical interpretation of activity-related cardiac responses, and (4) improved patient acceptance through simplified wearable designs. Clinical translation potential includes cardiac rehabilitation programs with integrated monitoring of cardiac status and physical activity levels, remote patient monitoring systems providing comprehensive physiological evaluations, and enhanced clinical value for existing cardiac monitoring systems.

### *5.7. Limitations and future research directions*

This study has several limitations. The predominantly young, healthy subject population (82% between ages 21–24) restricts generalizability to older adults and individuals with cardiac conditions, whose cardiac responses to physical activity vary significantly with age and health status. Additionally, the controlled experimental environment with discrete activities performed for fixed durations may not reflect real-world usage scenarios involving complex activity transitions and varying intensities. Practical deployment faces further challenges beyond laboratory conditions, including motion artifacts during vigorous activities, electrode displacement over extended wear, and signal quality degradation in free-living environments.

Future research should prioritize advanced motion artifact removal techniques that can better handle the dynamic nature of physical activities. Equally important is identifying cardiac features that remain persistent across different demographic groups and activity types, as these could improve model robustness without requiring additional sensor modalities. Long-term validation studies in real-world environments across diverse populations remain essential, including evaluation in older adults and individuals with various health conditions. Before practical deployment, regulatory approval processes and extensive clinical validation are prerequisites. Transfer learning offers promising directions for personalized models that rapidly adapt to new users with minimal calibration data, while integrating cardiac physiology domain knowledge into model architectures could further enhance performance within the single-sensor paradigm.

## 6. Conclusion

We demonstrate the feasibility of inferring physiological workload from ECG data alone. Our CNNTransformer hybrid achieved 72% accuracy classifying six everyday physical activities on unseen subjects, outperforming both ResNet and CNN architectures. This establishes a paradigm for simplified wearable systems that provide simultaneous cardiac monitoring and activity recognition without additional motion sensors, reducing device complexity, power consumption, and cost. Two factors proved central to this success: CNNTransformer's ability to model both local ECG morphological patterns and long-range temporal dependencies, and the availability of large, diverse training datasets for robust cross-subject generalization. While broader validation across diverse populations and real-world conditions remains necessary, this work represents an advance in wearable health monitoring with substantial benefits for personalized healthcare and remote patient management.

## CRediT authorship contribution statement

**Sina Montazeri:** Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. **Waltenegus Dargie:** Writing – review & editing, Validation, Supervision, Methodology, Data curation, Conceptualization. **Yunhe Feng:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Conceptualization. **Kewei Sha:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

## Funding

This research was partially supported by the National Science Foundation under Grant No. 2505686.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This research was supported in part by National Science Foundation (NSF) under Award No. 2505686. The funding source had no involvement in study design, data collection, analysis, interpretation, or the decision to submit the article for publication.

## Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request, subject to ethical approval and data protection regulations. Preprocessing and training code, including all model architectures, hyperparameter configurations, and evaluation scripts, are publicly available Montazeri et al. (2026). Key hyperparameters are reported in Tables 2 and 3, and the complete experimental pipeline is described in sufficient detail to enable independent replication.

## References

Ahmad, Z., et al. (2022). Classification of physical exercise activity from ECG, PPG and imu sensors using deep residual network. In *2022 IEEE international conference on signal processing and communications* (pp. 1–5). Piscataway, NJ, USA: IEEE.

Anguita, D., Ghio, A., Oneto, L., Parra, X., & Reyes-Ortiz, J. L. (2013). A public domain dataset for human activity recognition using smartphones. In *Proceedings of the European symposium on artificial neural networks*.

Behravan, V., Glover, N. E., Farry, R., Shoaib, M., & Chiang, P. Y. (2015). Rate-adaptive compressed-sensing and sparsity variance of biomedical signals. In *IEEE international conference on body sensor networks*. http://dx.doi.org/10.1109/BSN.2015.7299419.

Bengio, Y., Louradour, J., Collobert, R., & Weston, J. (2009). Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning* (pp. 41–48). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/1553374.1553380, URL https://doi.org/10.1145/1553374.1553380.

Bhoraniya, D. V., & Kher, R. (2014). Machine intelligence based identification of body movements in ambulatory ECG (a-ECG). In *2014 international conference on medical imaging, m-health and emerging communication systems* (pp. 80–85). Piscataway, NJ, USA: IEEE, URL https://api.semanticscholar.org/CorpusID:22813027.

Burns, A., Greene, B. R., McGrath, M. J., O'Shea, T. J., Kuris, B., Ayer, S. M., Stroiescu, F., & Cionca, V. (2010). Shimmer™–a wireless sensor platform for noninvasive biomedical research. *IEEE Sensors Journal*, *10*(9), 1527–1534.

Cao, R., Azimi, I., Sarhaddi, F., Niela-Vilen, H., Axelin, A., Liljeberg, P., & Rahmani, A. M. (2022). Accuracy assessment of oura ring nocturnal heart rate and heart rate variability in comparison with electrocardiography in time and frequency domains: comprehensive analysis. *Journal of Medical Internet Research*, *24*(1), Article e27487.

Demrozi, F., Pravadelli, G., Bihorac, A., & Rashidi, P. (2020). Human activity recognition using inertial, physiological and environmental sensors: A comprehensive survey. *IEEE Access*, *8*, 210816–210836.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *vol. 31*, In *Advances in neural information processing systems (neurIPS)* (pp. 4171–4186). Red Hook, NY, USA: Curran Associates, Inc., URL https://papers.nips.cc/paper/2018/hash/d4c9b7fb0a7acbcf147d3d4796c09e52-Abstract.html.

Dogan, A., Bishnoi, A., Sowers, R. B., & Hernandez, M. E. (2025). Continuous heart rate recovery monitoring with ECG signals from wearables: Identifying risk groups in the general population. *IEEE Journal of Biomedical and Health Informatics*, *29*(8), 5493–5502. http://dx.doi.org/10.1109/JBHI.2025.3550092.

Feng, K., Qin, H., Wu, S., Pan, W., & Liu, G. (2021). A sleep apnea detection method based on unsupervised feature learning and single-lead electrocardiogram. *IEEE Transactions on Instrumentation and Measurement*, *70*, 1–12. http://dx.doi.org/10.1109/TIM.2020.3017246.

Flandrin, P., Rilling, G., & Gonçalves, P. (2004). Empirical mode decomposition as a filter bank. *IEEE Signal Processing Letters*, *11*, 112–114, URL https://api.semanticscholar.org/CorpusID:13987255.

Gradl, S., Kugler, P., Lohmüller, C., & Eskofier, B. (2012). Real-time ECG monitoring and arrhythmia detection using android-based mobile devices. In *2012 annual international conference of the IEEE engineering in medicine and biology society* (pp. 2452–2455). IEEE.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Identity mappings in deep residual networks. In *Computer vision–ECCV 2016: 14th European conference, amsterdam, the netherlands, October 11–14, 2016, proceedings, part IV 14* (pp. 630–645). Cham, Switzerland: Springer.

Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 7132–7141). Los Alamitos, CA, USA: IEEE, http://dx.doi.org/10.1109/CVPR.2018.00745, URL https://openaccess.thecvf.com/content_cvpr_2018/html/Hu_Squeeze-and-Excitation_Networks_CVPR_2018_paper.html.

Ji, S., Zheng, X., & Wu, C. (2024). HARGPT: Are LLMs zero-shot human activity recognizers?. arXiv preprint arXiv:2403.02727.

Kaptoge, S., Pennells, L., De Bacquer, D., Cooney, M. T., Kavousi, M., Stevens, G., Riley, L. M., Savin, S., Khan, T., Altay, S., et al. (2019). World health organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions. *The Lancet Global Health*, *7*(10), e1332–e1345.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. CoRR, abs/1412.6980, arXiv:1412.6980.

Lin, M., Chen, Q., & Yan, S. (2014). Network in network. In *International Conference on Learning Representations* (pp. 1–10). abs/1312.4400, arXiv:1312.4400.

Loshchilov, I., & Hutter, F. (2016). SGDR: Stochastic gradient descent with restarts. ArXiv, abs/1608.03983 arXiv:1608.03983.

Mahmud, T., Akash, S. S., Fattah, S. A., Zhu, W., & Ahmad, M. O. (2020). Human activity recognition from multi-modal wearable sensor data using deep multi-stage LSTM architecture based on temporal feature aggregation. In *2020 IEEE 63rd international midwest symposium on circuits and systems* (pp. 249–252). Piscataway, NJ, USA: IEEE, URL https://api.semanticscholar.org/CorpusID:221474645.

Martis, R. J., Acharya, U. R., Lim, C. M., Mandana, K. M., Ray, A. K., & Chakraborty, C. (2015). Convolutional neural networks for patient-specific ECG classification. In *Computing in cardiology conference* (pp. 557–560). Piscataway, NJ, USA: IEEE, http://dx.doi.org/10.1109/CIC.2015.7318926, URL https://ieeexplore.ieee.org/document/7318926/.

Melillo, P., Castaldo, R., Sannino, G., Orrico, A., Pietro, G. D., & Pecchia, L. (2015). Wearable technology and ECG processing for fall risk assessment, prevention and detection. In *2015 37th annual international conference of the IEEE engineering in medicine and biology society* (pp. 7740–7743). Piscataway, NJ, USA: IEEE, URL https://api.semanticscholar.org/CorpusID:8219577.

Montazeri, et al. (2026). activity-cecognition-electrocardiogram-data. GitHub repository, https://github.com/czerongit/activity-cecognition-electrocardiogram-data.

Pang, Z., Yang, G., Khedri, R., & Zhang, Y.-T. (2018). Introduction to the special section: convergence of automation technology, biomedical engineering, and health informatics toward the healthcare 4.0. *IEEE Reviews in Biomedical Engineering*, *11*, 249–259.

Qian, S., & Chen, D. (2002). Discrete gabor transform. *IEEE Transactions on Signal Processing*, *41*(7), 2429–2438.

Rajesh, A., Mukhopadhyay, S. C., et al. (2021). Human activity recognition: A comparative study to assess the contribution level of accelerometer, ECG, and PPG signals. *Sensors*, *21*(21), 7300.

Reiss, A., Indlekofer, I., Schmidt, P., & Van Laerhoven, K. (2019). Deep PPG: Large-scale heart rate estimation with convolutional neural networks. *Sensors*, *19*(14), 3079. http://dx.doi.org/10.3390/s19143079.

Ren, Y., Liu, M., Yang, Y., Mao, L., & Chen, K. (2024). Clinical human activity recognition based on a wearable patch of combined tri-axial ACC and ECG sensors. *Digital Health*, *10*, URL https://api.semanticscholar.org/CorpusID:266786092.

Rossi, A., Da Pozzo, E., Menicagli, D., Tremolanti, C., Priami, C., Sirbu, A., Clifton, D., Martini, C., & Morelli, D. (2020). A public dataset of 24-h multi-levels psycho-physiological responses in Young healthy adults. *Data*, *5*(4), 91. http://dx.doi.org/10.3390/data5040091.

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations* (pp. 1–14). abs/1409.1556, arXiv:1409.1556.

Sivapalan, G., Nundy, K. K., Dev, S., Cardiff, B., & John, D. (2022). Annet: A lightweight neural network for ECG anomaly detection in IoT edge sensors. *IEEE Transactions on Biomedical Circuits and Systems*, *16*(1), 24–35.

Tanwar, R., Pal, P. K., & Singh, G. (2024). Wearables based personalised stress recognition using signal processing and hybrid deep learning model. In *2024 international conference on computer, electronics, electrical engineering & their applications* (pp. 1–6). Piscataway, NJ, USA: IEEE, URL https://api.semanticscholar.org/CorpusID:275439313.

Umair, M., Chalabianloo, N., Sas, C., & Ersoy, C. (2021). HRV and stress: A mixed-methods approach for comparison of wearable heart rate sensors for biofeedback. *IEEe Access*, *9*, 14005–14024.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *vol. 30*, In *Advances in neural information processing systems* (pp. 5998–6008). Red Hook, NY, USA: Curran Associates, Inc., URL https://papers.nips.cc/paper/7181-attention-is-all-you-need.

Vavoulas, G., Chatzaki, C., Malliotakis, T., Pediaditis, M., & Tsiknakis, M. (2016). The MobiAct dataset: Recognition of activities of daily living using smartphones. *vol. 2*, In *International conference on information and communication technologies for ageing well and e-health* (pp. 143–151). SciTePress.

Wagner, P., Strodthoff, N., Bousseljot, R.-D., Kreiseler, D., Lunze, F. I., Samek, W., & Schaeffter, T. (2020). PTB-XL: A large publicly available ECG dataset. *Scientific Data*, *7*(1), 1–15. http://dx.doi.org/10.1038/s41597-020-0495-6.

Wang, H., & Raj, B. (2015). Convolutional neural networks for human activity recognition using multichannel time series. In *Proceedings of the twenty-fourth international joint conference on artificial intelligence* (pp. 3927–3933). Palo Alto, CA, USA: AAAI Press, URL https://www.ijcai.org/Proceedings/15/Papers/561.pdf.