IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2023.1120000

Enhancing Human Activity Recognition in Wrist-Worn Sensor Data Through Compensation Strategies for Sensor Displacement

HUI WANG[1], XIN WANG[2], (Student Member, IEEE), CHENGGANG LU[3], MENGHAO YUAN[1], YAN WANG[1], (Member, IEEE), HONGNIAN YU[4], (Senior Member, IEEE), HENGYI LI[5], (Member, IEEE)

[1]School of Automation and Electrical Engineering, Zhongyuan University of Technology, 450007, Zhengzhou, China
[2]Graduate School of Science and Engineering, Ritsumeikan University, 5250054, Kusatsu, Japan
[3]School of Intelligent Sensing and Instrumentation, Zhongyuan University of Technology, 450007, Zhengzhou, China
[4]School of Computing, Engineering and Built Environment, Edinburgh Napier University, EH10 5DT, Edinburgh, U.K.
[5]Research Organization of Science and Technology, Ritsumeikan University, 5250054, Kusatsu, Japan

Corresponding author: Xin Wang (e-mail: gr0663vk@ed.ritsumei.ac.jp).

This work was supported by the Key Technologies R&D Program of Henan Province (232102211020, 222102210016), the National Natural Science Foundation (62073297), and in part by the Backbone Teacher Support Program (GG202414)

ABSTRACT
Human Activity Recognition (HAR) using wearable sensors, particularly wrist-worn devices, has garnered significant research interest. However, challenges such as sensor displacement and variations in wearing habits can affect the accuracy of HAR systems. Two compensation stratigies for sensor displacemnt are proposed to address these issues. The first strategy is hybrid data fusion, which involves merging sensor data collected from different displacement locations on the wrist. This technique aims to mitigate the discrepancies in data distribution that result from the multiple wearing positions along the wrist, thereby enhancing the overall accuracy of HAR models. The second strategy is cross-location transfer fine-tuning, which involves pretraining a model with data from typical wrist locations and then fine-tuning it with data from a new sensor location. This approach improves the model's ability to adapt and perform accurately when the sensor is placed in a different position, significantly enhancing its performance and generalization capabilities. To verify the effectiveness of these proposed compensation strategies, we built an LSTM baseline model and introduce a new Multi-stage Feature Extraction (MSFE) model that integrates 1D CNN and attention. Experiments on common activities such as walking, standing, using stairs, and lying down, with data recorded at multiple locations along the wrist, have shown that both hybrid data fusion and cross-location transfer fine-tuning strategies notably improve the recognition accuracy of HAR models. The proposed MSFE model achieves higher recognition accuracies than the LSTM model in all six experimental scenarios, particularly in Scenario 5 involving sensor displacement, with an improvement of up to 31.65%. Additionally, thecross-location transfer fine-tuning strategy enhances the recognition accuracy by 9.19% for Subject 3 with sensor displacement at the right wrist location. These advancements in handling sensor displacement and wearing variations are crucial for developing more reliable and versatile wearable technologies.

INDEX TERMS Sensor displacement compensation, human activity recognition, deep learning, wrist

I. INTRODUCTION

THE internet of things infrastructure enables the scalable deployment of wearable devices and sensors, such as smartwatches, wristbands, and others. This technology holds significant potential for enhancing the way we understand, monitor, and interact with human activities in various contexts [1]. Wearable sensor-based Human Activity Recognition (HAR) is a field that involves using sensors in wearable devices to monitor and interpret human activities. These devices, typically worn on the body, are equipped with various sensors, including accelerometers, gyroscopes, magnetometers, and sometimes additional sensors like heart rate monitors [2]. HAR is applicable in various fields due to its ability to analyze human activities using sensor data, encompassing health and fitness, security and surveillance, sports science, and assistive technologies [3].

The study in [4] highlighted that users' natural placement of sensors can significantly impact the recognition capabilities of the system. While multi-sensor configurations were effective in mitigating variations introduced by sensor deployment, not all models demonstrated equal tolerance to sensor displacement. Particularly, feature fusion methods experienced a notable decline in performance when users personally placed the sensors. This decrease in performance was even more pronounced when sensors were intentionally misplaced. Feature fusion multi-sensor models were especially sensitive to displacement effects, as changes in the signal space were naturally integrated into the aggregated feature vectors. Neither single-sensor nor feature fusion multi-sensor methods exhibited acceptable recognition capabilities when sensor displacement occurred.

The authors in [5] explored the feasibility of detecting the current on-body location of a wearable device in real-world scenarios using a single acceleration sensor for common activities. To assess the significance of location information, they created a substantial real-world dataset by recording seven different on-body locations of 15 participants engaged in eight distinct activities. Their experimental results strongly supported that incorporating on-body location information significantly enhanced activity recognition accuracy. Yet, their work treated the multiple locations as a multi-class problem without further investigating compensation for sensor drifting in the location-unaware scenarios. In [6], a proposed solution addressed the challenge of location-independent human activity recognition using wearable sensors. The researchers developed a set of linear and non-linear transformations for sensor data to minimize the inertial sensor's sensitivity to location and orientation, improving recognition performance in real-time scenarios.

Wrist is a commonly chosen location for HAR, well received by users [7]–[10]. Data from wrist-worn sensors provide rich movement information and are extensively used in recognizing periodic repetitive activities like step counting, calorie consumption, and heart rate measurement [11], [12], which are generally less sensitive to sensor displacement. However, for the recognition of more complex activities like walking, climbing stairs, or sitting, a greater amount or richer information from sensors is required. Many HAR studies based on wrist-worn sensors typically secure the sensor at a predetermined wrist location during data collection. Nevertheless, the loose placement of sensors or varying individual wearing habits can lead to relative movements between the sensor and the wrist. Moreover, wearers may place the sensors differently from the predefined location, further complicating accurate activity recognition. Using a model trained with data from the initially defined wrist location directly to recognize activities after sensor displacement can significantly degrade recognition performance.

Most wearable HAR systems operate under the assumption that training and testing data share the same distribution. However, in real-world scenarios, the distribution of sensor data is influenced by the locations where users wear the sensors. This disparity in distribution can result in a decline in model performance or a reduction in generalization [13]. Sensor displacement occurs when sensors are not in their ideal or predefined locations, which can involve rotation, translation, or both [4]–[6]. This displacement often accompanies a drift in the original data distribution, affecting the entire activity recognition process [14]. Such drift can lead to the failure of activity recognition systems designed for specific sensor locations. Imposing restrictions on how users wear wearable devices in real-world scenarios is impractical and may diminish user enthusiasm. To address the performance variations caused by sensor displacement, different strategies have been explored. The authors in [15] experimented with smartphone-embedded sensors placed in pockets, bags and hands to compare the performance differences in recognizing the same activity. They integrated sensor data from these three locations and extracted features to establish a location-independent activity recognition system with an accuracy of 80%. Yet, the system still required data from three different body locations. Similarly, the study in [16] employed recognition models according to different sensor locations to identify daily activities. Huang et al. [17] proposed a method capable of identifying the smartphone's holding orientation by selecting an appropriate model for activity state recognition, effectively addressing the issue of varying recognition accuracies stemming from different sensor placement locations.

The work highlighted a significant impact on recognition results due to sensor displacement on the wrist, especially when training and testing data originate from different locations along the wrist [7]. As depicted in Figure 1, the visualization of principal component analysis (PCA) shows that data from both subjects demonstrate distinct distributions based on four wrist-wearing locations. Developing a new classification model for each distribution would entail a substantial investment of time and computing resources. The key to mitigating the decline in recognition performance caused by sensor displacement is enabling the model to adapt to the data after displacement. Therefore, compensating for sensor displacement becomes crucial for maintaining robust recognition performance. Yang [18] conducted a comprehensive comparison of the impact of device locations on smartphone-based HAR. They eliminated differences in raw sensor data resulting from diverse device orientations through coordinate transformation. The effectiveness of using acceleration sensors for HAR was highly dependent on the smartphone-embedded accelerators' wearing location.

Deep learning has greatly enhanced the efficiency of feature extraction for HAR [19], [20]. Networks like Convolutional Neural Networks (CNNs) and Long Short-Term Memory networks (LSTM) are capable of automatically extracting valuable features from raw sensor data [21]–[24]. LSTM, in particular, excel in handling time series tasks. Wang et al. [25]

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3422256

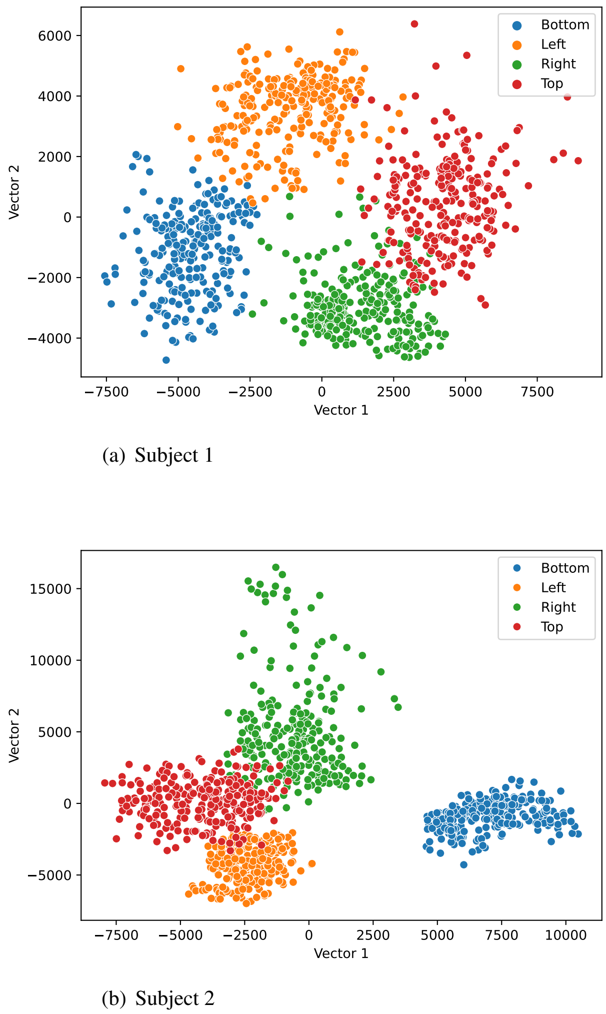


FIGURE 1. PCA dimensionality reduction visualization of data at different wrist positions

introduced a hierarchical deep model with two hidden layers of LSTM, leveraging LSTM's ability to memorize and store periods of time series data, thereby enhancing HAR accuracy. Chen Qing applied an LSTM model for gait sequence prediction to identify falls [26]. Tan Huixing utilized an improved LSTM model [27] to detect toe-hitting and heel-hitting actions in users' gait data collected from accelerometers. Also, CNNs have been employed to enhance HAR systems' performance by directly working with raw accelerometer and gyroscope data. For instance, Chen et al. [28] demonstrated the effectiveness of deep CNN in improving HAR accuracy, highlighting their ability to automatically learn relevant features from multi-channel sensor inputs.

On the other hand, attention mechanisms have been introduced to further enhance deep learning models' performance in HAR. By dynamically weighting the importance of different time steps, attention mechanisms enable the model to capture the most significant features for distinguishing between different activities [29]. A notable example is the work by [30], who applied attention-based models to HAR, demonstrating improved performance compared to traditional CNN and LSTM approaches.

Transfer learning stands as a crucial technique in deep learning, effectively addressing the challenge of distribution bias [31], [32]. Its principal objective is to expedite the application of existing algorithms, models, or parameters to new tasks. By minimizing the discrepancy between the source and target domains, transfer learning enhances models' accuracy during testing. Various techniques, such as sample weight transfer, feature transformation transfer, and model pre-training, are used to achieve domain adaptation. While transfer learning has been extensively applied in computer vision (ImageNet model) [33] and natural language processing (BERT model) [34], there currently exist no established general models tailored for time series HAR tasks.

Current studies on sensor displacement have predominantly focused on the impact of larger displacements across the entire body or employed traditional machine learning methods. There is a notable gap in research regarding the impact and compensation for smaller sensor displacements in specific body parts, particularly when employing deep learning approaches. Consequently, this study aims to address this gap by concentrating on compensating for smaller displacements of wrist-worn sensors. We gather activity data from wrist sensors and construct scenarios representing small displacements of the wrist-based sensors. Two sensor compensation schemes are proposed: one involves fusing data from all sensor locations, and the other employs cross-location transfer fine-tuning. To assess the effectiveness of these schemes, an LSTM and a Multi-stage Feature Extraction (MSFE) recognition model are developed. The primary contributions of this paper are outlined as follows:

(1) A Multi-stage Feature Extraction (MSFE) model for HAR is developed, consisting of three key modules: a sensor axis-wise attention layer, a self-attention module, and a CNN temporal module. These modules are designed to assign weights to sensor channels, capture time step representations, and extract local time-dependent features, respectively.
(2) Six distinct scenarios are devised to assess the impact and compensation of wrist sensor displacement. Scenario 1 evaluates recognition performance for each subject in four defined wrist-wearing locations sequentially, with no displacement. Scenario 2 involves mixing each subject's data from the four locations in advance for training and testing, assessing the model's general performance. In Scenario 3, data from all subjects in the same location are mixed to examine the model's cross-subject recognition performance. Scenario 4 combines all data from all subjects in the four locations for training and testing, validating the model's cross-subject and cross-location recognition performance. Scenario 5 Tests the model trained on three locations with data from the remaining one location

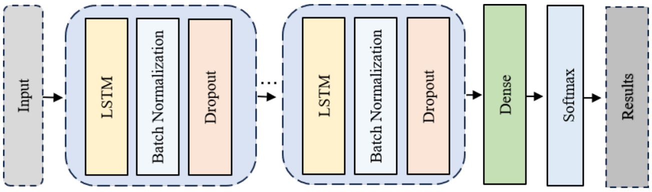


FIGURE 2. Framework of the established LSTM model

simulated more severe sensor displacement. Scenario 6 Evaluates model performance on new subjects using Leave-One-Subject-Out Cross-Validation (LOSO-CV) demonstrated the model's adaptability to new users.

(3) Two compensation strategies are proposed to improve model performance amid sensor displacement. Hybrid data fusion merges sensor data from different wrist displacement locations to mitigate disparities in data distribution. Cross-location transfer fine-tuning pretains a model using data from all conceivable wrist positions, followed by fine-tuning it with data from a newly chosen sensor location, thereby enhancing the model's adaptability and accuracy when the sensor is relocated to a new position.

The rest of this paper is structured as follows. Section II reviews the related work based on deep learning and transfer learning. Section III discusses the presented recognition models, including LSTM and MSFE. SectionIV presents experiments, results, and sensor displacement compensation strategies. The conclusion in Section V synthesizes the main findings and suggests future research directions in HAR.

II. RELATED WORKS

With the proliferation of wearable devices like smartwatches and fitness trackers, there is an increasing demand for accurate and robust HAR systems. Recent advancements in deep learning have substantially improved HAR accuracy by leveraging models such as CNN, LSTM, and attention mechanisms.

LSTM are particularly useful for modeling temporal dynamics, which is crucial for recognizing activities characterized by specific time-based patterns. Chen [35] proposed an adaptive structure model incorporating deep learning networks, including CNN and LSTM, with the goal of enhancing the accuracy of HAR based on accelerometer sensors with non-fixed wearing locations. Moreover, attention mechanisms have become integral in various deep learning applications [36], [37], spanning natural language processing, image recognition, and speech recognition. The attention mechanism enables the model to concentrate on the most relevant parts of the input data, dynamically assigning weights to different parts of the input data. This assists the model in capturing crucial features while disregarding irrelevant information, thus resulting in improved performance. Chen et al. [30] utilized LSTM to achieve superior performance in HAR by effectively handling the time dependencies inherent in wearable sensor data . Their study underscored the importance of capturing temporal features to accurately distinguish between various human activities.

Among various deep learning approaches, CNN have gained prominence for their ability to automatically extract meaningful features from raw sensor data. Specifically, 1D CNN models have shown considerable promise in wearable HAR due to their capability to process time-series data effectively. 1D CNN applies convolutional operations along one dimension, making them well-suited for analyzing sensor signals that vary over time. For instance, Wang et al. [38] utilized 1D CNN to develop a robust HAR system that processes accelerometer and gyroscope data from wearable devices. Their approach focused on optimizing the network architecture to balance performance and computational cost, achieving high accuracy in real-time activity recognition. Similarly, Chen et al. [28] proposed a 1D CNN-based framework for HAR that integrates multiple sensor modalities, including accelerometers and magnetometers. Their model effectively combines the strengths of different sensors, enhancing the overall recognition performance.

Attention mechanisms have also been incorporated into 1D CNN architectures to enhance their ability to focus on the most relevant parts of the input sequence. Khan et al. [39] proposed an attention-based multi-head model for HAR. This model includes three lightweight convolutional heads, each utilizing a 1D CNN to extract features from sensory data. The attention mechanism is integrated into the lightweight multi-head model to enhance the CNN's representation ability, enabling the automatic selection of salient features while suppressing irrelevant ones. The experimental results on two publicly available benchmark datasets, WISDM and UCI HAR, demonstrated the proposed framework's effectiveness in activity recognition, achieving higher accuracy and ensuring computational efficiency. Essa et al. [40] proposed two novel architectures for classifying sequences of human activity data from different sensors: the Convolution with Self-Attention Network (CSNet) and the Temporal-Channel Convolution with Self-Attention Network (TCCSNet). CSNet captures both local and global dependencies in the input data using 1D CNN convolution and self-attention. TCCSNet exploits temporal and inter-channel dependencies through two branches of 1D CNN convolutions and self-attentions, extracting time-wise and channel-wise information.

Despite these advancements, challenges remain in addressing the variability introduced by sensor displacement and different wearing habits. The study in [4] highlighted that users' natural placement of sensors can significantly impact the recognition capabilities of the system. While multi-sensor configurations were effective in mitigating variations introduced by sensor deployment, not all models demonstrated equal tolerance to sensor displacement. Feature fusion methods, in particular, experienced a notable decline in performance when users personally placed the sensors. This decrease in performance was even more pronounced when sensors were intentionally misplaced. Feature fusion multi-sensor models were especially sensitive to displacement effects, as changes in the signal space were naturally integrated into the aggregated feature vectors.

Neither single-sensor nor feature fusion multi-sensor methods exhibited acceptable recognition capabilities when sensor displacement occurred. The authors in [5] explored the feasibility of detecting the current on-body location of a wearable device in realworld scenarios using a single acceleration sensor for common activities. To assess the significance of location information, they created a substantial real-world dataset by recording seven different on-body locations of 15 participants engaged in eight distinct activities. Their experimental results strongly supported that incorporating on-body location information significantly enhanced activity recognition accuracy. Yet, their work treated the multiple locations as a multi-class problem without further investigating compensation for sensor drifting in the location-unaware scenarios. In [6], a proposed solution addressed the challenge of location independent human activity recognition using wearable sensors. The researchers developed a set of linear and nonlinear transformations for sensor data to minimize the inertial sensor's sensitivity to location and orientation, improving recognition performance in real-time scenarios.

A study in [41] highlighted the benefits of using transfer learning to adapt HAR models to varying sensor locations, thereby improving recognition accuracy. Subsequent research has continued to explore and validate these findings. For instance, Soleimani et al. [42] introduced a novel adversarial knowledge transfer method called Subject Adaptor GAN (SA-GAN), which employed the Generative Adversarial Network framework to facilitate cross-subject transfer learning in wearable sensor-based HAR. In some instances, it achieved up to 90% of the accuracy attainable through supervised training on the same domain data. Bursa [43] explored using transfer learning to create personalized models and model compression techniques to enable efficient DL algorithms on resource-constrained devices. They investigated the effects of various deep learning architectures, the extent of layer fine-tuning, the amount of user-specific training data, and transferring models to new datasets on HAR performance. The study compared the performance of transferred models with general and user-specific models in terms of F1 score, training time, and inference time.

In response to the sensor displacement issues in wearable HAR, this paper introduces two innovative compensation schemes: hybrid data fusion and cross-location transfer fine-tuning, aiming to enhance the accuracy of HAR especially using wrist-worn sensors. The proposed strategies integrate data from multiple locations and fine-tune deep learning models to adapt to a new sensor location. By harnessing the strengths of 1D CNNs and combining them with sensor axis-wise attention and self attention mechanisms, we also introduce a new Multi-stage Feature Extraction (MSFE) model to verify the efficacy of the compensation strategies.

III. RECOGNITION MODELS

A. MULTI-LAYER LSTM

In this study, the developed Multi-stage Feature Extraction model and a foundational multi-layer LSTM model [44] are employed to assess the proposed compensation schemes. The architecture of the LSTM model is depicted in Figure 2. It consists of three LSTM blocks with hidden units of 32, 128, and 64, respectively. Each block contains an LSTM layer, a batch normalization layer and a dropout layer, facilitating the training process and mitigating overfitting to some extent. Ultimately, two fully connected (dense) layers, with 64 and 4 units respectively, along with the softmax operation, yield the recognition results.

B. MULTI-STAGE FEATURE EXTRACTION MODEL (MSFE)

The Multi-stage Feature Extraction model (MSFE) introduced in this paper integrates the standard attention mechanism, self-attention mechanism and 1D CNN to extract more informative features from activity time series data. The MSFE is composed of three components: the sensor axis-wise attention layer, the self-attention module, and the CNN temporal module. Its structure is illustrated in Figure 3.

The sensor axis attention layer is employed to determine the attention weights for all sensor axes, indicating the significance of each axis's collected data for activity recognition. Subsequently, a one-dimensional convolution transforms the weighted representation into a vector along the time dimension. Following the convolution, a positional encoding layer is introduced for the time window, enabling the model to recognize the sequence order of the input time series and better comprehend their correlations.

The self-attention module focuses on learning the feature of the positional encoding data through scaled dot product operations. The CNN temporal module enhances the extraction of local time-dependent features from the outputs of the self-attention module, thereby obtaining the final feature representation. Ultimately, the fully connected layers, followed by a softmax layer, produce the probability for the final label category. The detailed principles of the three attention layers are elucidated below.

- Sensor axis-wise attention layer

The data gathered from wrist-worn sensors consist of multi-axis time series. Each axis exhibits varying sensitivity to activities, making a distinct contribution to the recognition of the target activity. The sensor axis-wise attention layer is crafted to discern attention weights for the provided axis data, thereby establishing the significance of each axis in HAR. A higher weight is assigned to an axis deemed more relevant to the target activity. The acquired attention weights are expressed as illustrated in equations (1) and (2).

$$s_k^{(t_i)} = \frac{exp\left(a_k^{(t_i)}\right)}{\sum_k exp\left(a_k^{(t_i)}\right)} \tag{1}$$

$$v_i = \sum_i s_k{}^{(t_i)} a_k{}^{(t_i)} \tag{2}$$

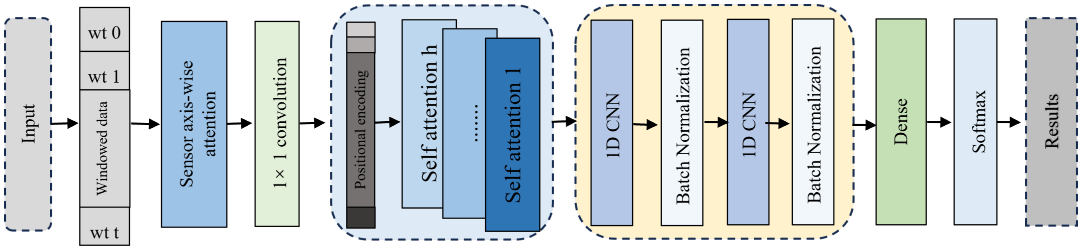


FIGURE 3. Framework of the established MSFE model

where k denotes the individual sensor axes, $a_k^{(t_i)}$ represents the sensor axis data at time t_i, $s_k^{(t_i)}$ stands for the attention weight for axis k, and the v_i parameter is the weighted representation produced by the sensor axis-wise attention layer.

- Self-attention module

The self-attention mechanism allocates distinct weights to each time step based on its similarity to other time steps. It then uses these weights to conduct matrix transposition between each time step and other relevant time steps to derive attention weights. In self-attention, the query, key, and value are the transformation vectors obtained by linearly transforming the input through a fully connected layer. The dot product of the query and key can be viewed as a comparison between each time step and others. The attention score for a given time step is determined by scaling and softmax normalizing the dot product value. The weighted representation of each time step is computed from the attention score and the value vector. These operations are executed concurrently through matrix-based operations, as depicted in Equation (3), where q, k, v refers to the query, key, and value, respectively, $f_{sa}^{(h_j)}$ is the output of the self-attention module j.

$$f_{sa}^{(h_j)}(q,k,v) = softmax\left(\frac{q \cdot k^T}{\sqrt{d_k}}\right) v \tag{3}$$

The MSFE utilizes multiple attention heads to better capture the input data, enabling the extraction of more valuable features. Each attention head generates unique outputs while computing queries, keys, and values in Equation (3). Subsequently, the output from each attention head is concatenated and transformed into the same dimension as a single attention head through matrix-wise operations outlined in Equation (4).

$$s_{mha} = W_o \cdot concat\left(f_{sa}^{(h_1)}, f_{sa}^{(h_2)} \ldots, f_{sa}^{(h_{n-1})}, f_{sa}^{(h_n)}\right) \tag{4}$$

where W^o is the parameter matrix learned during training, $f_{sa}^{(h_1)}, f_{sa}^{(h_2)} \ldots, f_{sa}^{(h_{n-1})}, f_{sa}^{(h_n)}$ represents the output of each self-attention block, and n is the number of self-attention blocks.

- The CNN temporal module

The CNN temporal module consists of two 1D CNN layers, each followed by a BatchNormalization layer. In the 1D CNN layer, the filter is moved along the direction of the time step thus extracting local time-dependent features among adjacent time step. This process enhances the extraction of temporal features from sensor time series signals. The Batch Normalization layer is introduced to significantly accelerate model training and enhance model stability by normalizing the input layer's features, reducing internal covariate shift.

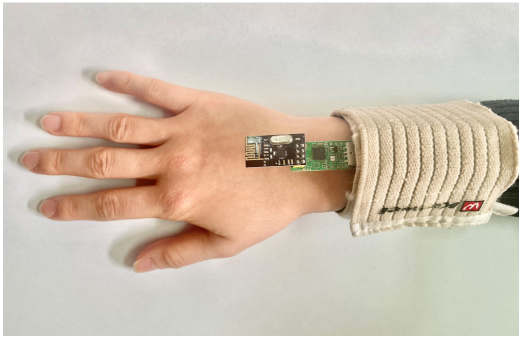


FIGURE 4. The wrist-worn sensor device

IV. EXPERIMENTS AND RESULTS

A. WRIST BEHAVIOR DATA

The defined activity data were gathered using a wrist-worn device equipped with an accelerometer, a gyroscope, and a magnetometer, as depicted in Fig. 4. This device was employed to construct a Wrist-Worn Human Activity Recognition (WWHAR) dataset. The WWHAR dataset encompasses data from five healthy individuals engaging in four types of daily activities: lying, climbing stairs, standing, and walking, with specific activity definitions outlined in Table 1. The dataset provides 12 channels of time series data, including Pitch, Yaw, Roll, three-axis accelerometer data (ACCx, ACCy, ACCz), three-axis gyroscope data (GYROx, GYROy, GYROz), and three-axis magnetometer data (Magx, Magy, Magz). Partially normalized raw data is depicted in Fig. 5, where "Channel" represents the data channel, "Sample" denotes the time sample number, and "Value" indicates the normalized value.

During the data collection process, the wearable device was securely placed on the dominant wrist of each user at four predefined locations: Top, Left, Right, and Bottom of the wrist, as shown in Fig. 6. Each participant underwent four sets of experiments based on the designated wearing locations, performing the four specified activities in each experiment set. The collection time for each activity was approximately 2.5 minutes. To mitigate the unwanted activities at the beginning and end of each round, the first and last 15 seconds of data series were excluded, generating an approximately two minutes of valid series. These valid data were manually labeled with corresponding activity labels.

TABLE 1. Activities defined in this study

Activity	Description
Lying	Resting on a mat without frequent turns
Stairs using	Ascending or descending stairs
Standing	Standing still
Walking	Walking around

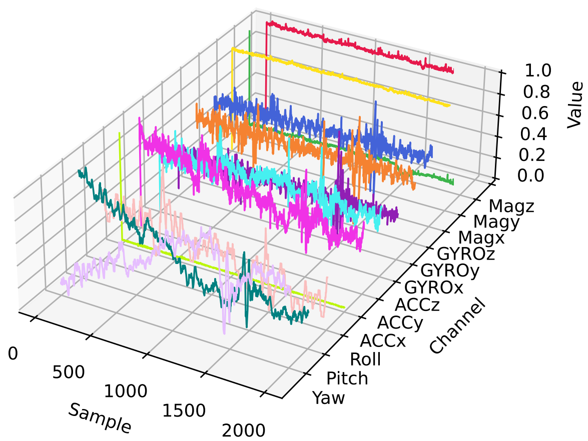


FIGURE 5. Visualization of partial gathered data

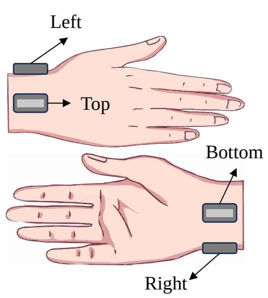


FIGURE 6. The Wrist-Worn Sensor Placement

B. COMPENSATION OF HYBRID DATA FROM ALL SENSOR LOCATIONS

To assess the impact of recognition accuracies following local sensor displacement along the wrist and apply appropriate compensation methods, we have devised six distinct scenarios. These scenarios define four specific wrist locations to mimic sensor displacement. The experiments, conducted across these scenarios, involve the selection of activity data from five participants. This exploration aims to understand the robustness of wrist-worn activity recognition from multiple perspectives, including wearing locations, individual differences among participants, and the overall performance across all participants.

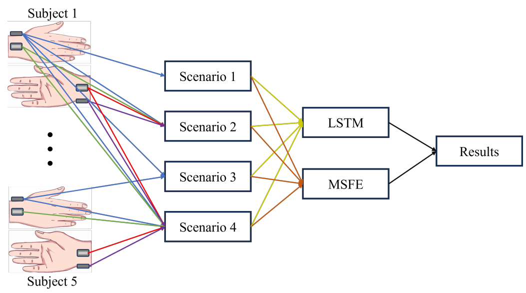


FIGURE 7. The setup of Scenarios 1-4

In Scenario 1, models are trained and tested using data from the same wrist location for each participant, validating the location-wise recognition performance individually. Scenario 2 entails training and testing the model by amalgamating data from all four locations, providing insight into the model's recognition performance across the defined locations for each participant separately. In Scenario 3, data from the same location of five participants is combined, and the model is trained and tested to evaluate its recognition performance across participants. In Scenario 4, all data from the four locations of five participants are combined, and the model is trained and tested to examine its recognition performance across both participants and wearing locations. In Scenario 5, data from three of the four locations are used to train the model and data from the remaining one is used to test the model. This scheme is used to evaluate the recognition performance of the model under extreme sensor displacement conditions. To evaluate the model's performance on new subjects, we employed Leave-One-Subject-Out Cross-Validation (LOSO-CV) in Scenario 6. The model is trained using the activity data from four subjects and tested on the data from the fifth subject. This process is repeated for each subject, and the final performance is the average result. The specific experimental settings are depicted in Figures 7 and 8.

In the six scenarios, each dataset undergoes windowing with a 50% overlap, employing a window width of 30. Each time window has a size of (30, 12), where the label for each window is assigned based on the most frequently occurring label among the 30 samples within the window. Each dataset is divided into training and testing sets in a 7:3 ratio. The training set has a size of (0.7#, 30, 12), while the testing set has a size of (0.3#, 30, 12), where # denotes the total number of time windows in the dataset. Through comprehensive comparative experiments and analysis, we find that the recognition performance of the model in the Multi-stage Feature Extraction model is optimal when the number of self-

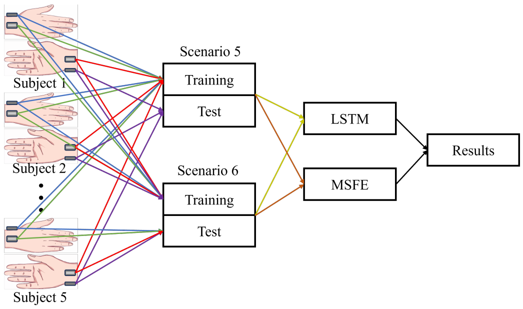


FIGURE 8. The setup of Scenarios 5-6.

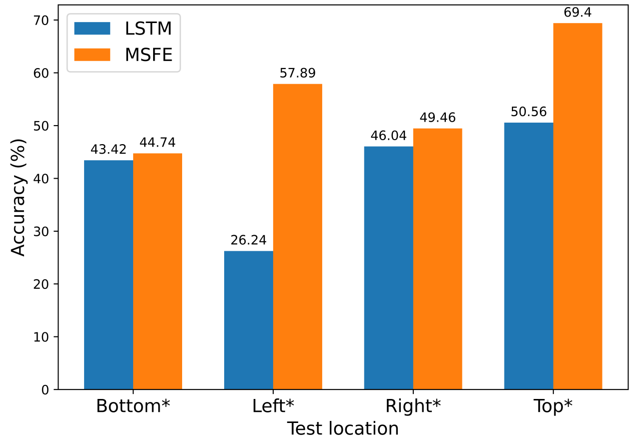


FIGURE 9. Results of the LSTM and MSFE models in Scenario 5

attention blocks is set to 4. Additionally, the value of d_k is set to 32 based on the input dimensionality of the wrist dataset.

Based on three rounds of cross-validation, the recognition accuracies of MSFE and LSTM in the first four experimental scenarios are presented in Table 2. Different background colors are used to highlight the results of Scenarios 1-4, aligning with the experimental scenarios. The light-blue region in the top left corner corresponds to the results of Scenario 1, the light-yellow region in the bottom left corner corresponds to the results of Scenario 2, and the light-green and light-grey regions in the top right and bottom right corners respectively represent the results of Scenarios 3 and 4. The recognition accuracies of the two models in each experimental scenario are compared, with higher recognition accuracies highlighted in bold.

The MSFE model achieves superior recognition accuracies in Scenarios 1-4, particularly at Right of Subject 2. The recognition accuracy of the LSTM model is 87.48%, while the MSFE model reaches 95.29%, showing an improvement of 7.81%. The MSFE model also demonstrates a significant improvement in Scenario 2 for Subject 3, with a recognition accuracy of 94.45%, an increase of 6.91% compared to the LSTM model's 87.54%. This improved performance indicates that the MSFE model can extract more robust feature representations from sensor data when the training and test data are from the same locations.

Next, we set up Scenario 5, where the test data and the training data are from different wrist-wearing locations, to explore the model's recognition performance at unknown wrist locations. It simulates to some extent the sensor displacement. Figure 9 presents the results of the LSTM and MSFE models in Scenario 5. Here, Bottom*, Left*, Right* and Top* represent the data at the current location used for testing the model, while data from the remaining three locations serve as the training set. Although the recognition accuracies of both models for unseen data drops drastically, the MSFE model still achieves a higher accuracy. In particular, the MSFE accuracy is 31.65% higher than the LSTM model when the data from the Left location is used for the test set. Despite the fact that wrist-worn devices should try to avoid such subversive displacement phenomenon in practical use, it can still be seen that the proposed MSFE exhibits better robustness in the face of displaced data.

The application of wrist-worn devices often involves new users; therefore, in Scenario 6, we evaluated the model's recognition performance on new subjects using the LOSO-CV criterion, as listed in Table 3. Although the computational complexity of the MSFE model exceeds that of the baseline model LSTM in terms of the parameter count and FLOPs, its LOSO-CV accuracy is 83.09%, which is 7.21% higher than that of LSTM. Currently, the training of the model relies on PCs, and the existing commonly used private PCs are basically able to satisfy the demand for such computational resources. Therefore, substantial performance improvement is the primary need of the model, and the subsequent deployment of the mobile model will fully consider the saving of computational resources.

We also compare the proposed MSFE model with previous work listed in Table 4. Mohsen et al. [45] proposed a k-nearest neighbor (KNN) algorithm, which achieved 90.46% activity recognition accuracy on the publicly available UCI-HAR dataset through parameter fine-tuning. Huang et al. [46] developed a shallow CNN model for HAR tasks, incorporating cross-channel communication within the same layer. This design allowed for comprehensive interaction between all channels, enabling the model to capture more differential features from the sensor input. Imran et al. [47] created the CNN-BiGRU model, which extracts multiscale features of activity data through two CNN modules integrating multiple different kernel sizes and temporal features through a bidirectional GRU layer. Similarly, Khatun et al. [48] proposed a CNN-LSTM model to extract spatial and temporal dependencies from the sensor data, introducing self-attention to enhances the model's ability to focus on relevant parts of the input sequence for better prediction. We evaluated the above models on data from Scenario 4, where the treatment of the data is consistent with the mainstream HAR evaluation approach. As listed in Table 4, the proposed MSFE model achieves the best recognition accuracy.

TABLE 2. Recognition accuracies in the Scenarios 1-4 with LSTM and MSFE

	Subject 1		Subject 2		Subject 3		Subject 4		Subject 5		All Subjects	
	LSTM (%)	MSFE (%)	LSTM (%)	MSFE (%)	LSTM (%)	MSFE (%)	LSTM (%)	MSFE (%)	LSTM (%)	MSFE (%)	LSTM (%)	MSFE (%)
B	91.56	**96.02**	94.37	**95.60**	97.64	**98.77**	96.46	**97.07**	94.49	**98.07**	95.99	**96.20**
L	98.36	**98.63**	96.09	**96.74**	93.00	**94.29**	92.42	**98.25**	94.09	**96.98**	94.35	**94.94**
R	95.59	**98.17**	87.48	**95.29**	83.11	**86.06**	90.91	**92.68**	98.37	**99.09**	92.77	**93.92**
T	97.53	**98.11**	94.86	**95.14**	90.38	**92.19**	91.94	**92.09**	96.85	**97.69**	93.85	**97.23**
All	95.31	**97.00**	93.48	**93.86**	87.54	**94.45**	93.64	**94.60**	95.23	**98.13**	94.06	**95.18**

B: Bottom; L: Left; R: Right; T: Top; All: Four locations

TABLE 3. The results of the LSTM and MSFE models in Scenario 6

Model	LOSO-CV Acc.(%)	Params. (M)	FLOPs. (M)
LSTM	75.88	**0.14**	**0.016**
MSFE	**83.09**	0.54	30.31

To further investigate the robust feature representations of the wrist activity data, the focus will now be on analyzing the results of the MSFE model in the Scenarios 1-4. For a more straightforward comparison, the results of the LSTM model are excluded from Table 2, and a new Table 5 is generated. Additionally, an 'Average' row is added, representing the average values of the results at the four locations. Table 2 and 5 use the same background colors to differentiate the experimental results of different scenarios, enhancing readability.

Firstly, in Scenario 1, three out of the five subjects achieved the highest recognition accuracy at Left, while the remaining two subjects achieved the highest accuracy at Low and Right. This validates the phenomenon of individual differences among subjects due to personal habits when performing activities. In Scenario 3, the highest recognition accuracy of 97.23% was obtained at Top of the wrist. The MSFE model trained in Scenario 3 could capture common features, acquiring a feature representation that contains characteristics from all subjects.

Secondly, the performance in Scenario 2 was slightly lower than the average accuracy in Scenario 1 (except for Subject 3). This could be attributed to the varied distribution of activity data across different wrist locations, and models trained with mixed data from multiple locations may experience a slight decrease in performance due to this variability. However, the advantage is an enhanced robustness, as evidenced by Subject 3 achieving 86.06% recognition accuracy at Right, which did not significantly impact the overall accuracy of 94.45% in Scenario 2. From the comparative analysis, two measures to enhance the robustness of the wrist-based HAR system during data collection can be derived: 1) increasing the number of subjects to extract richer common activity features, thereby improving the generalization ability of HAR; 2) increasing the local collection locations for wrist activity data to enhance the system's robustness to data distribution biases.

In addition, the recognition accuracy at Right for Subject 3 is relatively low (86.06%). However, when combining data from all four locations, the recognition accuracy achieved is comparable to that of other subjects (94.45%). Upon repeated verification of the original data, we have ascertained that there are no instances of label errors. The reason may be the occurrence of sensor displacement or local displacement when collecting data at Right of the wrist for Subject 3, leading to a change in the distribution of their activity data. Therefore, Scenario 4, where data from all wrist-worn locations are mixed, serves as a compensatory approach to allow the recognition model to preview the features of data after displacement. To address the negative impact of sensor displacement on the performance of the HAR system, this paper proposes the other compensation scheme.

C. CROSS-LOCATION TRANSFER FINE-TUNING FOR SENSOR DISPLACEMENT COMPENSATION

The compensation scheme of mixing data from all locations validates that the model, having early exposure to all possible data that may undergo displacement, acquires more location-specific activity information, thus overcoming the performance degradation when sensor displacement occurs. However, the model needs retraining when predicting activities in new locations. In contrast, transfer learning addresses domain adaptation problems by extracting useful information from data in a related domain and transferring it to the target task. In the compensation strategy based on transfer learning, data from all wrist locations of the five subjects are utilized as the source-domain data to pretrain the MSFE model, while the target model uses shifted sensor data. The focus is on transferring knowledge from the source model to the target model. Since the recognition tasks in both domains are identical, only the fully connected classification layer needs to be fine-tuned.

The data in both domains share the same pre-training layer structure and parameters, with fine-tuning of adding a fully connected layer to enhance the learning of the specific features in target domain. Subsequently, the pretrained parameters of the MSFE model are transferred to the model in the target domain as initial parameters for further training, reducing training difficulty and time consumption. The schematic dia-

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3422256

TABLE 4. Comparative results of the proposed MSFE model with some existing literature works.

Ref.	Model	Accuracy(%)
Mohsen et al. [45], 2021	KNN	67.78
Huang et al. [46], 2021	Shallow CNN model	93.53
Imran et al. [47], 2023	CNN-BiGRU	94.94
Khatun et al. [48], 2022	CNN-LSTM with self-attention	95.00
Proposed	**MSFE**	**95.18**

TABLE 5. Recognition accuracies in the Scenarios 1-4 with MSFE

Location	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	All Subjects
B	96.02%	95.60%	**98.77%**	97.07%	98.07%	96.20%
L	**98.63%**	**96.74%**	94.29%	**98.25%**	96.98%	94.94%
R	98.17%	95.29%	86.06%	92.68%	**99.09%**	93.92%
T	98.11%	95.14%	92.19%	92.09%	97.69%	**97.23%**
Average	**97.73%**	**95.19%**	92.32%	**95.02%**	**97.96%**	**95.57%**
All	97.00%	93.86%	**94.45%**	94.60%	98.13%	95.18%

B: Bottom; L: Left; R: Right; T: Top; All: Four locations

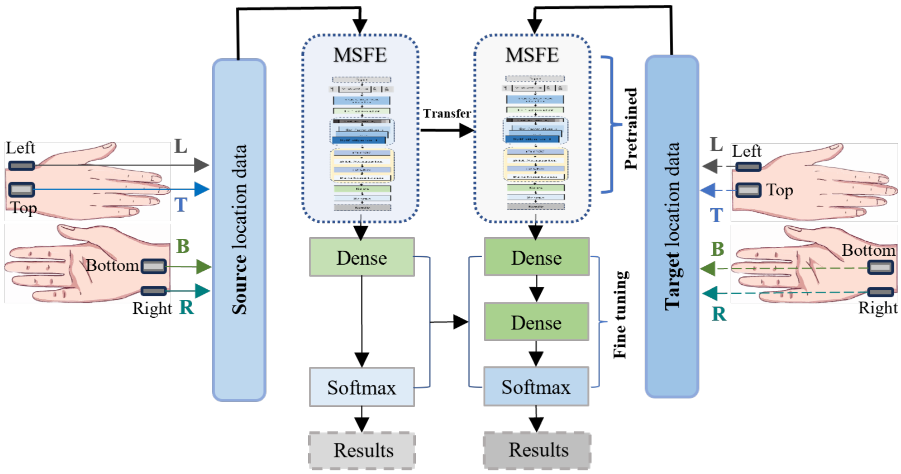
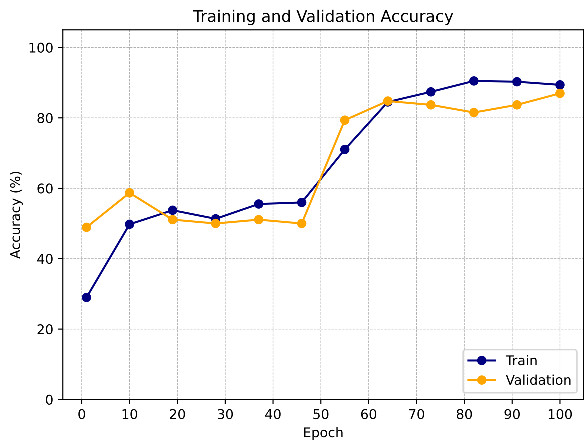


FIGURE 10. Compensation scheme based on transfer fine-tuning

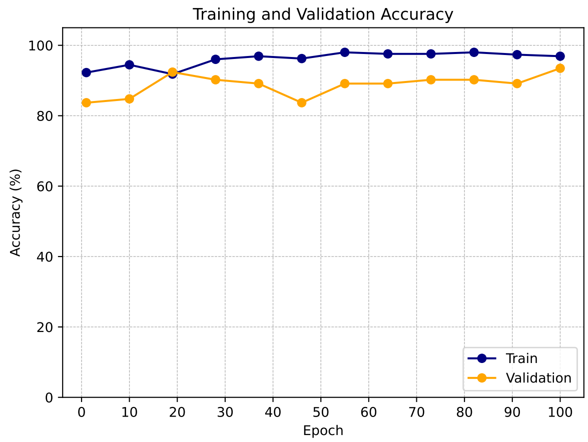
gram is depicted in Figure 10. During pretraining, the MSFE model extracts common features across all subjects and wrist locations. Then, these pretrained parameters are transferred to the locations where local sensor displacement occurs for further training. This allows the MSFE model to continue optimizing performance based on the target domain's activity data, obtaining more robust features while maintaining common features.

Taking Subject 3 as an example, the activity data at Right are employed as the target domain for transfer learning compensation. Figure 11 depicts the accuracy curves for training and validation before and after applying the fine-tuning compensation scheme based on the MSFE model. The accuracy achieved by the MSFE model with transfer learning is 95.07%, marking an improvement of 9.14% over the accuracy of 85.93% obtained without compensation. This underscores the effectiveness of the fine-tuning compensation strategy in significantly mitigating the impact of sensor displacement on the performance of the wrist-based HAR system. Additionally, the MSFE model with transfer learning also exhibits faster convergence, indicating that it can attain higher recognition accuracy in a shorter training time.

Furthermore, we conducted the aforementioned transfer learning compensation on the MSFE model for the Scenario 3 task, aiming to enhance its generalization in this specific context. This transfer scheme also utilizes activity data from all wrist locations of all subjects as the source domain for pre-training MSFE, with the difference that mixed data from the same one location of all subjects as the target domain for fine-tuning and retraining. To rigorously validate the efficacy of this compensation scheme in Scenario 3, we conducted the transfer experiment at the each of the four predetermined wrist locations. Table 6 presents the specific results, wherein 'T', 'R', 'L', and 'B' specifically represent the activity data



(a) Without transfer fine-tuning



(b) With transfer fine-tuning

FIGURE 11. Comparison of accuracy curves before and after applying compensation at Right of Subject 3

from current location of five subjects, utilized as the target domain for the experiment.

Additionally, to mitigate the impact of randomness of the model initialization on the recognition performance, three sets of the transfer experiments on all four wrist locations were executed, denoted as Transfer 1, Transfer 2, and Transfer 3, respectively. The averages of these three sets of cross-validation results are then adopted as the final recognition accuracies following the transfer compensation for the current location. It can be observed that the performance after applying transfer learning has been improved to varying degrees at the four locations, with the highest improvement observed at Left, reaching 2.20%. The MSFE model, trained through transfer learning, obtains more robust feature representations, resulting in an enhanced generalization ability.

TABLE 6. Results of the MSFE model before and after transfer compensation in Scenario 3

Experiment	B(%)	L(%)	R(%)	T(%)
Transfer 1	98.39	97.89	94.46	97.72
Transfer 2	97.85	96.69	94.61	97.58
Transfer 3	98.52	96.84	94.48	97.99
Average	98.25	97.14	94.52	97.76
With no transfer	96.20	94.94	93.92	97.23

B: Bottom; L: Left; R: Right; T: Top

V. CONCLUSIONS

Challenges such as sensor displacement and variations in wearing habits can affect the accuracy of HAR systems, prompting the proposal of two strategies in this paper to address these issues. One approach is hybrid data fusion, which merges sensor data collected from different displacement locations along the wrist. This aims to mitigate the discrepancies in data distribution that result from the sensor being worn in different positions. By integrating data from multiple locations, we create a more robust and consistent dataset for activity recognition, thereby enhancing the overall accuracy of HAR models. Another innovative strategy is cross-location transfer fine-tuning, which involves pretraining a model with data from all typical wrist locations and then fine-tuning it with data from a new sensor location. This approach improves the model's ability to adapt and perform accurately when the sensor is placed at a different position, significantly enhancing its performance and generalization capabilities.

Future work will further focus on testing the compensation strategies at random wearing positions along the wrist to better reflect more real usage. Additionally, we plan to incorporate pruning techniques to compress model parameters, reduce computational complexity, then explore the feasibility of edge deployment.

REFERENCES

[1] A. Yazici, D. Zhumabekova, A. Nurakhmetova, Z. Yergaliyev, H. Y. Yatbaz, Z. Makisheva, M. Lewis, and E. Ever, "A smart e-health framework for monitoring the health of the elderly and disabled," *Internet of Things*, vol. 24, p. 100971, 2023.
[2] B. Fu, N. Damer, F. Kirchbuchner, and A. Kuijper, "Sensing technology for human activity recognition: A comprehensive survey," *Ieee Access*, vol. 8, pp. 83791–83820, 2020.
[3] Y. Wang, S. Cang, and H. Yu, "A survey on wearable sensor modality centred human activity recognition in health care," *Expert Systems with Applications*, vol. 137, pp. 167–190, 2019.
[4] O. Banos, M. A. Toth, M. Damas, H. Pomares, and I. Rojas, "Dealing with the effects of sensor displacement in wearable activity recognition," *Sensors*, vol. 14, no. 6, pp. 9995–10023, 2014.
[5] T. Sztyler and H. Stuckenschmidt, "On-body localization of wearable devices: An investigation of position-aware activity recognition," in *2016 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pp. 1–9, IEEE, 2016.
[6] N. Qamar, N. Siddiqui, M. Ehatisham-ul Haq, M. A. Azam, and U. Naeem, "An approach towards position-independent human activity recognition model based on wearable accelerometer sensor," *Procedia Computer Science*, vol. 177, pp. 196–203, 2020.
[7] X. Wang, Y. Wang, C. Lu, H. Yu, H. He, and Z. Li, "A case study: effect of wrist sensor displacement on har performance using lstm and attention

mechanism," in *2021 International Conference on Advanced Mechatronic Systems (ICAMechS)*, pp. 103–108, IEEE, 2021.

[8] T. G. Pavey, N. D. Gilson, S. R. Gomersall, B. Clark, and S. G. Trost, "Field evaluation of a random forest activity classifier for wrist-worn accelerometer data," *Journal of science and medicine in sport*, vol. 20, no. 1, pp. 75–80, 2017.

[9] F. Masiero, I. Fagioli, L. Truppa, A. Mannini, L. Cappello, and M. Controzzi, "Looking for synergies in healthy upper limb motion: a focus on the wrist," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 1248–1257, 2023.

[10] O. R. A. Almanifi, I. M. Khairuddin, M. A. M. Razman, R. M. Musa, and A. P. A. Majeed, "Human activity recognition based on wrist ppg via the ensemble method," *ICT Express*, vol. 8, no. 4, pp. 513–517, 2022.

[11] P. Alinia, R. Fallahzadeh, C. P. Connolly, and H. Ghasemzadeh, "Paralabel: Autonomous parameter learning for cross-domain step counting in wearable sensors," *IEEE Sensors Journal*, vol. 20, no. 23, pp. 13867–13879, 2020.

[12] S. Levikari, A. Immonen, M. Kuisma, H. Peltonen, M. Silvennoinen, H. Kyröläinen, and P. Silventoinen, "Improving energy expenditure estimation in wrist-worn wearables by augmenting heart rate data with heat flux measurement," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–8, 2021.

[13] Y. Du, J. Wang, W. Feng, S. Pan, T. Qin, R. Xu, and C. Wang, "Adarnn: Adaptive learning and forecasting of time series," in *Proceedings of the 30th ACM international conference on information & knowledge management*, pp. 402–411, 2021.

[14] B. Barshan and A. Yurtman, "Classifying daily and sports activities invariantly to the positioning of wearable motion sensor units," *IEEE Internet of Things Journal*, vol. 7, no. 6, pp. 4801–4815, 2020.

[15] O. Durmaz Incel, "Analysis of movement, orientation and rotation-based sensing for phone placement recognition," *Sensors*, vol. 15, no. 10, pp. 25474–25506, 2015.

[16] D. Rodriguez-Martin, A. Sama, C. Perez-Lopez, A. Catala, J. Cabestany, and A. Rodriguez-Molinero, "Svm-based posture identification with a single waist-located triaxial accelerometer," *Expert Systems with Applications*, vol. 40, no. 18, pp. 7203–7211, 2013.

[17] Y. Huang, H. Lei, R. Zhou, and N. Sang, "Grip mode judgment and motion status recognition based on mobile phone sensors," *Journal of University of Electronic Science and Technology of China*, vol. 46, no. 2, pp. 399–406, 2017.

[18] R. Yang, "Research and application of device location in smartphone behavior recognition," Master's thesis, Nanjing University of Information Science & Technology, 2017.

[19] S. Mohsen, A. Elkaseer, and S. G. Scholz, "Industry 4.0-oriented deep learning models for human activity recognition," *IEEE Access*, vol. 9, pp. 150508–150521, 2021.

[20] S. A. El-Nabi, W. El-Shafai, E.-S. M. El-Rabaie, K. F. Ramadan, F. E. Abd El-Samie, and S. Mohsen, "Machine learning and deep learning techniques for driver fatigue and drowsiness detection: a review," *Multimedia Tools and Applications*, vol. 83, no. 3, pp. 9441–9477, 2024.

[21] H. Li, X. Yue, C. Zhao, and L. Meng, "Lightweight Deep Neural Network from Scratch," *Applied Intelligence*, vol. 53, pp. 18868–19886, August 2023.

[22] W. Huang, L. Zhang, W. Gao, F. Min, and J. He, "Shallow convolutional neural networks for human activity recognition using wearable sensors," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.

[23] M. M. Hassan, M. Z. Uddin, A. Mohamed, and A. Almogren, "A robust human activity recognition system using smartphone sensors and deep learning," *Future Generation Computer Systems*, vol. 81, pp. 307–313, 2018.

[24] C. A. Ronao and S.-B. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks," *Expert systems with applications*, vol. 59, pp. 235–244, 2016.

[25] L. Wang and R. Liu, "Human activity recognition based on wearable sensor using hierarchical deep lstm networks," *Circuits, Systems, and Signal Processing*, vol. 39, pp. 837–856, 2020.

[26] Q. Chen, B. Liang, and J. Wang, "A comparative study of lstm and phased lstm for gait prediction," *Int J Artificial Intelli & App*, vol. 10, no. 4, pp. 57–66, 2019.

[27] H. X. Tan, N. N. Aung, J. Tian, M. C. H. Chua, and Y. O. Yang, "Time series classification using a modified lstm approach from accelerometer-based data: A comparative study for gait cycle detection," *Gait & posture*, vol. 74, pp. 128–134, 2019.

[28] Y. Chen and Y. Xue, "A deep learning approach to human activity recognition based on single accelerometer," in *2015 IEEE international conference on systems, man, and cybernetics*, pp. 1488–1492, IEEE, 2015.

[29] X. Wei and Z. Wang, "Tcn-attention-har: human activity recognition based on attention mechanism time convolutional network," *Scientific Reports*, vol. 14, no. 1, p. 7414, 2024.

[30] K. Chen, D. Zhang, L. Yao, B. Guo, Z. Yu, and Y. Liu, "Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities," *ACM Computing Surveys (CSUR)*, vol. 54, no. 4, pp. 1–40, 2021.

[31] M. Iman, H. R. Arabnia, and K. Rasheed, "A review of deep transfer learning and recent advancements," *Technologies*, vol. 11, no. 2, p. 40, 2023.

[32] J. Wang and Y. Chen, "Introduction to transfer learning," *Technologies*, 2021.

[33] M. A. Morid, A. Borjali, and G. Del Fiol, "A scoping review of transfer learning research on medical image analysis using imagenet," *Computers in biology and medicine*, vol. 128, pp. 104–115, 2021.

[34] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[35] X. Chen, "Research on activity recognition method based on placement-independent ization," Master's thesis, Harbin Institute of Technology, 2019.

[36] J. Hu and W. Zheng, "Multistage attention network for multivariate time series prediction," *Neurocomputing*, vol. 383, pp. 122–137, 2020.

[37] S. P. Singh, M. K. Sharma, A. Lay-Ekuakille, D. Gangwar, and S. Gupta, "Deep convlstm with self-attention for human activity decoding using wearable sensors," *IEEE Sensors Journal*, vol. 21, no. 6, pp. 8575–8582, 2020.

[38] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: A survey," *Pattern recognition letters*, vol. 119, pp. 3–11, 2019.

[39] Z. N. Khan and J. Ahmad, "Attention induced multi-head convolutional neural network for human activity recognition," *Applied soft computing*, vol. 110, p. 107671, 2021.

[40] E. Essa and I. R. Abdelmaksoud, "Temporal-channel convolution with self-attention network for human activity recognition using wearable sensors," *Knowledge-Based Systems*, vol. 278, p. 110867, 2023.

[41] Y. Chen, J. Wang, M. Huang, and H. Yu, "Cross-position activity recognition with stratified transfer learning," *Pervasive and Mobile Computing*, vol. 57, pp. 1–13, 2019.

[42] E. Soleimani and E. Nazerfard, "Cross-subject transfer learning in human activity recognition systems using generative adversarial networks," *Neurocomputing*, vol. 426, pp. 26–34, 2021.

[43] S. O. Bursa, O. D. Incel, and G. I. Alptekin, "Personalized and motion-based human activity recognition with transfer learning and compressed deep learning models," *Computers and Electrical Engineering*, vol. 109, p. 108777, 2023.

[44] X. Wang, Y. Wang, C. Lu, H. Yu, H. He, and Z. Li, "A case study: effect of wrist sensor displacement on har performance using lstm and attention mechanism," in *2021 International Conference on Advanced Mechatronic Systems (ICAMechS)*, pp. 103–108, 2021.

[45] S. Mohsen, A. Elkaseer, and S. G. Scholz, "Human activity recognition using k-nearest neighbor machine learning algorithm," in *Proceedings of the International Conference on Sustainable Design and Manufacturing*, pp. 304–313, Springer, 2021.

[46] W. Huang, L. Zhang, W. Gao, F. Min, and J. He, "Shallow convolutional neural networks for human activity recognition using wearable sensors," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.

[47] H. A. Imran, Q. Riaz, M. Hussain, H. Tahir, and R. Arshad, "Smart-wearable sensors and cnn-bigru model: A powerful combination for human activity recognition," *IEEE Sensors Journal*, 2023.

[48] M. A. Khatun, M. A. Yousuf, S. Ahmed, M. Z. Uddin, S. A. Alyami, S. Al-Ashhab, H. F. Akhdar, A. Khan, A. Azad, and M. A. Moni, "Deep cnn-lstm with self-attention model for human activity recognition using wearable sensor," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 10, pp. 1–16, 2022.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3422256

HUI WANG received the master's degree from the School of Physics, Zhengzhou University, in 2011. She is currently a senior lecturer at the School of Automation and Electrical Engineering, Zhongyuan University of Technology, Zhengzhou,China.

XIN WANG (Student Member, IEEE) is currently a Ph.D. student at the Graduate School of Science and Engineering, Ritsumeikan University, Japan. He received his M.S. degree in School of Electric and Information, Zhongyuan University of Technology in 2023. His research interests encompass the Deeping learning, Human Activity Recognition, and Bioinformation analysis.

CHENGGANG LU received the master's degree from the School of Physics, Zhengzhou University, in 2011. He is currently a Lab instructor at the School of Intelligent Sensing and Instrumentation, Zhongyuan University of Technology, Zhengzhou,China.

MENGHAO YUAN received the B.E. degree from the Zhongyuan University of Technology in Zhengzhou, China, in 2019, and he is currently pursuing the master's degree in the School of Automation and Electrical Engineering. His research interests are gait recognition based on deep learning.

YAN WANG (Member, IEEE) received the Ph.D. degree from the Department of Computing, Bournemouth University, Bournemouth, U.K., in 2018. She is currently an Associate Professor at the School of Automation and Electrical Engineering, Zhongyuan University of Technology, Zhengzhou, China. Her major interests include pattern recognition, machine learning with a particular focus on healthcare and robots.

HONGNIAN YU (Senior Member, IEEE) is a Professor at the School of Computing, Engineering and Built Environment, Edinburgh Napier University, Edinburgh, U.K. He has authored more than 200 journal and conference research papers. His research covers the two main areas: 1) robotics with applications in the rescue and recovery operations, and healthcare and 2) ICT enabled healthcare including assistive technologies in supporting elderly and people with dementia, and activity recognition of elderly people. Dr. Yu is a member of the Engineering and Physical Sciences Research Council Peer Review College, a fellow of the Institution of Engineering and Technology, and a fellow of the Royal Society of Arts.

HENGYI LI (Member, IEEE) is currently a Senior Researcher at the Research Organization of Science and Technology, Ritsumeikan University, Japan. He received his Ph.D. degree in Advanced Electrical, Electronic and Computer Systems from Ritsumeikan University in 2023. His research interests encompass the High-Performance Computing, Computer Architecture, Deep Learning, FPGA-based Accelerator Design for Artificial Intelligence (AI), and Edge Computing.

• • •