

Article

Domain Adaptation for Sensor-Based Human Activity Recognition with a Graph Convolutional Network

Jing Yang ^{1,†} , Tianzheng Liao ^{2,†}, Jingjing Zhao ³, Yan Yan ⁴ , Yichun Huang ⁵, Zhijia Zhao ¹, Jing Xiong ^{4,*} and Changhong Liu ^{1,*}

¹ School of Mechanical and Electrical Engineering, Guangzhou University, Guangzhou 510006, China

² Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou 519041, China

³ School of Biological Science and Medical Engineering, Southeast University, Nanjing 211189, China

⁴ Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

⁵ School of Mechatronic Engineering and Automation, Foshan University, Foshan 528010, China

* Correspondence: jing.xiong@siat.ac.cn (J.X.); lch@gzhu.edu.cn (C.L.)

† These authors contributed equally to this work.

Abstract: Sensor-based human activity recognition (HAR) plays a fundamental role in various mobile application scenarios, but the model performance of HAR heavily relies on the richness of the dataset and the completeness of data annotation. To address the shortage of comprehensive activity types in collected datasets, we adopt the domain adaptation technique with a graph neural network-based approach by incorporating an adaptive learning mechanism to enhance the action recognition model's generalization ability, especially when faced with limited sample sizes. To evaluate the effectiveness of our proposed approach, we conducted experiments using three well-known datasets: MHealth, PAMAP2, and TNDA. The experimental results demonstrate the efficacy of our approach in sensor-based HAR tasks, achieving impressive average accuracies of 98.88%, 98.58%, and 97.78% based on the respective datasets. Furthermore, we conducted transfer learning experiments to address the domain adaptation problem. These experiments revealed that our proposed model exhibits exceptional transferability and distinguishing ability, even in scenarios with limited available samples. Thus, our approach offers a practical and viable solution for sensor-based HAR tasks.



Citation: Yang, J.; Liao, T.; Zhao, J.; Yan, Y.; Huang, Y.; Zhao, Z.; Xiong, J.; Liu, C. Domain Adaptation for Sensor-Based Human Activity Recognition with a Graph Convolutional Network. *Mathematics* **2024**, *12*, 556. <https://doi.org/10.3390/math12040556>

Academic Editor: Marjan Mernik

Received: 2 December 2023

Revised: 30 January 2024

Accepted: 5 February 2024

Published: 12 February 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: sensor-based human activity recognition (HAR); graph convolutional network; domain adaptation; transfer learning

MSC: 68T05; 68T07

1. Introduction

Human activity recognition (HAR) has emerged as a captivating research area with broad implications. HAR involves the classification of human activities by processing and learning information obtained from diverse sources such as cameras, sensors, radar, WiFi signals, and more. These sources capture human movements, encompassing activities like walking, running, ascending stairs, falling, sitting still, and standing up. The continuous advancement of technology and the improvement of living standards have fostered the expansion of application scenarios in the field of human motion recognition. Notably, HAR has found significance in domains including medical rehabilitation, physical games, and sports competitions. Sensor-based HAR, in particular, plays a pivotal role in this field. Compared to alternative methods of acquiring human motion signals, sensors offer several notable advantages. Sensors are less prone to environmental interference, ensuring the collection of continuous and accurate signals. This characteristic makes sensor-based HAR applicable in a wide range of scenarios, thereby enhancing the effectiveness and reliability of activity recognition. However, the process of acquiring human motion sensor signals

from multiple points simultaneously presents challenges. It is a time-consuming and laborious endeavor, contributing to the pervasive issue of limited sample availability in the field of sensor-based HAR. Collecting a substantial number of samples for network training purposes is arduous, with only a few publicly available datasets currently accessible for learning and research endeavors. Moreover, the advancement of augmented reality (AR), virtual reality (VR), mixed reality (MR), and related technologies has sparked growing interest in the field of recognizing human behaviors using wearable robotics and wearable devices. At present, the realm of human motion recognition using wearable robots faces a substantial obstacle concerning the attainment of precise cross-domain recognition outcomes for motion data characterized by diverse formats or representations. Consequently, this predicament imposes a rigorous requirement on the quality of motion input data for wearable robots. The efficacy of HAR is contingent upon the quality and quantity of the training data. Nonetheless, the process of data annotation presents a considerable challenge owing to its demanding and time-consuming nature. Hence, current scholarly investigations are predominantly concentrated on exploring methodologies that effectively exploit existing data and expand its applicability across diverse domains or analogous disciplines. In [1], a methodology was proposed to tackle the challenge of insufficient data by reutilizing pre-existing user behavior models from websites in different domains. This approach involves the selection of diverse datasets from websites spanning various domains, followed by the utilization of transfer learning techniques to establish user behavior models by leveraging the data from these domains. On the other hand, [2] introduced the Dimension-Adaptive Neural Architecture (DANA), a novel solution that empowers deep neural networks to dynamically adapt to changes in input data dimensions. This adaptation capability allows DANA to effectively address challenges such as sensor availability and adaptive sampling during inference. Notably, DANA offers a unified training model that achieves and sustains high classification accuracy across a wide range of practical scenarios. Consequently, the requirement for individual classifiers for each specific scenario is obviated. The advancement of smartphones has significantly facilitated the continuous acquisition of human activity data. By virtue of their non-invasive nature and the incorporation of diverse built-in sensors, smartphones enable versatile and precise monitoring of bodily movements [3]. Presently, extensive research is being conducted on HAR methods that are based on smartphone technology. This research holds substantial significance across diverse application domains, encompassing areas such as fitness tracking, health management, and beyond [4]. Within the domain of human activity recognition (HAR), [5] presented a novel approach that is based on random projection (RP) for smartphone-based HAR. This approach employs the Subsampled Randomized Hadamard Transformation (SRHT) technique to accomplish both dimensionality reduction and classification tasks, thereby effectively augmenting the classifier's capacity for generalization. Reference [6] introduced an innovative framework known as CAEL-HAR, which amalgamates convolutional neural networks (CNNs), Autoencoder models, and long short-term memory (LSTM) networks. By leveraging the complementary characteristics of these three constituents, CAEL-HAR facilitates the recognition of human activities based on motion data acquired from smartphones. Reference [7] further enhanced the efficacy of activity monitoring by integrating attention mechanisms into multi-head CNNs, thereby enabling refined feature extraction and selection processes within the LSTM network. Reference [8], through a comparative analysis of four hybrid architectures, established a pioneering two-dimensional CNN-LSTM hybrid architecture with the capability to extract a wider range of data features, resulting in an improved accuracy of activity recognition. In the domain of multi-modal wearable sensor data, Mahmud et al. [9] proposed a comprehensive global feature optimizer network, comprising multiple LSTM layers and a sequence of dense connection layers. This network seamlessly integrates multi-modal features to extract global features, effectively harnessing feature information across diverse temporal scales and significantly enhancing the accuracy of activity recognition.

Convolutional neural networks (CNNs) are widely recognized for their efficacy in extracting features from well-structured matrix data. However, CNNs face limitations when processing non-Euclidean structured data, as they are unable to directly perform convolution operations on such data using convolution kernels of the same size. This raises an important question: how can we efficiently extract spatial features for machine learning on complex data structures like topological graphs? In the domain of human activity recognition (HAR), the interdependence of motion information across different body parts exhibits variations. The data acquired from various joints of the human body can be seen as graph-structured data residing in a non-Euclidean space. To address this challenge, graph convolutional neural networks (GCNNs) integrate the strengths of CNNs and graph theory, enabling the establishment of neural networks that effectively process graph-structured data and overcome the intricacies of non-Euclidean spaces. In order to capture the interrelationships among sensor information originating from diverse channels more effectively, the present study employs the Chebyshev Graph Convolutional Network (ChebNet) for feature extraction from human motion data.

To address the aforementioned issues, we present a novel approach called the Graph Domain Adaptation (GDA) network. In contrast to conventional deep models, our GDA network adeptly captures the non-Euclidean graph structure relationships inherent in sensor signals. To mitigate the issue of over-smoothing in multi-layer graph convolution, we incorporate a local residual structure into the network architecture. This addition enables the network to effectively learn the local features of the graph and comprehend the intricate relationships within the graph structure of sensor-based human activity. Consequently, our approach yields more precise and generalizable results. Significantly, by integrating an adaptation layer into the network, we can effectively address the disparities in data distribution between the source and target domains. This adaptation layer facilitates the transfer of the model trained in the source domain to the target domain, leading to enhanced classification accuracy. Moreover, our proposed approach overcomes the challenges associated with training small sample datasets alone, such as low classification accuracy, and extends the limitations of traditional transfer learning techniques when applied to small sample sizes.

The subsequent sections of this paper are structured as follows: Section 2 presents an overview of the existing literature and current research efforts in the field. Section 3 outlines the architecture of our proposed framework, followed by a detailed account of its implementation in Section 4. In Section 5, we present the evaluation results of our approach. Finally, the possible future works are discussed in Section 6 and the paper is concluded in Section 7.

2. Related Work

Over an extended duration, as a typical problem in pattern recognition, numerous conventional machine learning algorithms have been employed to address challenges related to human activity recognition (HAR). These algorithms encompass decision trees, random forests (RF), support vector machines (SVM), Bayesian networks (BYS), among others [10]. In the field of machine learning, data samples are typically represented as feature vectors, which serve as inputs for subsequent machine learning algorithms. The process of feature extraction is known to be time-consuming and labor-intensive. Consequently, the concept of feature learning or representation learning has emerged as an intriguing research area, as it enables systems to automatically uncover the representations essential for classification directly from raw data. Deep learning is an integral component of a broader set of machine learning techniques that aim to jointly learn both data features and model parameters from a given dataset.

In 2008, Sharma et al. [11] employed an artificial neural network (ANN) to classify basic activities, and their experiments demonstrated that the average classification rate improved when using a neural network classifier. Subsequently, convolutional neural networks (CNNs) emerged as a significant advancement in image classification through the

utilization of convolutional layers for feature extraction. Krizhevsky et al. [12] introduced deep convolutional neural networks (DCNNs) for large-scale image classification, achieving state-of-the-art results in the ImageNet LSVRC-2010 competition and captivating the attention of researchers. This breakthrough significantly contributed to the advancement of contemporary CNNs. In 2014, Ming Zeng et al. [13] showcased the superiority of CNNs in human motion recognition over existing methods. They accomplished this by employing CNN to extract local dependency features and scale-invariant features from accelerometer time series, followed by comprehensive comparative experiments. Sojeong et al. [14] proposed CNNs (specifically CNNpf and CNN-pff) to address the challenge of feature interference among multiple sensors in HAR. Notably, CNN-pff, designed for multimodal data, demonstrated the feasibility of this multimodal processing approach by achieving improved recognition performance. Vijay et al. [15] focused on gait activity classification and devised a combination of deep learning models, including Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM), CNN-Gated Recurrent Unit (CNN-GRU), LSTM-CNN, and LSTM-GRU. They conducted a comparative analysis of various models, such as CNN, LSTM, GRU, CNN-LSTM, LSTM-CNN, CNN-GRU, and GRU-CNN, using the MHealth dataset and the HAG dataset. Wan et al. [16] developed a system architecture for human activity recognition based on the inertial accelerometer of smartphones. They proposed a CNN-based approach to extract cross-domain knowledge of human activity features, aiming to capture the dissimilarities between similar activities. Li et al. [17] extracted activity features from raw CNN training data and enhanced human activity features by combining the features obtained from inverse CNN data through processes such as de-pooling, de-rectification, and de-convolution. It is worth noting that deep neural network models like CNNs eliminate the need for additional feature extraction. However, they require extensive training data, which can be time-consuming and resource-intensive. Another limitation of CNNs is their disregard for structural information between samples, as they solely consider the features of individual samples. In sensor-based HAR applications, the utilization of sensors from different body parts can offer a more comprehensive representation of human motion. Nevertheless, effectively representing the features and connections between different channels remains an ongoing challenge in current research.

Presently, deep learning has demonstrated notable performance in tasks involving Euclidean space data. However, challenges arise when processing non-Euclidean data using traditional deep learning approaches. Consequently, an expanding body of researchers has started to investigate neural networks designed specifically for graph data processing. Since 2018, graph convolutional networks (GCN) have received much attention, with a large number of GCN-related papers being published.

In the realm of graph neural networks (GNNs), two primary research streams have emerged. The first approach, known as spatial graph convolution, involves convolving each spatial node within the graph. This process entails aggregating information from neighboring nodes, resembling the traditional convolution applied in convolutional neural networks (CNNs). The second approach, termed spectral graph convolution, is a variant of graph convolution grounded in the convolution theorem and graph spectral theory. It leverages the eigenvalues and eigenvectors of the graph's Laplacian matrix to analyze graph properties. Initially, Bruna et al. [18] conducted significant research on spectral convolutional neural networks (SCNN). However, their parametric approach exhibited certain drawbacks, primarily relating to high computational complexity, a large number of parameters, and non-local connections. Subsequently, Michaël Defferrar et al. [19] improved upon GCNN by proposing the Chebyshev Graph Convolutional Network (ChebNet). They defined the Chebyshev polynomial of the diagonal matrix of eigenvectors as a filter and employed it to approximate the convolution kernel, effectively reducing computational complexity. As the field of graph networks continues to advance, numerous new models and applications have emerged, predominantly based on the aforementioned spatial and spectral approaches. Mondal et al. [20] introduced an end-to-end fast graph neural network capable of capturing information not only from individual sensor channels but also from the

relationships with other samples in the form of an undirected graph structure. Mohamed et al. [21] presented HAR-GCCN, a deep graph convolutional neural network model that simultaneously incorporates correlations between different sensor channels. They also devised a novel training strategy that leverages known activity labels to predict missing activity labels through the HAR-GCCN model. Yang et al. [22] proposed a multi-sensor integration mechanism aimed at enhancing the representation of correlations between human activities and adjacent signals. Their approach involved a three-step sorting algorithm to generate an optimal activity graph aligning the width and height of adjacent signals. Additionally, they utilized a deep convolutional neural network to automatically extract discriminative features from the graph. Nian et al. [23] modeled time-series data from sensors as fully connected subgraphs using a sliding window size. They employed spectral graph convolution to extract human motion information from these subgraphs. Notably, their HGCNN method achieved an impressive classification accuracy of 99.54% on the Extra-Sensory dataset.

With the escalating demand and evolving requirements for real-time human activity recognition, an increasing number of experiments have utilized modified versions of the GCNN model in the context of multi-channel sensor-based HAR. However, it is worth noting that many of these experiments depend on distinct data samples from the same dataset for both training and validation purposes. In real-world applications, stringent criteria are imposed on data collection devices, data transmission formats, and data formats themselves, thereby raising concerns regarding the model's ability to generalize effectively. To address these concerns, this paper proposes an approach termed Domain Adaptation GCNN. In addition, a residual network is incorporated to enhance the model's capacity to represent features extracted from sensor signals.

3. Methodology

3.1. Domain-Invariant Feature Learning by GCN

3.1.1. Sensor Networks as Graph

This paper focuses on spectral graph convolution, a method grounded in graph spectral theory and the convolution theorem, to classify human activity signals collected from sensors. In this study, the signals are treated as inherently multivariate time series, serving as the foundation for constructing the underlying graph structure. To quantify the relationships between different channels of signals, Pearson's correlation coefficient metric is employed, Pearson correlation measures the correlation between two signals, a and b , which is defined in (1).

$$\rho(a, b) = \frac{E[(a - \mu_a)(b - \mu_b)]}{\sigma_a \sigma_b} \quad (1)$$

where $E[\cdot]$ denotes the covariance, while σ denotes the respective standard deviations. We use the pre-processed sensor signals to perform sliding-window segmenting and Pearson's correlation coefficient computing to build the graph data for the HAR tasks. In the graph, each node has a dimensional feature vector F , and the feature $X \in \mathbb{R}^{N \times F}$ represents the feature matrix of all N nodes.

3.1.2. Graph Convolutional Network

A l -layer GCN means that it is composed of l layers, while each convolutional layer constructs the embedding l of each node through the previous layer, as shown in (2). The embedding of N nodes at the l -th layer as $X^{(l+1)} \in \mathbb{R}^{N \times F_l}$.

$$Z^{(l+1)} = \tilde{A}X^{(l)}W^{(l)}, X^{(l+1)} = f(Z^{(l+1)}) \quad (2)$$

where X is represented as $X^{(0)} = X$ at the 0-th layer. $\tilde{A} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$, and D is a degree matrix—that is, a diagonal matrix representing the number of adjacent edges. The specific formula is shown in (3).

$$D_{ii} = \sum_j A_{ij} \quad (3)$$

where A_{ij} is the i -th row and j -th column of the adjacency matrix A . $W^{(l)} \in \mathbb{R}^{F(l) \times F(l+1)}$ represents the learned weight matrix, and $f(\cdot)$ represents the activation function. Here, we use the ReLU activation function.

The spectral convolutional neural network (SCNN) combines spectral theory and the convolution theorem to extend the traditional Fourier transform to the graph Fourier transform.

$$U \left((U^T x) \odot (U^T h) \right) = U \text{diag} [\hat{h}(\lambda_1), \dots, \hat{h}(\lambda_N)] U^T x \quad (4)$$

In Equation (4), the eigenvector matrix U is obtained by means of spectral decomposition of the Laplacian matrix. The input node features of the graph are denoted as x , and h is a trainable convolution kernel used to extract the spatial features of the graph structure. The parameters of the same layer are shared. The SCNN replaces the diagonal elements $\hat{h}(\lambda_n)$ of the $\text{diag} [\hat{h}(\lambda_1), \dots, \hat{h}(\lambda_N)]$ diagonal matrix in Equation (3) with learnable parameters θ_n to extract the features. The formula for the SCNN is shown in (5).

$$y = f \left(U g_{\theta}(\wedge) U^T x \right) \quad (5)$$

where $g_{\theta}(\wedge)$ is the convolution kernel, f is the activation function, x is the feature vector of each node in the graph, and y is the output of each node's GCN. The above is the initial GCN, which has the disadvantages of having to perform feature decomposition on the Laplacian matrix every time it is calculated, repeating this step during forward propagation, and consuming excessive computational resources and time when the graph scale is enormous. The ChebNet approximates the convolution kernel using a K -th order Chebyshev polynomial. The Chebyshev polynomial is substituted into the Fourier transform of the graph, and the final result does not require feature decomposition. The formula is shown in (6).

$$\begin{cases} y = f \left(\sum_{k=0}^{K-1} \beta_k T_k(\tilde{L}) x \right) \\ \tilde{L} = \frac{2}{\lambda_{max}} L - I \end{cases} \quad (6)$$

where β_k represents the feature extraction of the k -th layer with parameters. K can be regarded as the receptive field of the convolution kernel. Each central node updates its own feature representation by aggregating adjacent nodes within K . L is the Laplacian matrix and λ_{max} is the largest eigenvalue of the matrix. After the $\tilde{L}$ transformation of Formula (6), the eigen decomposition of the Laplacian matrix is no longer needed in the subsequent calculation process, which can effectively reduce the amount of calculation. The global convolution has become a local convolution. The recursive definition of the Chebyshev polynomial is shown in (7).

$$\begin{cases} T_0(\tilde{L}) = I \\ T_1(\tilde{L}) = \tilde{L} \\ T_{n+1}(\tilde{L}) = 2\tilde{L}T_n(\tilde{L}) - T_{n-1}(\tilde{L}) \end{cases} \quad (7)$$

Based on the aforementioned derivation, the number of parameters in the ChebNet convolution kernel has been effectively reduced from N to K , resulting in improved spatial localization of the convolution kernel.

Considering these advancements, this paper adopts ChebNet as the fundamental model and further enhances its capabilities.

3.1.3. Residual Structure

In order to address the over-smoothing phenomenon that arises when using a multi-layer graph convolutional network, the introduction of a residual network has been proposed [24]. The residual element encompasses two main aspects: shortcut connections and identity mapping. The inclusion of shortcut connections enables the preservation of residual information, while identity mapping facilitates the deepening of the network. The fundamental concept involves adding identity mapping to transform the original function $H(x)$ that requires learning into $F(x) + x$. This idea draws inspiration from residual vector coding in image processing [25]. By reformulating the problem into multiple-scale direct residual problems, effective optimization of training can be achieved. In Figure 1, a residual structure based on four ChebNet Layers is depicted, demonstrating the utilization of Chebyshev graph filtering, normalization, and activation.

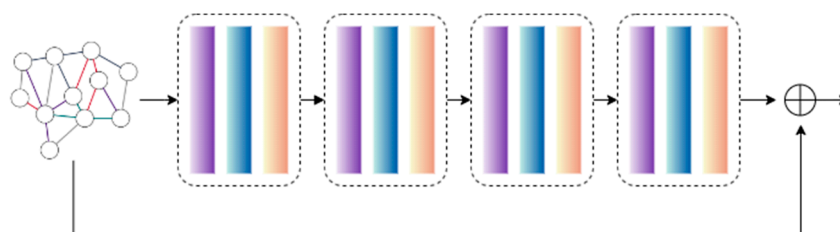


Figure 1. The ResGCN framework.

3.2. Domain Adaptation

3.2.1. Domain Adaptation for Transfer Learning

Traditional machine learning approaches typically assume that data are independently and identically distributed, under the assumption that both the training and test sets are generated from the same underlying data distribution. Consequently, the generalization error is bounded by the training error and the availability of training sample data. However, in practical scenarios, we frequently encounter situations characterized by either limited data samples or variations in data types. To address these challenges, transfer learning methods have emerged as effective solutions.

Domain adaptation is a specialized configuration within the domain of transfer learning. A notable computational hurdle in the realm of domain adaptation pertains to the mitigation of distribution divergence between the source domain and the target domain data. Data distribution adaptation includes marginal distribution adaptation, conditional distribution adaptation, and joint distribution adaptation. In this work, we adopt the marginal distribution adaptation approach, which reduces the distance between the source and target domain's marginal probability distributions, thereby enriching the model and knowledge of the side with fewer samples. When $x = \{x_1, \dots, x_n\} \in X$ denotes a set of learning samples, a domain consists of two fundamental components: the feature space, which encapsulates the input x , and the marginal probability distribution associated with the input $P(x)$. Mathematically, the dissimilarity between the source domain $D_S = \{(x_{S_1}, y_{S_1}), \dots, (x_{S_m}, y_{S_m})\}$ and the target domain $D_T = \{(x_{T_1}, y_{T_1}), \dots, (x_{T_n}, y_{T_n})\}$ is quantitatively estimated by evaluating the distance between the probability distributions $P(x_S)$ and $P(x_T)$.

$$\text{DISTANCE}(D_S, D_T) \approx \|P(x_S) - P(x_T)\| \quad (8)$$

3.2.2. Maximum Mean Discrepancy

In early research, Yang et al. [26] proposed transfer component analysis. The idea is to find a feature mapping function ϕ so that the data can achieve distribution $P\phi((x_S)) \approx P\phi((x_T))$ after passing through the mapping. If the marginal distributions of the two domains are

close, the conditional distributions will also be close. There are an infinite number of such mapping functions. Eric et al. [27] proposed the deep domain adaptation model based on deep transfer learning. An adaptation layer is introduced into the already trained AlexNet network, and a new loss function of maximum mean discrepancy (MMD) is introduced to the original loss to effectively reduce the difference between the source domain and the target domain and achieve a new result. MMD is a kernel learning method that can measure the distance between two data distributions in a reproducing kernel Hilbert space. MMD is shown in (9).

$$MMD(\mathcal{F}, X, Y) = \left\| \frac{1}{m} \sum_{i=1}^m \phi(x_i, x_j) - \frac{1}{n} \sum_{j=1}^n \phi(x_i, x_j) \right\| \quad (9)$$

where m denotes the number of samples in the source domain data, while n represents the number of samples in the target domain data. The kernel function is denoted by the symbol ϕ .

3.3. GCN-Based Domain Adaptation Framework

This paper employs the aforementioned method in the context of HAR. Specifically, a fundamental Chebyshev graph network is constructed, accompanied by the incorporation of a deep network adaptation layer to adapt the parameter distribution across different source data. Additionally, a novel loss value, the MMD, is introduced. The loss function formula for the deep adaptive model in the graph is presented in Equation (10).

$$\ell = \ell(D_S, Y_S) + \lambda MMD^2(D_S, D_T) \quad (10)$$

In the above equation, $\ell(D_S, Y_S)$ represents the cross-entropy loss function, $\lambda MMD^2(D_S, D_T)$ represents the MMD loss, and the sum of the two is the total loss function.

The adaptive processing of the GDA network is illustrated in Figure 2.

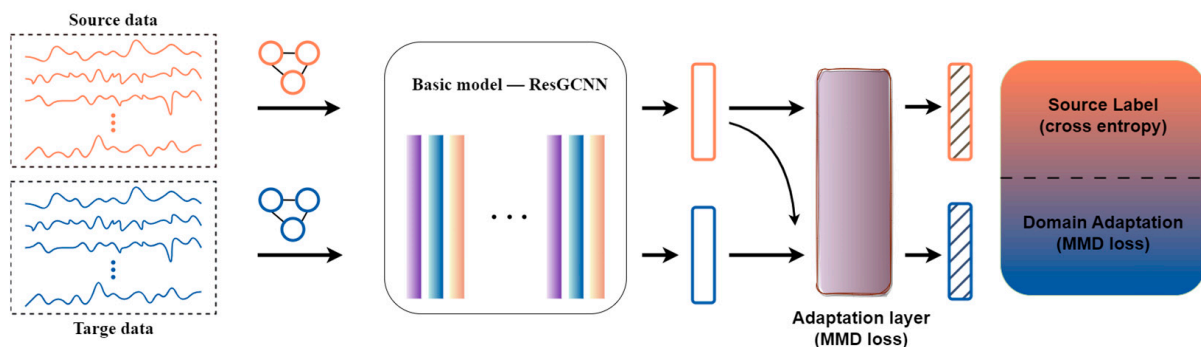


Figure 2. Adaptive processing of GDA network. After the source domain data and target domain data pass through ResGCNN, the cross-entropy loss is calculated after the source domain data passes through a classification layer. At the same time, the source domain data and a small amount of target domain data enter the adaptive layer and calculate the MMD loss. Finally, the total loss of the model is formed.

4. Experiment

4.1. Dataset

In this study, three public datasets were utilized: the MHealth dataset [28], consisting of 12 activities, the PAMPA2 dataset [29], comprising 12 activities, and the TNDA-HAR dataset [30], containing 8 actions. The specific details of these datasets are presented in Table 1. Notably, the MHealth and PAMPA2 datasets are widely employed for evaluating sensor-based human activity recognition (HAR) models. For the purpose of transfer learning experiments, we selected seven commonly occurring action types from these

three datasets for transfer training and recognition. These actions include lying, sitting, standing, walking, running, cycling, and climbing stairs.

Table 1. Summary of the HAR datasets used for our tests.

Dataset	Sensor Placement	Channel	Subject	Activities Classes	Activities	IMU Frequency
MHealth	IMU: chest, right wrist, left ankle 2-lead ECG	24	10	12	Standing still, sitting and relaxing, lying down, walking, climbing stairs, waist bends forward, frontal elevation of arms, knees bending, cycling, jogging, running, jumping forwards and backwards Lying, sitting, standing, walking, running, cycling, Nordic walking, watching TV, computer work, car driving, ascending stairs, descending stairs, vacuum cleaning, ironing, folding laundry, house cleaning, playing soccer, rope jumping	50 Hz
PAMAP2	IMU: chest, right wrist, right ankle HR-monitor	54	9	12	Sitting, standing, laying, walking, running, cycling, walking upstairs, walking downstairs	100 Hz
TNDA	IMU: left ankle, left knee, back, right wrist, right arm	46	50	8		50 Hz

4.1.1. MHealth Dataset

The Mobile HEALTH (MHealth) dataset comprises recordings of physical movements and vital signs from ten volunteers engaged in 12 sports activities. In order to capture the movements experienced by different body parts, inertial sensors were placed on the volunteer's chest, right wrist, and left ankle. These sensors measured various parameters, including acceleration, angular velocity, and magnetic field orientation. The data from all sensor modes were collected at a sampling rate of 50 Hz. The dataset includes the following specific activities: standing still (1 min), sitting and relaxing (1 min), lying down (1 min), walking (1 min), climbing stairs (1 min), waist bends forward (20 repetitions), frontal elevation of arms (20 repetitions), knees bending (crouching) (20 repetitions), cycling (1 min), jogging (1 min), running (1 min), and jumping forwards and backwards. The dataset consists of 24 columns, where "min" represents the duration of the exercises.

4.1.2. PAMAP2 Dataset

The PAMAP2 dataset encompasses activities performed by nine participants, who were instructed to engage in a total of 12 diverse activities. These activities include both daily activities and various exercise activities such as walking, cycling, and playing football. To capture the participants' movements and physiological states, three inertial measurement units (IMUs) comprising accelerometers, gyroscopes, and magnetometers were employed. Additionally, temperature and heart rate data were recorded. The dataset was collected at a sampling frequency of 100 Hz. As a result, the dataset comprises 54 dimensions. The specific activities included in the dataset are standing still, sitting and relaxing, lying down, walking, climbing stairs, waist bends forward, frontal elevation of arms, knees bending, cycling, jogging, running, and jumping forwards and backwards.

4.1.3. TNDA Dataset

The TNDA dataset encompasses data collected from 50 subjects engaged in eight distinct types of movements. To capture the movements, five nine-axis inertial sensors were affixed to specific body parts of the subjects, including the left ankle, left knee, back, right wrist, and right arm. The volunteers were given instructions to perform eight actions while maintaining fixed positions. These actions comprised sitting, standing, lying down,

walking, running, cycling, walking upstairs, and walking downstairs. Each activity had a duration of 2 min. The recorded data for each sensor included three-axis acceleration from the accelerometer, three-axis angular velocity from the gyroscope, and three-axis magnetic field data from the magnetometer.

4.2. Experiment Settings

In the experiment involving graph models based on Euclidean space, several pre-processing steps were undertaken. Firstly, all sensor time series were normalized and resampled at a frequency of 50 Hz following noise filtering. Subsequently, data samples were prepared using a sliding window approach with a fixed length of 128 and an overlap rate of 50%. Considering the variations in sampling frequency across different datasets, it was established that each window in the MHEALTH dataset had a duration of 2.56 s, whereas the windows in the PAMAP2 and TNDA datasets had durations of 1.28 s. As a consequence, a total of 13,092, 5361, and 11,784 activity time series segments were obtained from TNDA-HAR, MHEALTH, and PAMAP2, correspondingly. Each of these segments was regarded as an individual sample. To serve as input for the GCN model, a graph was constructed for each sample. In this graph, each sensor channel was treated as a distinct node. The Pearson correlation coefficient was employed to calculate the correlation between each node, resulting in a correlation coefficient matrix. Nodes exhibiting a correlation coefficient exceeding 0.2 were considered to possess a strong correlation and were subsequently connected.

In this work, we use the Pytorch framework with the geometric package to build the neural network model. For each evaluation, we performed 5-fold cross-validation and 80% of the samples were used as training data while 20% were used as the testing data. In the experiments, the learning rate of the model was set as 0.001, and the batch size was set as 64. We used a maximum of 100 training rounds and the Adam stochastic optimization algorithm to optimize the network parameters.

We conducted two experiments, denoted as Exp. #1 and Exp. #2, respectively. In Exp. #1, the focus was on processing and converting the sensors into graph-structured data without considering the source domain and target domain elements. Sensor-based human activity recognition (HAR) tasks were performed separately on each dataset to assess the learning capability of the ResGCNN base model. The specific parameters of the GCNN base model can be found in Table 2. In Exp. #2, we employed a GCN-based domain adaptation model to address small sample learning tasks. Public activity labels were utilized for evaluating transfer learning. The MHealth dataset was utilized as the target domain, whereas the TNDA and PAMAP2 datasets were employed as the source domains. Initially, pre-training was conducted on the source domains (TNDA or PAMAP2) using the base GCN model. The resulting parameters were subsequently retained and utilized in subsequent transfer learning tasks. To optimize the parameters for transfer learning, fine-tuning was performed using a mere 1% of the samples from the target domain. Furthermore, to evaluate the performance of human activity recognition under different target and source domain configurations, the PAMAP2 dataset and TNDA dataset were employed as target domains, while the remaining two datasets served as the source domains. This allowed us to conduct the same experiment as described above. In addition to the proposed GDA-based transfer learning approach, we also assessed transfer learning without the utilization of domain adaptation techniques. A comparative analysis was then conducted among fine-tuning methods without domain adaptation techniques, conventional fine-tuning methods, and fine-tuning methods employing GDA.

Table 2. GCN structure parameters.

Block	Layer	Operator	Graph Filtering Parameters' Size	Norm Size
1st	1	ChebNet Layer	$128 \times 256 \times 1$	256
2nd	2	ChebNet Layer	$256 \times 512 \times 1$	512
3rd	3	ChebNet Layer	$512 \times 256 \times 1$	256
4th	4	ChebNet Layer	$256 \times 128 \times 1$	128
-	5	FC Layer	128×64	-
-	6	FC Layer	$64 \times \text{No. Labels}$	-

4.3. Evaluation Metrics

For a more comprehensive evaluation of the proposed method, four diverse evaluation metrics, including precision, recall, F-measure, and accuracy were utilized, which are mathematically formulated as

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (11)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (12)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (13)$$

$$F1 - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (14)$$

where TP , FN , FP , and TN represent true positives, false negatives, false positives, and true negatives, respectively.

5. Results

5.1. Sensor-Based HAR with Base Model

We evaluated our GDA model based on three different datasets: MHealth, PAMAP2, and TND. In the HAR experiment, our model achieved accuracies of 98.88%, 97.78%, and 98.58% on these respective datasets. Comparing these results with several other machine learning classification algorithms, such as SVM, RF, BYS, and extreme gradient boosting (XGB), our model outperformed them in HAR applications, as illustrated in Table 3.

Table 3. The accuracy of activity recognition under different methods.

Dataset	SVM	RF	BYS	XGB	CNN	LSTM	CNN-LSTM	Ours
Mhealth	90.8%	85.27%	90.80%	96.15%	91.94% [14]	86.89% [15]	91.66% [15]	98.88%
Pamap2	84.93%	73.88%	84.93%	89.44%	96.68% [17]	85.86% [16]	96.97% [17]	98.58%
TNDA	89.41%	83.91%	89.41%	91.29%	-	-	-	97.78%

Currently, our model surpasses some state-of-the-art (SOTA) methods. In previous studies using the MHealth dataset, Sojeong et al. [14] employed partial weight sharing and full weight sharing mechanisms with 2D-CNNs for HAR and achieved an average accuracy of 91.94%. Vijay et al. [15] proposed a hybrid deep learning model using an ensemble learning approach, attaining an average accuracy of 94.00%. They also compared their work to previous studies utilizing CNN, LSTM, gated recurrent unit (GRU), CNN-LSTM, CNN-GRU, and GRU-CNN models, which achieved average recognition accuracies of 91.66%, 86.89%, 81.77%, 91.66%, 91.66%, and 92.53%, respectively.

Regarding the PAMAP2 dataset, the multiscale deep feature learning method proposed by Tang et al. [31] achieved a recognition accuracy of 93.75% on the PAMAP2 dataset. The

Actual Fusion within Virtual Fusion (AFVF) method, proposed by Nguyen et al. [32], achieved an accuracy of 96.72%. Wan et al. [16] used a three-layer CNN structure and LSTM for action classification recognition, resulting in accuracies of 91.00% and 85.86%, respectively. Li et al. [17] employed a three-layer CNN structure and a three-layer reverse CNN structure to extract features from motion data. Based on domain knowledge, they selected a CNN (three-layer)-LSTM (one-layer) structure as the deep learning classifier. They compared different combined features in the PAMAP2 dataset to assess their impact on experimental results, achieving average accuracies ranging from 96.68% to 97.14%, and 96.97% to 97.37%, respectively.

Moreover, as depicted in Figure 3, the GDA model demonstrates remarkable recognition results for each distinct labeled action in the MHealth, PAMAP2, and TNDA datasets. Within the MHealth dataset, the GDA achieves a minimum recognition accuracy of 97% for each action. Similarly, in the PAMAP2 dataset, the GDA attains a minimum recognition accuracy of 96% for each action. Furthermore, in the TNDA dataset, the GDA demonstrates a minimum recognition accuracy of 93% for each action.

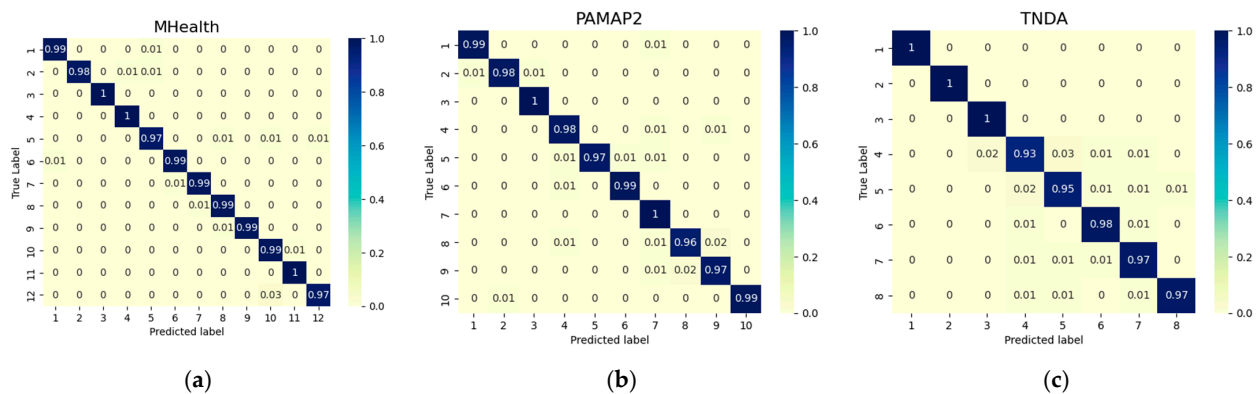


Figure 3. Results of the base model in Exp. #1. (a–c) show the human activity recognition results of our model on MHealth, PAMAP2, and TNDA datasets, respectively.

5.2. Cross Domain Evaluation by Transfer Learning with GDA

In practical applications, encountering challenges related to insufficient data samples or disparate data sources is a common occurrence. Consequently, enhancing learning capabilities under conditions of limited data samples becomes a pivotal concern. To assess the transferability of our graph domain adaptation approach, we conducted comparative experiments utilizing three distinct datasets: the MHealth dataset, the PAMAP2 dataset, and the TNDA dataset. We employed six different combinations to evaluate the data transferability of the three models, and the recognition accuracy was measured as a percentage (%). Table 4 provides a comprehensive overview of the experimental outcomes. The first column illustrates the experimental settings, which describe the configurations of the source domain dataset and the target domain dataset. The second column presents the experimental results of the base GCN model without transfer learning, denoted as “None TF”, under different dataset settings. The third column showcases the recognition accuracy outcomes of the base GCN model under various experimental settings when employing the fine-tuning strategy, labeled as “Fine-tuning TF”. The fourth column exhibits the experimental results of our GDA method utilizing the fine-tuning strategy, referred to as “GDA-TF”. Notably, upon observation of this table, the recognition accuracy achieved by GDA-TF exceeds that achieved by the other two methods most of the time in the context of human activity recognition, while keeping the experimental settings unchanged.

Table 4. Transfer learning experiments and comparisons in few-shot learning.

Settings	None TF (%)	Fine-Tuning TF (%)	GDA-TF (%)
PAMAP2 to MHealth TNDA to MHealth	79.72/78.81/80.55/79.72	70.75/70.44/72.49/70.75 71.83/71.29/75.82/71.83	85.49/85.46/87.21/85.49 88.57/88.42/89.45/88.57
MHealth to PAMAP2 TNDA to PAMAP2	75.44/75.59/76.03/75.44	74.75/74.95/75.77/74.75 79.25/79.18/79.23/79.25	75.43/75.19/75.56/75.43 81.49/81.40/81.80/81.49
PAMAP2 to TNDA MHealth to TNDA	84.75/84.89/85.66/84.75	84.95/85.11/85.80/84.95 82.49/82.44/83.05/82.49	87.15/87.18/87.39/87.15 85.64/85.78/86.42/85.64

Each value in the table represents accuracy/F1-score/precision/recall, with the unit expressed as a percentage (%).

Initially, the PAMAP2 and TNDA datasets were employed as the source datasets, while the MHealth dataset served as the target dataset for our evaluation of transfer learning. The public activity labels were utilized for this evaluation. We used 1% of the data samples in MHealth as the fine-tuning training set, while 80% of the MHealth data samples were used as the validation set (activity samples share similar distribution). The outcomes presented in Table 4 demonstrate the effectiveness of different approaches. In the non-transfer case, the base GCN model was trained using all the data samples from the PAMAP2 and TNDA datasets and subsequently used for recognition tasks in the MHealth dataset. The achieved accuracy for this scenario was 79.72%. Subsequently, the traditional fine-tuning strategy was applied, utilizing only 1% of the MHealth dataset samples. The resulting accuracies for TNDA and PAMAP2 were 70.75% and 71.83%, respectively, indicating a decline in recognition performance compared to the non-transfer case. Finally, we employed GDA-based transfer learning techniques, which yielded transfer learning accuracies of 85.49% for PAMAP2 and 88.57% for TNDA. These figures represent notable enhancements of 5.77% and 8.85%, respectively, compared to the results obtained without transfer learning. The experimental outcomes depicted in Figure 4a–e correspond to the results obtained by utilizing the MHealth dataset as the target domain data, aligning with the first and second rows in Table 4. In Figure 4a, it is evident that the non-transfer case exhibits a significant deviation in the recognition result for climbing stairs, with only 27% accuracy and 49% misclassification as walking. After implementing fine-tuning, the recognition effect for climbing stairs significantly improves, while the recognition performance for other actions such as standing, walking, running, and cycling decreases. Figure 4b,c demonstrate that in the PAMAP2 to MHealth case, 36% of walking instances were mistakenly recognized as climbing stairs, while in the TNDA to MHealth case, 44% of walking instances were misclassified as climbing stairs, with an accuracy rate only 3% lower than the correct recognition rate. However, as depicted in Figure 4d,e, when employing our GDA network, significant improvements in the accuracy of recognizing climbing stairs were achieved compared to the previous calculation methods, with accuracy rates of 68% and 89%, respectively. Additionally, we observed that, compared to the non-fine-tuning case, the fine-tuning method led to better recognition of sitting activities, while the recognition accuracy for standing decreased, and there was a certain probability of misidentification as sitting. This outcome can be attributed to the similarity between sitting and standing, both being static activities, which makes them prone to misclassification. Therefore, further research is needed to develop improved recognition methods for static actions.

Subsequently, the PAMAP2 dataset was chosen as the target domain dataset, while the MHealth and TNDA datasets were employed as the source datasets. For fine-tuning training, only 1% of the data samples from PAMAP2 were utilized, and the remaining 80% served as the validation set. A comparative analysis was conducted to evaluate the transferability of the three models. The overall recognition results and the recognition capabilities pertaining to seven distinct activity types were represented using the accuracy values in Table 4 and the corresponding confusion matrices in Figure 5a–e. In the case of the base GCN model, the recognition accuracy without transfer learning stood at 75.44%, as demonstrated by the confusion matrix depicted in Figure 5a. When the fine-tuning strategy

was employed, the recognition results for MHealth to PAMAP2 were 74.75%, and for TNDA to PAMAP2 were 79.25%. These results correspond to Figure 5b,c, respectively. However, upon training our GDA-based neural network, the recognition results improved to 75.43% for MHealth to PAMAP2, and increased to 81.49% for TNDA to PAMAP2. These results correspond to Figure 5d,e, respectively. The GDA network outperformed the Fine-tuning TF case in both scenarios. Furthermore, considering the limited number of subjects in the PAMAP2 dataset, the model's recognition performance may be poorer when using only 1% of the PAMAP2 dataset as the target data. The four confusion matrices, ranging from Figure 5b–e, reveal that when conducting transfer performance verification with PAMAP2 as the target dataset, the recognition performance of the motion types standing, walking, and climbing stairs is generally poorer compared to that for other motion types. This difference may be attributed to the fact that standing involves static data, which can be easily confused with sitting when using only IMU data as the input. Consequently, there is a higher likelihood of misclassifying standing as sitting during recognition. Additionally, walking and climbing stairs exhibit similarities in terms of motion types. As a result, in Figure 5d, 26% of the walking data are misclassified as climbing stairs, and 17% of the climbing stairs data are misclassified as walking. Likewise, in Figure 5e, 12% of the walking data are misclassified as climbing stairs, and 13% of the climbing stairs data are misclassified as walking.

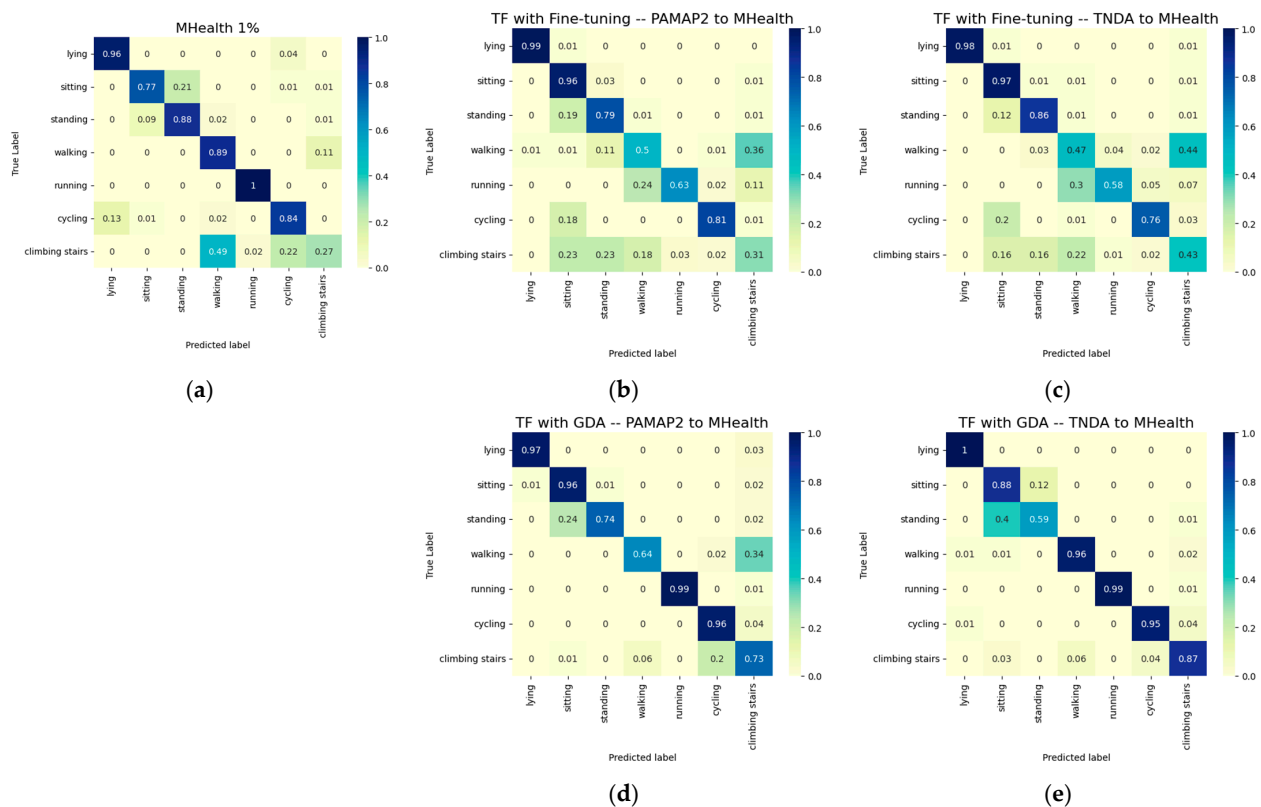


Figure 4. The results of the GDA model in Exp. #2, with MHealth as the target domain. (a) Non-transfer (b) Fine-tuning strategy when PAMAP2 dataset is source domain data (c) Fine-tuning strategy when TNDA dataset is source domain data (d) GDA-based transfer learning when PAMAP2 dataset is source domain data (e) GDA-based transfer learning when TNDA dataset is source domain data.

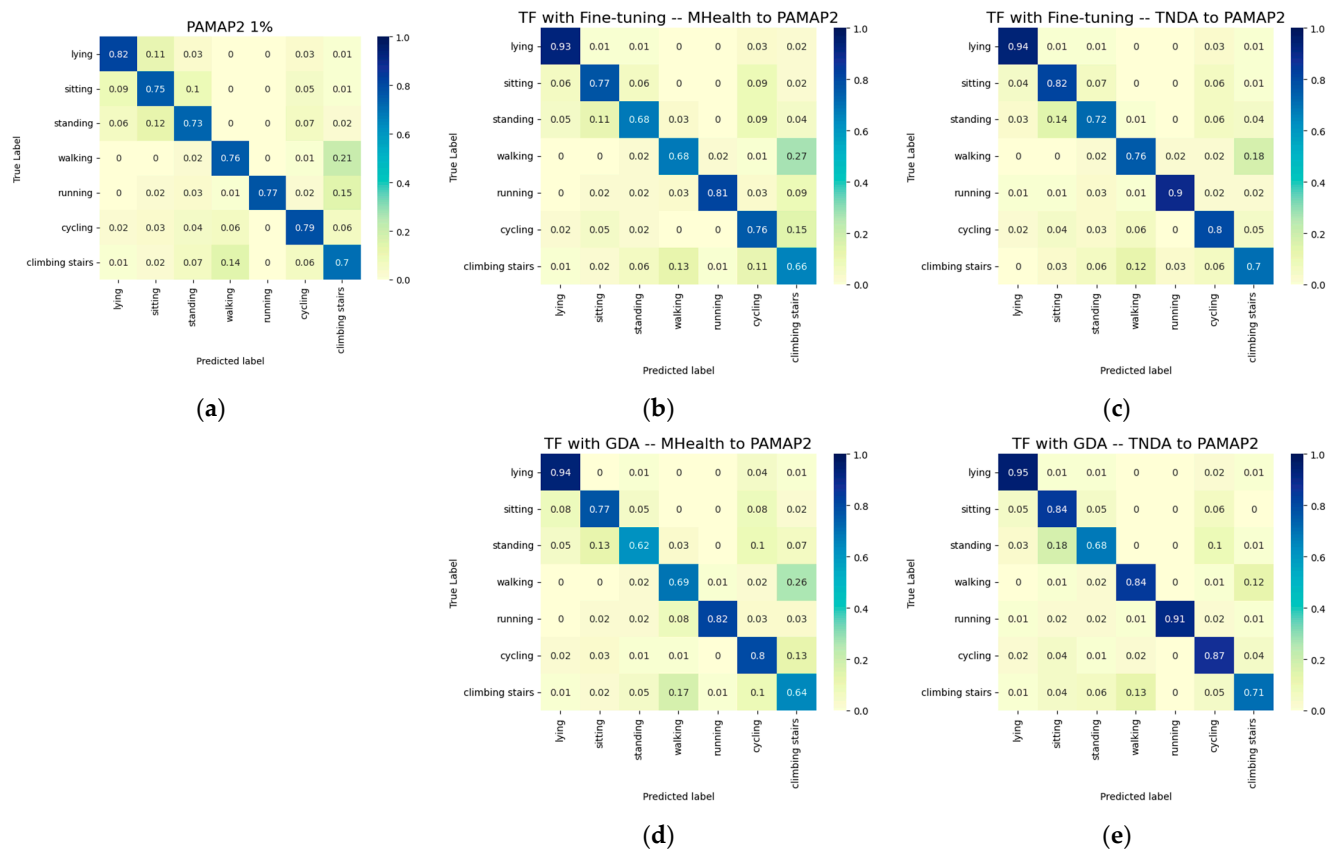


Figure 5. The results of the GDA model in Exp. #2, with PAMAP2 as the target domain. (a) Non-transfer (b) Fine-tuning strategy when MHealth dataset is source domain data (c) Fine-tuning strategy when TNDA dataset is source domain data (d) GDA-based transfer learning when MHealth dataset is source domain data (e) GDA-based transfer learning when TNDA dataset is source domain data.

Finally, using 1% of the TNDA dataset as the fine-tuning training set, we separately performed fine-tuning with 1% of the TNDA dataset using PAMAP2 and MHealth as the source domains. We used 80% of the TNDA dataset as the validation set for the None TF, Fine-tuning TF, and GDA-TF models. The recognition results are displayed in the last two rows of Table 4, and the confusion matrices can be found in Figure 6a–e. In the non-transfer scenario, we trained the base GCN model using all data samples from PAMAP2 and MHealth, achieving an accuracy of 84.75% for both datasets. The confusion matrix is shown in Figure 6a. Then, we applied traditional fine-tuning strategies, obtaining recognition accuracies of 84.95% for PAMAP2 to TNDA and 82.49% for MHealth to TNDA. The recognition performance for each motion type under these two fine-tuning strategies is shown in Figure 6b,c, respectively. Finally, we utilized GDA-based transfer learning techniques, achieving transfer learning results of 87.15% for PAMAP2 and 85.64% for MHealth. The corresponding confusion matrices can be found in Figure 6d,e. These results are better than the Fine-tuning TF case. Compared to using the base GCN, our GDA improved the recognition accuracy by 2.2% for PAMAP2 to TNDA and 3.15% for MHealth to TNDA in the two different experimental setups, demonstrating the good generalization performance of GDA. In Figure 6a, 24% of the lying data were recognized as sitting, which is a clear limitation when using only IMU for human activity recognition. Similarly, there was an approximately 12% misclassification between standing and sitting. When we introduced fine-tuning, the recognition ability for lying improved, but the recognition performance for other action types declined. As shown in Figure 6c, the accuracy of sitting recognition was only 67%, a decrease of 17% compared to Figure 6a. The accuracy of cycling recognition decreased by 13%, and the accuracy of climbing stairs recognition decreased by 7%. Finally, when we changed the trained model to our GDA, as shown in Figure 6d,e,

the recognition performance for various motion types generally improved, addressing the significant deficiencies of the previous two methods in recognizing lying and sitting.

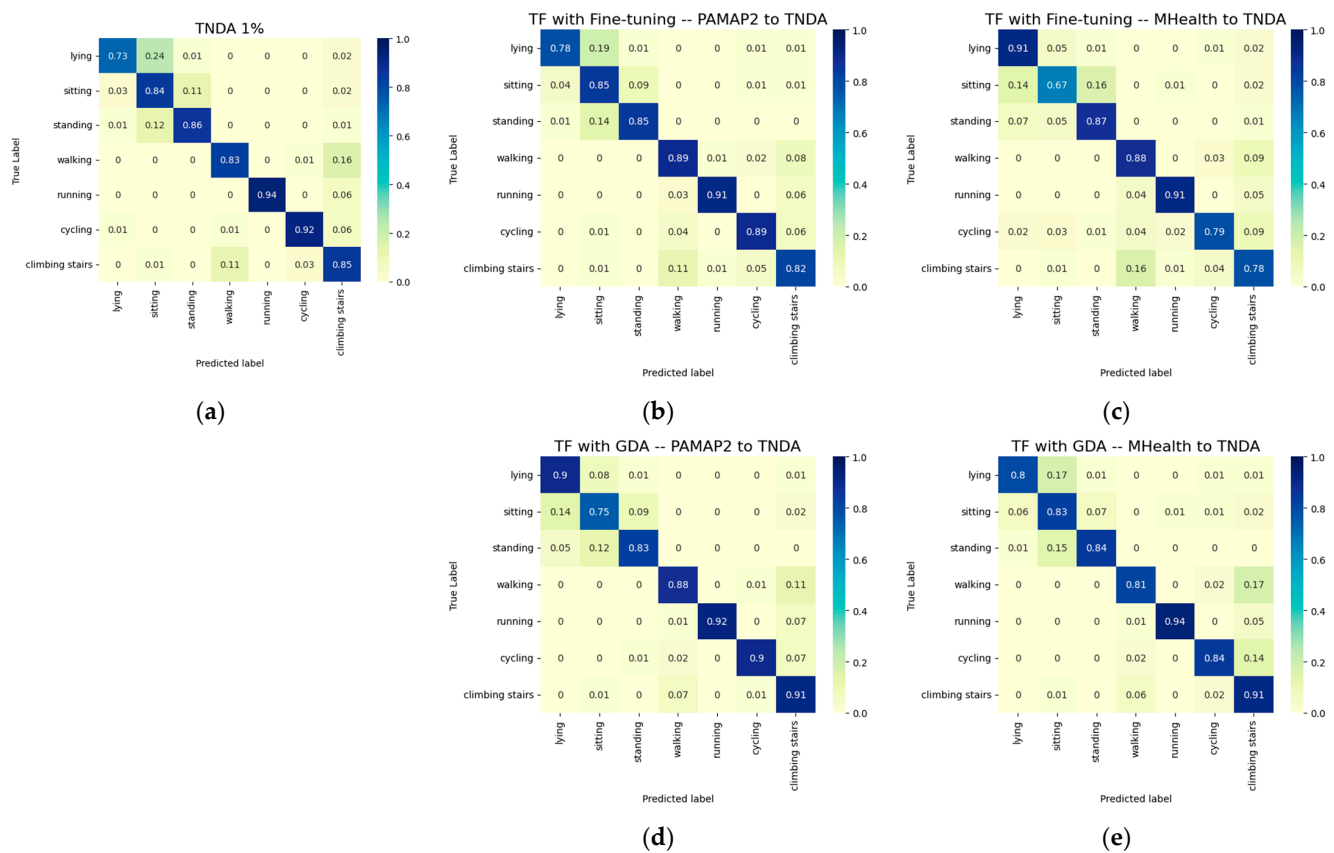


Figure 6. The results of the GDA model in Exp. #2, with TND as the target domain. (a) Non-transfer (b) Fine-tuning strategy when PAMAP2 dataset is source domain data (c) Fine-tuning strategy when MHealth dataset is source domain data (d) GDA-based transfer learning when PAMAP2 dataset is source domain data (e) GDA-based transfer learning when MHealth dataset is source domain data.

Based on the above results, it is evident that our GDA model has higher transferability. In the six comparative experiments comparing non-transfer learning and base GCN fine-tuning, GDA-TF demonstrated significant recognition performance, with only one experiment showing slightly lower effectiveness compared to the other models.

6. Discussion

In this work, we developed a GNN mode to make improvements to the HAR method based on inertial measurement Unit data, resulting in significant recognition performance on the MHealth dataset, PAMAP2 dataset, and TND dataset, respectively. Moreover, restricted by the dataset variation and subject differences, we proposed a GDA approach for the transfer learning task in using the model for small sample learning applications. Our transfer learning experiments reveal that the GDA model showcases distinct advantages in transfer performance when compared to the non-transfer base GCN model and the base GCN model employing the fine-tuning strategy. GDA effectively tackles cross-domain recognition challenges, as shown in the above section.

6.1. Learning Process Evaluation with GDA

The learning loss curve and test accuracy curve based on the training epoch are shown in Figures 7 and 8, respectively. As can be seen from Figure 7, negative transfer occurred when using fine-tuning without GDA, and the accuracy after fine-tuning was lower than the original with the increase in the number of iterations. When the GDA network was

added, the negative transfer problem was effectively solved. GDA enhanced the recognition accuracy of small samples by adjusting the data distribution. In addition, as shown in Figure 7, the train loss also decreased significantly. Compared with non-transfer learning, the GDA network after fine-tuning has better learning ability and convergence speed. As illustrated in Figure 8, the test accuracies with GDA transfer learning are 72.5% and 63.45% in the 0-th epoch case when TNDA is used as the source dataset and PAMAP2 is used as the source dataset, respectively, which are better than the non-transfer case with 14.10%. When TNDA is used as the source dataset, the average recognition accuracy reaches 80% after only one epoch, while when PAMAP2 is used as the source dataset, the accuracy reaches 80% after only eleven epochs. On the contrary, in the non-transfer case, the accuracy is still lower than 80% after 30 epochs.

Transfer Learning

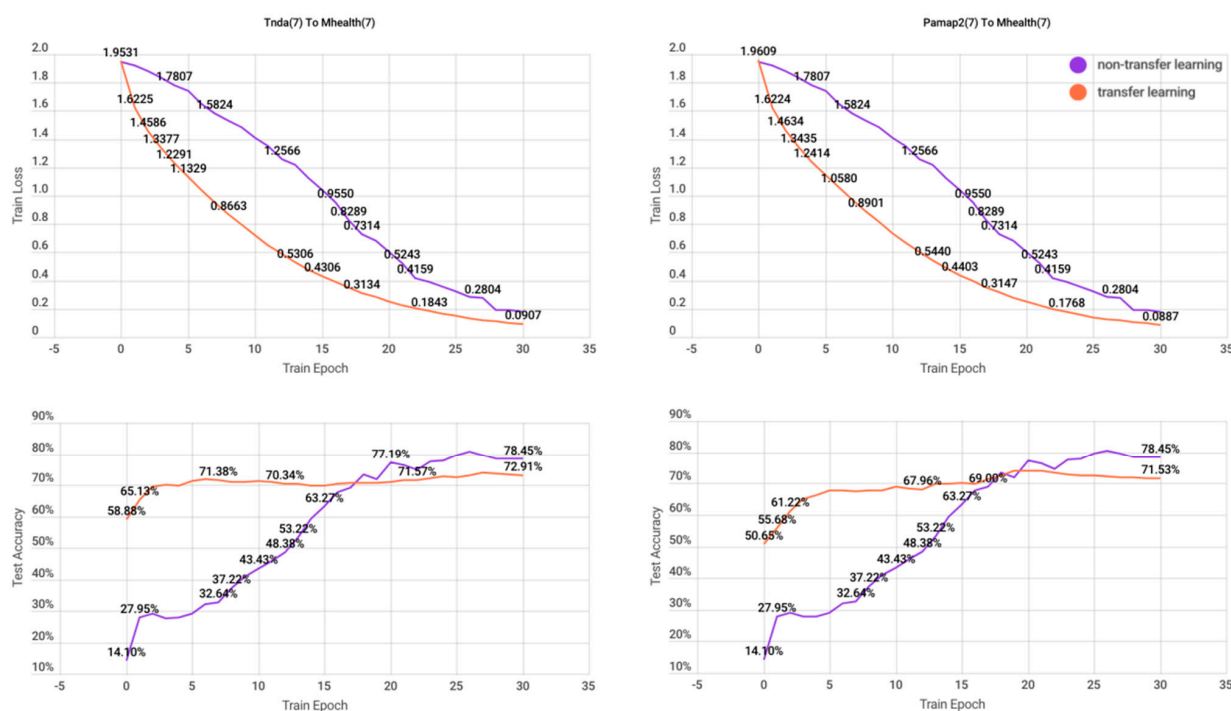


Figure 7. Finetuning transfer learning, with MHealth as the target domain.

6.2. Limitations and Future Directions

In our transfer experiments, we observed that certain static actions, particularly standing and sitting, may be misclassified as other static actions during recognition. Similarly, periodic actions such as walking and climbing stairs also have a tendency for misclassification between them. These issues can be attributed to two main reasons. Firstly, the limited amount of training data used in the experiment can be addressed by incorporating more source domain data during the pre-training stage. Secondly, the experiment solely relies on inertial sensor data, which primarily captures the motion and posture of key body joints, including acceleration and angular changes. As a result, similar types of movements may face challenges regarding misclassification. Thus, it is beneficial to include other sensor data, such as heart rate and electromyography data during human movement, in addition to the currently used IMU data, which would enable the assessment of human motion types from multiple dimensions. However, the proposed GDA mechanism still presents significant challenges in dealing with the generalization to real-world applications.

GDC Transfer Learning

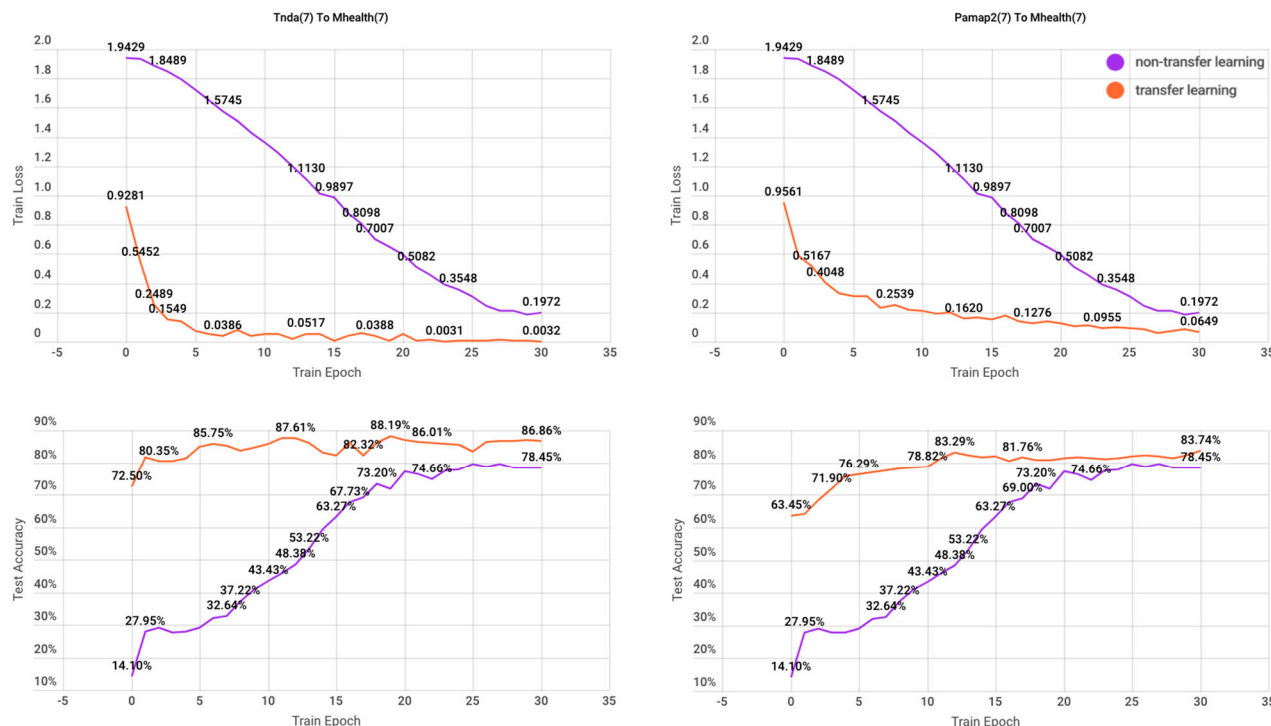


Figure 8. GDA transfer learning, with MHealth as the target domain.

7. Conclusions

Sensor-based HAR is a representative mobile computing scenario in biomedical information acquisition and analysis. We propose a GCN-based domain adaptation model for sensor-based HAR tasks, namely the GDA framework. The GDA network takes a multi-layer neural network composed of GNNs with Chebyshev filter functions and residual structures as the basic model. It adjusts the learnable parameters through an adaptation layer with MMD loss for domain adaptation to learn sensor signal representations and recognize human activities. The significance of this work lies in (1) the GDA network showing excellent classification ability in HAR tasks and (2) great transferability being shown for small-sample learning tasks. The GDA-based model shows solid small-sample learning (few-shot learning) ability in distinguishing activity categories. These findings provide a feasible solution for practical applications of classification tasks with insufficient data samples and differences in data types. Our model would be a good choice for sensor-based HAR tasks and mobile learning applications. By employing a limited number of annotated samples for training, one can accomplish cross-domain recognition of human behaviors by capitalizing on pre-existing knowledge from diverse domains. This methodology significantly diminishes the expenses associated with manual annotation and establishes a robust technological framework for tackling future challenges in cross-domain recognition concerning wearable robotics.

Author Contributions: Conceptualization, J.Y.; methodology, T.L.; software, T.L. and J.Z.; validation, J.Y.; formal analysis, J.Y. and T.L.; investigation, J.Z.; resources, C.L. and J.X.; data curation, Y.Y. and Z.Z.; writing—original draft preparation, J.Y.; writing—review and editing, Y.Y., C.L. and J.X.; visualization, Y.H.; supervision, C.L. and Y.Y.; project administration, C.L. and J.X.; All authors have read and agreed to the published version of the manuscript.

Funding: The authors acknowledge the funding of the following science foundations: the Science and Technology Planning Project of Guangzhou, China (202102010392), the Science and Technology Planning Project of Guangdong Province, China (2020A1414050067), National Key Research

and Development Program of China under Grant (2020YFC2007200), Guangdong Basic and Applied Basic Research Foundation (2022B1515020042), Shenzhen Science and Technology Program (JCYJ20210324115606018), and Shenzhen Engineering Laboratory for Diagnosis & Treatment Key Technologies of Interventional Surgical Robots (XMHT20220104009).

Data Availability Statement: Data will be made available upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Bakaev, M.; Speicher, M.; Heil, S.; Gaedke, M. I Don't Have That Much Data! Reusing user behavior models for websites from different domains. In *International Conference on Web Engineering*; Springer International Publishing: Cham, Switzerland, 2020; pp. 146–162.
2. Malekzadeh, M.; Clegg, R.; Cavallaro, A.; Haddadi, H. Dana: Dimension-adaptive neural architecture for multivariate sensor data. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2021**, *5*, 1–27. [\[CrossRef\]](#)
3. Thakur, D.; Biswas, S. Online change point detection in application with transition-aware activity recognition. *IEEE Trans. Hum.-Mach. Syst.* **2022**, *52*, 1176–1185. [\[CrossRef\]](#)
4. Krishnaprabha, K.K.; Raju, C.K. Predicting human activity from mobile sensor data using CNN architecture. In Proceedings of the 2020 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA), Cochin, India, 2–4 July 2020; pp. 206–210.
5. Thakur, D.; Pal, A. Subsampled Randomized Hadamard Transformation based Ensemble Extreme Learning Machine for Human Activity Recognition. *ACM Trans. Comput. Healthc.* **2023**, *5*, 1–23. [\[CrossRef\]](#)
6. Thakur, D.; Ro, S.; Biswas, S.; Ho, E.S.L. A Novel Smartphone-Based Human Activity Recognition Approach using Convolutional Autoencoder Long Short-Term Memory Network. In Proceedings of the 2023 IEEE 24th International Conference on Information Reuse and Integration for Data Science (IRI), Bellevue, WA, USA, 4–6 August 2023; pp. 146–153.
7. Thakur, D.; Guzzo, A.; Fortino, G. Attention-based Multihead Deep Learning Framework for online activity monitoring with Smartwatch Sensors. *IEEE Internet Things J.* **2023**, *10*, 17746–17754. [\[CrossRef\]](#)
8. Koşar, E.; Barshan, B. A new CNN-LSTM architecture for activity recognition employing wearable motion sensor data: Enabling diverse feature extraction. *Eng. Appl. Artif. Intell.* **2023**, *124*, 106529. [\[CrossRef\]](#)
9. Mahmud, T.; Akash, S.S.; Fattah, S.A.; Zhu, W.P.; Ahmad, M.O. Human activity recognition from multi-modal wearable sensor data using deep multi-stage LSTM architecture based on temporal feature aggregation. In Proceedings of the 2020 IEEE 63rd International Midwest Symposium on Circuits and Systems (MWSCAS), Springfield, MA, USA, 9–12 August 2020; pp. 249–252.
10. Wang, J.; Chen, Y.; Hao, S.; Peng, X.; Hu, L. Deep learning for sensor-based activity recognition: A survey. *Pattern Recognit. Lett.* **2019**, *119*, 3–11. [\[CrossRef\]](#)
11. Sharma, A.; Lee, Y.-D.; Chung, W.-Y. High Accuracy Human Activity Monitoring Using Neural Network. In Proceedings of the 2008 Third International Conference on Convergence and Hybrid Information Technology, Busan, Republic of Korea, 11–13 November 2008; pp. 430–435.
12. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012, Lake Tahoe, NV, USA, 3–6 December 2012.
13. Zeng, M.; Nguyen, L.T.; Yu, B.; Mengshoel, O.J.; Zhu, J.; Wu, P.; Zhang, J. Convolutional Neural Networks for human activity recognition using mobile sensors. In Proceedings of the 6th International Conference on Mobile Computing, Applications and Services, Austin, TX, USA, 6–7 November 2014; pp. 197–205. [\[CrossRef\]](#)
14. Ha, S.; Choi, S. Convolutional neural networks for human activity recognition using multiple accelerometer and gyroscope sensors. In Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 24–29 July 2016; pp. 381–388.
15. Semwal, V.B.; Gupta, A.; Lalwani, P. An optimized hybrid deep learning model using ensemble learning approach for human walking activities recognition. *J. Supercomput.* **2021**, *77*, 12256–12279. [\[CrossRef\]](#)
16. Wan, S.; Qi, L.; Xu, X.; Tong, C.; Gu, Z. Deep learning models for real-time human activity recognition with smartphones. *Mob. Netw. Appl.* **2020**, *25*, 743–755. [\[CrossRef\]](#)
17. Li, X.; Nie, L.; Si, X.; Ding, R.; Zhan, D. Enhancing Representation of Deep Features for Sensor-Based Activity Recognition. *Mob. Netw. Appl.* **2021**, *26*, 130–145. [\[CrossRef\]](#)
18. Bruna, J.; Zaremba, W.; Szlam, A.; LeCun, Y. Spectral networks and deep locally connected networks on graphs. In Proceedings of the 2nd International Conference on Learning Representations, Banff, AB, Canada, 14–16 April 2014.
19. Defferrard, M.; Bresson, X.; Vandergheynst, P. Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering. In Proceedings of the Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, Barcelona, Spain, 5–10 December 2016.
20. Mondal, R.; Mukherjee, D.; Singh, P.K.; Bhateja, V.; Sarkar, R. A New Framework for Smartphone Sensor-Based Human Activity Recognition Using Graph Neural Network. *IEEE Sens. J.* **2021**, *21*, 11461–11468. [\[CrossRef\]](#)

21. Mohamed, A.; Lejarza, F.; Cahail, S.; Claudel, C.; Thomaz, E. HAR-GCNN: Deep Graph CNNs for Human Activity Recognition From Highly Unlabeled Mobile Sensor Data. In Proceedings of the 2022 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops), Pisa, Italy, 21–25 March 2022; pp. 335–340. [\[CrossRef\]](#)
22. Yang, P.; Yang, C.; Lanfranchi, V.; Ciravegna, F. Activity Graph Based Convolutional Neural Network for Human Activity Recognition Using Acceleration and Gyroscope Data. *IEEE Trans. Ind. Inform.* **2022**, *18*, 6619–6630. [\[CrossRef\]](#)
23. Nian, A.; Zhu, X.; Xu, X.; Huang, X.; Wang, F.; Zhao, Y. HGCNN: Deep Graph Convolutional Network for Sensor-Based Human Activity Recognition. In Proceedings of the 2022 8th International Conference on Big Data and Information Analytics (BigDIA), Guiyang, China, 24–25 August 2022; pp. 422–427. [\[CrossRef\]](#)
24. Li, Q.; Han, Z.; Wu, X.-M. Deeper insights into graph convolutional networks for semi-supervised learning. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; Volume 32.
25. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
26. Pan, S.J.; Tsang, I.W.; Kwok, J.T.; Yang, Q. Domain adaptation via transfer component analysis. *IEEE Trans. Neural Netw.* **2010**, *22*, 199–210. [\[CrossRef\]](#)
27. Tzeng, E.; Hoffman, J.; Zhang, N.; Saenko, K.; Darrell, T. Deep domain confusion: Maximizing for domain invariance. *arXiv* **2014**, arXiv:1412.3474.
28. Banos, O.; Garcia, R.; Holgado-Terriza, J.A.; Damas, M.; Pomares, H.; Rojas, I.; Saez, A.; Villalonga, C. mhealthdroid: A novel framework for agile development of mobile health applications. In Proceedings of the Ambient Assisted Living and Daily Activities: 6th International Work-Conference, IWAAL 2014, Belfast, UK, 2–5 December 2014; pp. 91–98.
29. Reiss, A.; Stricker, D. Introducing a new benchmarked dataset for activity monitoring. In Proceedings of the 2012 16th International Symposium on Wearable Computers, Newcastle, UK, 18–22 June 2012; pp. 108–109.
30. Yan, Y.; Chen, D.; Liu, Y.; Zhao, J.; Wang, B.; Wu, X.; Jiao, X.; Chen, Y.; Li, H.; Ren, X. Tnda-har. *IEEE Dataport* **2021**. [\[CrossRef\]](#)
31. Tang, Y.; Zhang, L.; Min, F.; He, J. Multiscale deep feature learning for human activity recognition using wearable sensors. *IEEE Trans. Ind. Electron.* **2022**, *70*, 2106–2116. [\[CrossRef\]](#)
32. Nguyen, D.A.; Pham, C.; Le-Khac, N.A. Virtual Fusion with Contrastive Learning for Single Sensor-based Activity Recognition. *arXiv* **2023**, arXiv:2312.02185.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.