IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 12 August 2022, accepted 4 September 2022, date of publication 6 September 2022, date of current version 15 September 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3204739

APPLIED RESEARCH

Comparing Cross-Subject Performance on Human Activities Recognition Using Learning Models

ZHE YANG[1], MENGJIE QU[1], YUN PAN[1], (Member, IEEE), AND RUOHONG HUAN[2]
[1]College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China
[2]College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China

Corresponding author: Yun Pan (panyun@zju.edu.cn)

This work was supported in part by the Zhejiang Provincial Key Research and Development Program of China under Grant 2021C03027, and in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LY19F020032.

ABSTRACT Human activities recognition (HAR) plays a vital role in fields like ambient assisted living and health monitoring, in which cross-subject recognition is one of the main challenges coming from the diversity of various users. Although recent studies have achieved satisfactory results in a non-cross-subject condition, the recognition performance has significant degradation under the cross-subject criterion. In this paper, we evaluate three traditional machine learning methods and five deep neural network architectures under the same metrics on three popular HAR datasets: mHealth, PAMAP2, and UCIDSADS. The experimental results show that traditional machine learning approaches are generally more robust to the new subject scenarios under strict leave-one-subject-out cross-validation. Extra analysis indicates that hand-crafted features are one major reason for the better performance of traditional machine learning on cross-subject HAR, while deep learning is more prone to learning subject-dependent features under an end-to-end training process. A novel training strategy for decision-tree-based methods is also proposed in this paper, resulting in an improvement on the random forest model which achieves competitive performance at an average F1-score (accuracy) of 94.49% (95.09%), 91.64% (92.21%), and 92.70% (93.29%) on the three datasets, compared with state-of-the-art solutions for cross-subject HAR.

INDEX TERMS Cross-subject, deep learning, human activity recognition, leave one subject out, traditional machine learning.

I. INTRODUCTION

Human activities recognition (HAR) has been a popular research topic and widely used in the field of ambient assisted living [1], health monitoring [2], human-machine interaction [3], etc. With the significant growth of commercially available wearable devices, HAR using inertial measurement unit (IMU) [4], [5], [6] with the accelerometer, gyroscope, and magnetometer equipped has gained more attention recently on account of the ability to provide a portable, private, continuous, non-invasive, and low-cost recognition service, compared to the vision-based HAR [7] which has some challenges in privacy protection, resource consumption, and blind areas. A typical framework of HAR is shown in Fig. 1, where the general process of the HAR algorithm includes four stages: sensor data acquisition, data pre-processing, offline feature extraction and model training, and online activity classification. In the data acquisition stage, IMU sensors can be found in glasses [8], phones [9], watches or wrist bands [10], chest patches [11], shoes [12], etc., directly reflecting the subject's behavior tightly related to physical locations throughout the body. Since measured signals suffer from inherent sensor drift and subject's unconscious movements, median filter and low-pass filter are common methods for data cleaning in the pre-processing stage to eliminate noisy interference and redundant information [5], [13], [14]. Besides, continuous data segmentation is also necessary for this stage by dividing the signal into sliding windows with or without overlaps [15]. Feature extraction and model training stage plays a vital role to detect significant low-dimension

The associate editor coordinating the review of this manuscript and approving it for publication was Siddharth Tallur.

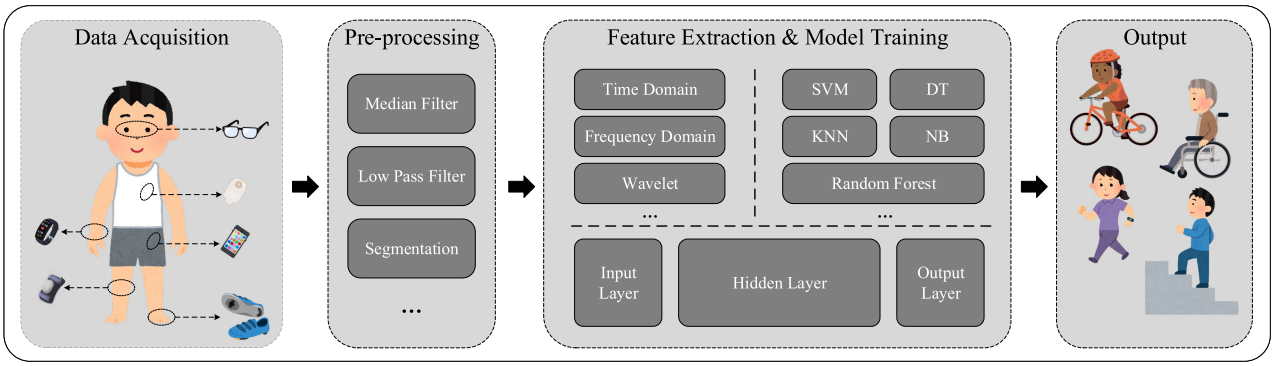


FIGURE 1. The overview of a typical HAR pipeline.

patterns from raw high-dimension sensor input. According to different feature extraction methods, current HAR solutions can be divided into two categories: hand-crafted feature extraction with traditional machine learning like naive Bayes (NB), decision tree (DT), k-nearest neighbor (KNN), support vector machine (SVM), etc., and deep learning using unsupervised features that automatically mined by the machine using an end-to-end training process.

Cross-subject (or inter-subject) recognition is one of the main challenges in HAR research [16], which comes from the limited size of datasets, the diversity of human bodies and habits, and in particular the diversity of devices' wearing modes. Thus, pre-trained models can be significantly user-dependent on the training sets and hard to be promoted to new users in practical applications. Although traditional machine learning and deep learning have achieved satisfactory results in a non-cross-subject (or intra-subject) testing where different samples from the same user appear in both the training set and testing set, the performance has significant degradation in new user scenarios. Most of the current studies on HAR, however, pay less attention to the robustness of the model in cross-subject scenarios and only cover an non-cross-subject test under a given dataset, lacking a standard approach that enables models effectively generalize over heterogeneous datasets performed by different users [17]. They either use data of all subjects indiscriminately for training and testing [18], [19], [20], [21], [22], or only designate one or a group of subjects as fixed testing set [23], [24], which is prone to producing biased evaluation results.

In order to explore the actual cross-subject performance on HAR, in this paper we evaluate three traditional machine learning methods and five deep neural network architectures under the same metrics on three popular HAR datasets: mHealth [25], PAMAP2 [26], [27], and UCIDSADS [28], considering the data size, the similarity and complexity of the activities, and the number of subjects. Hand-crafted features with KNN [4], SVM [5], and random forest [14] are selected as traditional learning frameworks, while the deep learning competitors are convolutional neural network (CNN) [18], [19], long short-term memory (LSTM) [20], [29], and their hybrid variants [21], [22]. These models are surveyed as most common used for HAR by [30], and are trained and tested in this paper under strict Leave One Subject Out (LOSO) cross-validation for a comprehensive examination of cross-subject recognition ability. This paper has the following contributions to the existing studies:

1. This paper has conducted a comprehensive strict cross-subject evaluation of traditional machine learning models and common-used deep learning models in new subject scenarios of HAR applications. We have performed experiments using traditional machine learning and deep learning models on three publicly available datasets, and the impact of hand-crafted features is further analyzed and discussed.

2. A novel training criterion for decision-tree-based learning models is proposed, which tries to discriminate different classes while ignoring the diversity of various subjects. This improvement increases the recognition accuracy of random forest and shows comparable performance with state-of-the-art cross-subject HAR solutions.

The rest of this paper is organized as follows. The related works of this paper are presented in Section II. Section III explains the chosen datasets, evaluation criteria, and the settings of traditional machine learning and deep learning models. Section IV presents the experimental results of different models in cross-subject activity recognition with detailed analysis and discussion. Finally, Section V concludes this paper.

II. RELATED WORK

A. HAR BASED ON TRADITIONAL MACHINE LEARNING AND DEEP LEARNING

Simple time domain and frequency domain features are commonly used in HAR [31], [32], [33] like harmonic mean, standard deviation, Pearson correlation coefficient, etc. These hand-crafted features are trained to build a recognition model like random forest, decision tree, SVM, and KNN as shown in Fig. 1. Casale *et al.* [34] utilized a set of 20 computationally efficient features to recognize 5 basic daily activities. The use of random forest reached a 94% accuracy for recognition, which outperformed the decision tree alone and boosting of trees. With the aid of feature selection and sensor data fusion techniques, Ayman *et al.* [35] were able

to recognize activities on PAMAP2 with a 99.03% accuracy using a random forest classifier. Mekruksavanich *et al.* [36] proposed a framework for recognizing activity based on accelerometer, gyroscope, and surface electromyography data, achieving 99% accuracy using a decision tree model. Arif *et al.* [4] extracted time-domain statistical features from the accelerometer and achieved 97.9% average classification accuracy on the PAMAP2 dataset using the KNN model. Hsu *et al.* [5] proposed a wearable inertial sensor network and an SVM-based behavior recognition algorithm, reaching a recognition rate of 98.23% and 99.55% on 10 common family activities (such as walking, running, up and down stairs, etc.) and 11 sports activities (such as table tennis, badminton, tennis, etc.) respectively. A fast feature dimension reduction method was proposed in [6], which used only 11% of the selected features in the UCIHAR dataset [14], achieving a 98.72% accuracy by random forest classifier. Helmi *et al.* [37] also showed that under properly optimized feature selection methods, SVM classifier can achieved an average accuracy of 98% on UCI-HAR dataset.

Recent advances in deep learning promote the development of deep-feature-based methods, which significantly outperform the hand-crafted features on other learning tasks like object tracking [38], image classification [39], speech recognition [40], etc. One-dimensional [18] and two-dimensional CNN [19] can automatically extract features from IMU for behavior recognition. In 2D CNN cases multiple sequences from multiple sensors are assembled into dynamic images, thus the model will not only considers the dependence within a single temporal signal, but also counts dependencies between signals from different axis and sensors. In order to achieve a significantly reduced execution time while the model performance remained, Gholamrezaii *et al.* [41] proposed a convolutional layer only architecture by removing the pooling layer and adding strides. An ensemble of CNN streams was proposed in [42], and the multi-modal and multi-temporal approach outperformed some state-of-the-art studies. On the other hand, the recurrent neural network (RNN) is another deep model that is often used for speech recognition, natural language processing, and other sequential tasks with various length sequences of inputs, of which the LSTM is a unique structure variant that is suitable for processing and predicting important events with long intervals and delays in time series. Ullah *et al.* [29] proposed a stacked network consisting of five LSTM layers for HAR from smartphone data, with an accuracy of 93.13% achieved in UCIHAR dataset. Hernandez *et al.* [20] improved the distinction between walking up and down stairs using a bi-directional LSTM (BLSTM) network, which can cope with the past and future information of signals.

CNN and RNN have their respective advantages in extracting temporal and spatial features. Therefore many studies have designed hybrid models based on CNN and RNN for better performance on HAR. Ordonez *et al.* [21] proposed a general network framework consisting of a four-layer CNN and a two-layer LSTM for behavior recognition coined as DeepConvLSTM, which achieved a 7.4% and 3.2% performance improvement over the original CNN baseline model in the Skoda [43] and Opportunity dataset [44], respectively. Huan *et al.* [22] proposed a hybrid CNN and BLSTM network based on multi-feature fusion and a novel feature selection method. Experiments on PAMAP2 and UT-data [10] obtained F1-scores at 92.23% and 98.07%, respectively. Lv *et al.* [45] introduced a margin mechanism to enhance the discriminative ability for deep learning, which was proved to be effective for different kinds of deep architectures and their hybrid variants. In addition, Li *et al.* [46] found that the features obtained by hybrid deep-learning architectures involving CNN and LSTM, had advantages to discover both short-term and long-term temporal relationship in the data.

B. CROSS-SUBJECT STUDIES OF HAR

The heterogeneity introduced by different subjects can significantly reduce the accuracy of activity recognition. Ravi *et al.* [47] made an experiment on 2 subjects wearing an accelerometer on the waist and recorded eight daily activities on different dates. They found that over 99% accuracy was achieved on cross-validation when two subjects' data were mixed for training and testing, while only 65% accuracy when the subjects' data were divided and used as either training or testing set. Janidarmian *et al.* [33] evaluated different traditional machine learning methods on HAR using accelerometer data from 14 public datasets containing 8 independent positions and 8 daily activities (walking, running, jogging, biking, standing, sitting, lying, up and down the stairs). In the non-cross-subject 10-fold evaluation, the average classification accuracy of the 8 positions was 96.44%±1.62%, however the number decreased to 79.92%±9.68% in the LOSO cross-subject evaluation.

Recent efforts on cross-subject HAR focus on transfer learning, manifold learning [48], and data augmentation [49]. Transfer learning with domain adaptation and domain generalization have been the most effective method to solve this problem, in which training subject data can be regarded as the source domain, while testing subject data are the target domain. According to whether the target domain data are labeled or not, the domain adaptation can be regarded as supervised and unsupervised.

(1) Supervised domain adaptive method was adopted in [50], [51] to update the pre-trained model with labeled source domain data using few-shot fine-tuning. Akbari *et al.* [52] achieved transfer learning using variational autoencoder (VAE) to identify the vital unlabeled samples and extract domain-invariant features. Since the labels from testing set are leaked, this method is not suitable for ready-to-use HAR solution that must immediately infer activity classes for new subjects without fine-tuning.

(2) Unsupervised domain adaptation aligns the feature distributions between source and target domains by means of distance minimization [53], [54], [55], [56] or generative adversarial networks (GAN) [57], [58], [59], [60]. Hosseini *et al.* [53] designed a BLSTM to extract

representative features and minimize confusion between source and target domains through maximum mean discrepancy loss. Zhang *et al.* [54] proposed a cross-subject adaptive method called gaussian-guided feature alignment as distance minimization metrics. For soft label and coarse-grained problems in class-to-class and set-to-set distribution alignment, a trade-off local domain adaptive method was proposed in [55] as fine-grained cluster-to-cluster distribution alignment between source and target domains. On the other hand, some researchers use GAN to automatically learn the implicit metric function between source and target domain. Soleimani *et al.* [57] took labeled and unlabeled data from different subjects as GAN input. In the training process, the feature extractor and domain discriminator were trained against each other to learn the domain-invariant features. Chakma *et al.* [58] proposed a multi-source adversarial domain adaptive framework to select the most relevant feature from multiple source domains and establish the mapping to the target domain. In unsupervised domain adaptation cases, the original or the summary of training data must be saved in the system to perform a distribution alignment with new targets, thus the occupation of memory increases as the data from new subjects are continuously added to the system.

Meanwhile, the adversarial domain generalization method was used in [17] and [60] for cross-subject recognition. Only the labeled data of training subjects were used to extract domain invariant features which were independent of subjects through adversarial learning, thus the model had good generalization performance on different but similar domains. In this case, the labels from testing set have no leakage and the distribution summary of training set will not be kept in the system, however the model is fixed like traditional machine learning methods and can not be fine-tuned. Once the model needs updating, the system will be re-trained from the beginning using the whole dataset.

C. COMPARISON STUDIES OF HAR

Sensors configuration, datasets selection, window length, testing method, and other factors directly affect the performance of the HAR model in the experiment, so there is no standard comparison benchmark among different studies. Many researchers have conducted comparative studies on existing methods under the same evaluation metric from different perspectives. Wan *et al.* [61] compared the advantages and disadvantages of CNN, LSTM, BLSTM, multilayer perceptron (MLP), and SVM algorithms in HAR on UCIHAR and PAMAP2 datasets under non-cross-subject evaluation. Hou *et al.* [62] compared the performance of HAR among traditional machine learning methods (SVM, KNN, and random forest) and deep learning methods (CNN and LSTM), and they found that when the size of HAR datasets is small, traditional structures are more likely to obtain satisfactory results, while deep learning methods are better choices when the dataset has a large scale. Leonardis *et al.* [63] comprehensively evaluated the effectiveness of five traditional machine learning classifiers (SVM, DT, KNN, NB, and MLP) on self-labeled activity recognition datasets, and focused on discussing the real-time performance of different classifiers on wearable devices. Angerbauer *et al.* [64] examined the traditional machine learning model and two commonly used deep learning models (CNN and LSTM) on HAR in terms of accuracy, memory consumption, real-time performance, etc. They found that random forest is the best model for memory-limited applications, while the best model considering complexity and performance is linear kernel SVM. The two deep neural networks are comparable in performance, but their increasing complexity makes it hard for real use cases. Gholamiangonabadi *et al.* [41] compared the cross-subject HAR performance between the feed forward neural network and CNN, and the results showed that CNN architecture with two convolutions and one-dimensional filter had the best generalization ability.

D. SUMMARY

With the growth of deep learning research, recent HAR studies focus on the improvement of recognition accuracy using complex deep architectures or transfer learning [53], [54], [55], [56] rather than traditional solutions [4], [5], [14]. However, some studies [64], [65] discovered the phenomenon that the traditional solutions outperform deep methods under the same metric on HAR, and the reason remained unclear. In this paper, we conduct a comprehensive comparison between traditional machine learning and deep learning methods on HAR under strict LOSO validation, and make a further analysis to the result of the experiment. Different from studies like [45], [46], the hyper-parameter settings of traditional machine learning is clarified in detail in this paper, together with the explicit definition of strict LOSO cross-validation.

III. MATERIALS AND METHODS

A. DATASETS

To comprehensively evaluate the cross-subject activity recognition performance of traditional machine learning and deep learning, we selected 3 datasets with different scales, containing multiple subjects and covering simple, complex, and similar activities.

The mHealth dataset contains body motion and vital signs recordings from 10 subjects. Each subject performed 12 activities in an out-of-lab environment without any constraints. 3 IMU sensors were placed on the subject's chest, right wrist, and left ankle to measure the 3-axis acceleration (m/s^2), 3-axis angular velocity (deg/s), and 3-axis magnetic field (G/s), respectively. Besides, the sensor placed on the chest also provides 2-lead ECG measurements. The sampling frequency of all sensors is 50 Hz.

The PAMAP2 dataset is a benchmark for daily activity recognition. It was recorded by 9 subjects (8 males and 1 female, aging from 24 to 32), wearing three IMUs placed on the arm, chest, and ankle, respectively, consisting of 12 activities including simple activities (such as sitting, running, etc.)

and complex activities (such as cleaning, ironing, etc.). The sensor data were recorded at 100 Hz.

The UCIDSADS dataset was specially constructed for daily and sports activities recognition. It comprises 19 activities, covering multiple groups of similar activities such as walking on a treadmill with different inclination angles, cycling in a vertical or horizontal position, etc. Each activity was performed by 8 subjects for 5 minutes in their style without any constraints. 5 IMU sensors on the torso and the four limbs were calibrated to acquire data at the sampling frequency of 25 Hz.

Only IMU data from the 3 datasets are used in the experiment. The raw sensor data are cleaned according to the procedures specified in papers that described the datasets [25], [26], [27], [28]. Linear interpolation is used to cope with missing data, and 10 seconds from the beginning and the end of each labeled activity is deleted to void dealing with eventual transient activities, as mentioned in [26]. In detail, a median filter and a fifth-order Butterworth low-pass filter with the cut-off frequency at 11 Hz are applied to reduce the noise. Before feature extraction, the sensor data are segmented by a sliding window with an appropriate length. A smaller window size may not accurately capture all the features of the activity, while a larger window size may introduce interference from other actions. In this paper, a fixed length of one-second sliding window with 50% overlap is used to perform segmentation on the 3 datasets. The label distribution of the PAMAP2 dataset is uneven, especially for subject 9 who lacks most of the samples after data cleaning, and thus only the data from subjects 1 to 8 are used in the experiment. Fig. 2 and Table 1 show the statistic details of the 3 datasets, including the composition and proportion of each activity.

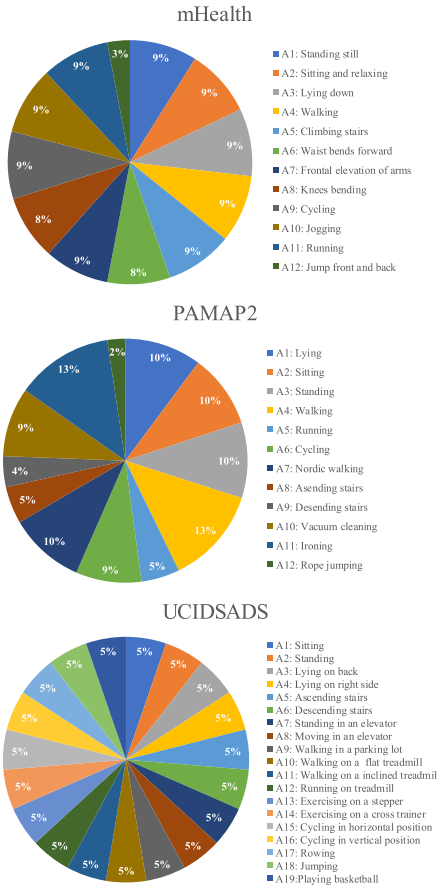


FIGURE 2. Activity distribution of the 3 datasets: mHealth, PAMAP2, and UCIDSADS.

B. EVALUATION CRITERIA

Strict cross-subject LOSO test: To simulate new subject scenarios and evaluate the cross-subject recognition performance of the model, we adopt a strict cross-subject LOSO cross-validation as followed: First, all samples of subject i are taken from N subjects from the dataset as the testing set, and the remaining $N-1$ subjects are used as the training set, in which the optimal hyper-parameters are grid-searched using LOSO cross-validation as well. After determining the optimal hyper-parameters, the model is re-trained on the entire training set, and the classification performance is tested on the testing set consisting of subject i. The process above is iterated N times until each subject has been taken as the testing set once, and the cross-subject recognition performance is obtained by averaging the results from N iterations.

Non-cross-subject 5-fold test: In HAR-related research, the non-cross-subject test is usually used to verify the performance of the model regardless of subject labels, in which the training set and testing set may contain different samples from the same subject, thus the classification models can achieve fairly high recognition accuracy on the testing set. To simplify the training process and maintain a unified comparison benchmark, this paper directly uses the N groups of hyper-parameters obtained through the cross-subject LOSO cross-validation mentioned above as the model configuration (i.e., the validation process is skipped), and performs a non-cross-subject 5-Fold cross-testing (80% samples for the training set and the rest 20% samples for the testing set) on the dataset. Finally, the average classification performance of N groups of hyper-parameters is taken as the non-cross-subject recognition result of the model.

C. METHODS

Model design and hyper-parameter selection need to avoid overfitting to overcome the impact of new subject scenarios. For traditional machine learning models, this paper incorporates the parameters related to overfitting into the hyper-parameter search space, such as the maximum tree depth in the random forest, the regularization parameter of SVM, etc. For deep learning models, effective generalization methods such as dropout, batch normalization, and L2 regularization are fully utilized in the network structure design. For HAR, a lightweight deep learning model is sufficient to achieve a satisfactory recognition performance [66], while too many trainable parameters often have the risk of

TABLE 1. Statistics of the 3 datasets: mHealth, PAMAP2, and UCIDSADS.

Dataset	Number of Subjects	Number of Activities	Frequency	Window size	Number of IMUs	Number of Samples
mHealth	10	12	50	48	3	14285
PAMAP2	8	12	100	96	3	36041
UCIDSADS	8	19	25	24	5	94992

TABLE 2. Evaluated hyper-parameters for traditional machine learning models in this paper.

Classifier	Hyperparameter	Candidate	
SVM	Kernel	RBF	Linear
	γ	$\frac{1}{n}$, $\frac{1}{n \cdot var}$	\
	C	0.01, 0.1, 1, 1.5, 10, 100	
Random Forest	Splitting criterion	Gini	Entropy
	Maximum splitting number	$\log_2 n$, $\sqrt{n}$	
	Number of trees	30, 50, 75, 100	
	Maximum tree depth	8, 12, 24, 32	
KNN	Number of neighbors	5, 10, 20, 30, 50	
	Weights function	Uniform, Distance	

overfitting, so the network model is preferably designed with fewer network layers.

1) TRADITIONAL MACHINE LEARNING

In this paper, the three most widely used traditional machine learning models, SVM, random forest, and KNN, are selected for recognition performance evaluation. According to the criteria defined in section III-B, each of those models has been tuned to extract the best possible cross-subject performance for the given dataset using grid search over the defined hyper-parameter space shown in Table 2. For example, the number of neightbors in KNN has five choices, while the weights function has two, then we have 5×2 hyper-parameter sets for grid search. Note that "\" means not applicable for the candidate.

All parameter n in Table 2 are the number of the input features. The RBF in SVM denotes the radial basis function kernel; γ is the kernel coefficient where *var* is the variance of the features; C is the regularization parameter in SVM. In KNN, the weights function "uniform" means all points in each neighborhood are weighted equally, while "distance" means points are weighted by the inverse of their distance. In the random forest, the splitting criterion is the function to measure the quality of the feature split in tree nodes. The Gini impurity is calculated as:

$$Gini = 1 - \sum_i p_i^2 \tag{1}$$

while the entropy (information gain) is obtained by:

$$Entropy = -\sum_i p_i \log_2 p_i \tag{2}$$

where p_i is the probability of class i from all data in current node.

Traditional machine learning relies on good feature engineering to express the original data. After pre-processing and sliding window procedure, we performed dimension augmentation on the input data. First, the amplitude value M was calculated as:

$$M = \sqrt{s_x^2 + s_y^2 + s_z^2} \tag{3}$$

to reduce the influence of orientation variation, where s_x, s_y, and s_z are data from the 3-axis of each sensor in each time window, respectively. Then, the original data and amplitude data were converted to the frequency domain by applying the short-time Fourier transform. Table 3 lists the hand-crafted features in the time domain and frequency domain used in this paper, where mean, harmonic mean, median, etc., measure the central tendency of the data, while standard deviation, absolute median deviation, and interquartile range describe the distribution of data for each time window. Note that $\boldsymbol{x} = \{x_1, x_2, \ldots, x_i, \ldots, x_n\}$ are sample points from one axis of the sensor within a single time window, and n is the window length. The Pearson correlation coefficient represents the correlation between data from different axes of the sensor. If the harmonic mean and Pearson correlation coefficient encounter a zero division, they are both set to 0 directly. For each IMU device, the accelerometer, gyroscope, and magnetometer all have 3 axes, thus all the feature components mentioned in Table 3 have 9 dimensions.

TABLE 3. Hand-crafted features in the time and frequency domain used in this paper.

Feature	Description
Mean	$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$
Harmonic mean	$\frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$
Standard Deviation	$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2}$
Max	$\max(x_1, x_2, ...x_n)$
Min	$\min(x_1, x_2, ...x_n)$
Peak-to-Peak Amplitude	$\max(\boldsymbol{x}) - min(\boldsymbol{x})$
Median	$\omega = median(x_1, x_2, ...x_n)$
Median absolute deviation	$median(\lvert\boldsymbol{x} - \omega\rvert)$
Interquartile range	$quartile(\boldsymbol{x}, 75) - quartile(\boldsymbol{x}, 25)$
Sum of area	$\sum_{i=1}^{n} \lvert x_i \rvert$
Signal mean energy	$\frac{1}{n}\sum_{i=1}^{n}(x_i)^2$
Skew	$\frac{1}{n\sigma^3}\sum_{i=1}^{n}(x_i - \mu)^3$
Kurtosis	$\frac{1}{n\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^4$
Pearson's Correlation Coefficient	$\frac{cov(\boldsymbol{x},\boldsymbol{y})}{\sigma_{\boldsymbol{x}}\sigma_{\boldsymbol{y}}}$

We extracted corresponding features mentioned in Table 3 from the data frame after dimension expansion (including original time-domain data, amplitude time-domain data, original frequency-domain data, and amplitude frequency-domain data), which were further normalized into a normal distribution with mean 0 and variance 1 according to (4), where f_μ and f_σ are the mean and standard deviation of the input feature $\boldsymbol{f}$. Before the normalization, we delete the features that are not distinct enough with $f_\sigma < 0.01$. The extracted and actually used numbers of features on the three datasets are listed in Table 4. Finally, the concatenated features are used

TABLE 4. Extracted and used feature number on the three datasets.

	mHealth	PAMAP2	UCIDSADS
Extracted feature number	805	1035	1725
Used number ($f_\sigma \geq 0.01$)	741	986	1549

as the input of traditional machine learning classifiers listed in Table 2.

$$\frac{f - f_\mu}{f_\sigma} \tag{4}$$

2) DEEP LEARNING

In this paper, 5 commonly used deep neural network architectures in the field of HAR are chosen for experiments, namely Conv1d-CNN, Conv2d-CNN, LSTM, BLSTM, and CNN-LSTM. The overall architecture of the Conv1d-CNN and Conv2d-CNN is shown in Fig. 3, consisting of 3 convolutional layers, 3 max-pooling layers, and a fully connected layer. The batch normalization is used between each convolutional layer to speed up convergence and improve generalization, while the dropout is used to prevent overfitting before the fully connected layer. Conv1d-CNN and Conv2d-CNN have the same network structure, but use different convolution kernels: Conv1d-CNN regards the original data as a multi-channel continuous time series and uses a one-dimensional convolution kernel; Conv2d-CNN regards the original data as single-channel image data, using a 2D convolution kernel.

The LSTM and BLSTM network architectures used in this paper are shown in Fig. 4. The model consists of 3 stacked LSTM/BLSTM layers and a fully connected layer, with dropouts added between each layer to avoid overfitting.

The CNN-LSTM structure used in this paper is shown in Fig. 5. The feature extraction network consists of 4 Conv1d-CNN layers with batch normalization between each layer and 2 stacked LSTM layers to extract temporal-spatial features of human activities.

All the stacked CNN used the same number of kernel and kernel filters, while the LSTM layers shared the same number of hidden channels, and the number of neurons in the final fully connected layer is determined according to the feature dimension output from the feature extraction network. We normalized the filtered data by (4) before feeding it into the deep neural network models. To fine-tune the deep learning models depicted above, we evaluate the hyper-parameter ranges in Table 5, where C denotes the number of axis, which is 9 times the number of IMUs in Table 1. The parameter L is the window size defined in Table 1. Note that "\" means not applicable for the candidate. Both training and testing are performed according to the criteria defined in Section III-B.

IV. RESULTS AND DISCUSSION

A. PERFORMANCE COMPARISON

In the experimental result section, all the testing results are evaluated by F1-score and accuracy, defined as:

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

TABLE 5. Evaluated hyper-parameters for deep learning models in this paper.

	Conv1d / 2d-CNN	LSTM / BLSTM	CNN-LSTM
Optimizer algorithm	Adam		
Loss function	CrossEntropyLoss		
Batch size	100		
Learning rate	0.001		
Epoch	15		
Dropout	0.1, 0.3, 0.5, 0.7		
Weight decay	$10^{-2}, 10^{-3}, 10^{-4}$	$10^{-4}, 10^{-5}, 10^{-6}$	$10^{-2}, 10^{-3}, 10^{-4}$
Input size	1d: $100 * C * L$ 2d: $100 * 1 * L * C$	$100 * L * C$	$100 * C * L$
Kernel	1d: 3, 5 (stride 1) 2d: $3 * 3, 5 * 5$ (stride 1)	\	3, 5 (stride 1)
Kernel filters	12, 24, 36, 48, 64	\	24, 36, 48, 64
Max pool	1d: 2 (stride 2) 2d: $2 * 2$ (stride 2)	\	\
Hidden size	\	16, 32, 64, 128	24, 36, 48, 64, 128

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

$$F1\text{-}score = \frac{2}{Precision^{-1} + Recall^{-1}} \tag{8}$$

where TP, TN, FP, and FN denote the number of true positive, true negative, false positive, and false negative classification, respectively. For a multi-class problem as HAR, the precision and recall are calculated for each class independently, and the total value are weighted average according to the number of true instances for each class i as:

$$Precision_{weighted} = \frac{1}{N}\sum_{i}^{N} Precision_i * N_i \tag{9}$$

$$Recall_{weighted} = \frac{1}{N}\sum_{i}^{N} Recall_i * N_i \tag{10}$$

where $N = \sum_i N_i$ are the total number of all labels.

Fig. 6 and 7 show the box plot of F1-score and accuracy of the traditional machine learning and deep learning models on 3 datasets, where the box extends from the first quartile to the third quartile of the data, with a line at the median. Note that the blue boxes are non-cross-subject results, while the orange boxes are for cross-subject tests. Table 6 and 7 demonstrate the average of accuracy and F1-score of the traditional machine learning and deep learning models on 3 datasets, together with the 95% confidence limits. Since the number of cross-validation is small in LOSO, we use t-distribution for an unbiased 95% confidence interval as:

95% confidence interval

$$= \left[\mu - t(n-1)\frac{S}{\sqrt{n}}, \mu + t(n-1)\frac{S}{\sqrt{n}}\right] \tag{11}$$

$$S^2 = \frac{1}{n-1}\sum_i (x_i - \mu)^2 \tag{12}$$

where n denotes the number of users in different datasets, and μ is the average of samples $x_1, \ldots, x_i$. The following insights can be obtained: (1) Under the non-cross-subject test, all models except LSTM achieved nearly perfect performance, and traditional machine learning models got the

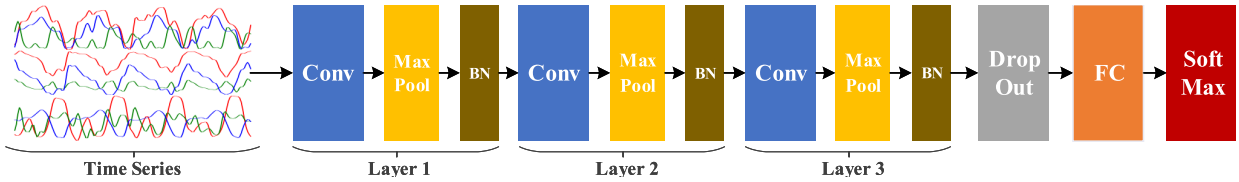


FIGURE 3. The architecture of CNN model in this paper.

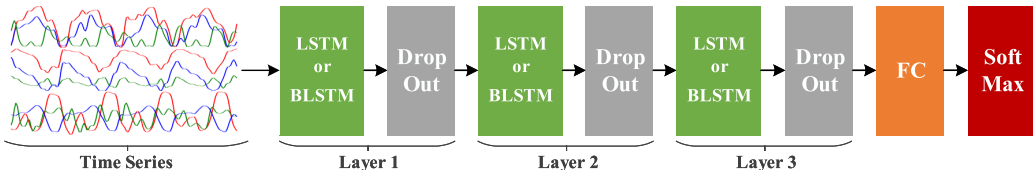


FIGURE 4. The architecture of LSTM and BLSTM model in this paper.

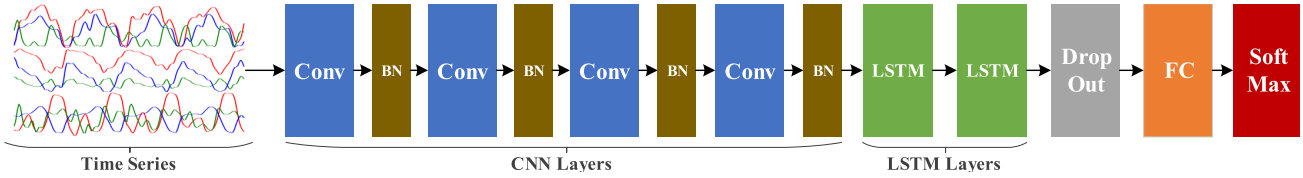


FIGURE 5. The architecture of CNN-LSTM model in this paper.

highest F1-score and accuracy on all datasets. On the smaller mHealth dataset, the traditional machine learning models generally outperform the deep learning models on F1-score and accuracy, while on the larger dataset like UCIDSADS, the deep learning models become comparable. (2) Compared with the non-cross-subject tests, all models have different degrees of performance loss in cross-subject conditions, and the traditional machine learning models, especially SVM and random forest, show better generalization ability on the three datasets. In detail, the average F1-scores loss of the traditional machine learning models are 5.45%, 8.20%, and 7.88% on mHealth, PAMAP2, and UCIDSADS, respectively, while the numbers for the deep learning models are 15.23%, 13.77%, and 15.52%. (3) The deviation of F1-score and accuracy are quite small in non-cross-subject test, while the numbers increase significantly in LOSO test, indicating the unstable performance over different subjects. On the smaller mHealth dataset, the traditional machine learning models share hardly any overlap in deviation with deep learning. However, in PAMAP2 and UCIDSADS the overlap becomes noticeable, which means the deep learning models have better performance on some subjects. Random forest has the smallest deviation among all the datasets, which is the robustest solution for cross-subject HAR.

Fig. 8 shows the average confusion matrix of deep learning (DL) models (except for LSTM) and traditional machine learning (TML) models under the cross-subject LOSO testing condition. It is worth noting that two simple static activities, standing and sitting, are easily confused with other activities in deep learning models over three datasets, which is however much improved in traditional machine learning models. In addition, for most periodic activities, such as walking (A4), running (A5), cycling (A6), nordic walking (A7), rope jumping (A12) in PAMAP2; exercising on a stepper (A13), exercising on a cross-trainer (A14), jumping (A18) in UCIDSADS, etc., traditional machine learning models have better classification performance. Nevertheless, for the confusion between similar activities, such as jogging (A10) and running (A11) in mHealth; standing and moving in an elevator (A7, A8), walking on different planes (A9, A10, A11) in UCIDSADS, traditional machine learning methods do not take more advantages.

In addition, by analyzing the confusion matrix of each subject, we found that in cross-subject activity recognition, deep learning models are more likely to misclassify some activities almost entirely, resulting in a significant drop in overall recognition accuracy. For instance, the static activity, A1: standing still of subject 1, are all wrongly classified as A8: knees bending in the mHealth dataset using deep learning models, as shown in Fig. 9(a), (c), and (e). While in traditional machine learning cases, the classification remains accurate, which can be seen in Fig. 9(b), (d), and (f).

Nevertheless, some traditional machine learning models can also make a totally wrong recognition. For example, the A7: standing in an elevator in UCIDSADS is incorrectly classified as A8: moving in an elevator for subject 4 using SVM, as shown in Fig. 10(d), which is similar to the behavior of BLSTM and CNN-LSTM in Fig. 10(c) and (e), while the

TABLE 6. The average accuracy and 95% confidence limits of different learning models on the three datasets using non-cross- and cross-subject evaluation criterion. The best accuracy among all the methods is highlighted.

	mHealth		PAMAP2		UCIDSADS	
	Non-cross	Cross	Non-cross	Cross	Non-cross	Cross
Conv1d	98.74±0.31	83.42±6.78	97.41±0.34	84.15±15.35	97.88±0.83	85.4±4.84
Conv2d	99.2±0.54	83.6±6.21	98.16±0.34	90.07±6.23	98.75±0.13	88.8±3.06
LSTM	91.14±1.38	83.26±4.12	86.95±2.03	73.62±16.26	97.49±0.41	82.97±4.85
BLSTM	95.16±0.97	84.04±5.67	93.98±0.75	78.87±17.81	97.98±0.58	82.14±5.44
CNN-LSTM	98.31±0.13	83.31±5.88	95.34±1.11	80.08±12.86	97.5±0.79	84.48±2.5
KNN	99.16±0.01	**94.68±2.85**	97.2±0.34	88.83±6.87	98.69±0.12	90.18±3.07
SVM	99.43±0.14	94.41±3.82	**98.86±0.0**	91.72±7.85	**99.34±0.16**	**93.05±3.93**
RF	**99.54±0.03**	94.33±3.1	98.57±0.09	**92.17±4.53**	98.7±0.38	92.2±2.49

TABLE 7. The average F1-Score and 95% confidence limits of different learning models on the three datasets using non-cross- and cross-subject evaluation criterion. The best F1-score among all the methods is highlighted.

	mHealth		PAMAP2		UCIDSADS	
	Non-cross	Cross	Non-cross	Cross	Non-cross	Cross
Conv1d	98.7±0.36	80.99±7.82	97.4±0.36	83.16±16.68	97.7±1.23	83.0±5.66
Conv2d	99.09±0.79	80.76±7.34	98.13±0.39	89.73±6.82	98.74±0.14	87.31±3.47
LSTM	89.62±1.88	79.74±4.63	84.03±2.85	70.15±17.85	97.29±0.55	79.95±6.34
BLSTM	94.66±1.17	81.44±6.63	93.86±0.81	77.9±19.5	97.86±0.79	78.97±6.9
CNN-LSTM	98.3±0.13	81.28±6.7	95.32±1.14	78.98±14.2	97.51±0.79	82.24±2.97
KNN	99.16±0.01	**94.28±3.27**	97.18±0.35	87.83±8.61	98.67±0.13	89.33±3.52
SVM	99.43±0.14	93.9±4.44	**98.86±0.0**	90.59±9.65	**99.34±0.16**	**92.27±4.72**
RF	**99.54±0.03**	93.59±3.71	98.57±0.09	**91.6±5.42**	98.69±0.39	91.45±3.38

classification is relatively correct using Conv2d-CNN, KNN and random forest as shown in Fig. 10(a), (b), and (f).

B. ANALYSIS AND VALIDATION

1) GENERALIZATION ABILITY OVER HAND-CRAFTED FEATURES

The heterogeneity among subjects is the main reason for the decline of cross-subject recognition performance. As shown in Fig. 11 (a), the t-SNE (t-distributed stochastic neighbor embedding) algorithm based on euclidean distance metric is used to map the raw data of the mHealth dataset to a two-dimensional space for visualization, where the perplexity and the maximum number of optimizing iterations are set to 30 and 1000, respectively. From Fig. 11 (a) most of the subject data in one certain activity are clustered individually, which means that the subject data has its unique input distribution. It is worth noting that the two-dimensional data points corresponding to the A1 (standing still) and A2 (sitting and relaxing) activities of different subjects are quite scattered and mixed with other activities, which is consistent with the overall confusion results of deep learning models in Fig. 8. On the other hand, the hand-crafted features used by traditional machine learning models are designed based on domain knowledge and do not depend on specific subjects, thus these time-frequency domain statistical features can reduce the data distribution differences between different users, as shown in Fig. 11 (b). This is one of the major reasons why traditional machine learning models generalize better on cross-subject scenarios, since the end-to-end trained deep learning models automatically extract features based on training data, from which subject-related features are easy to learn and the models are more susceptible to training distribution. Moreover, HAR datasets often have a small scale compared with tasks like computer vision and natural language processing, making the deep learning models with a large amount of parameters hard to extract general features.

TABLE 8. The average accuracy and F1-score of MLP model with hand-crafted features.

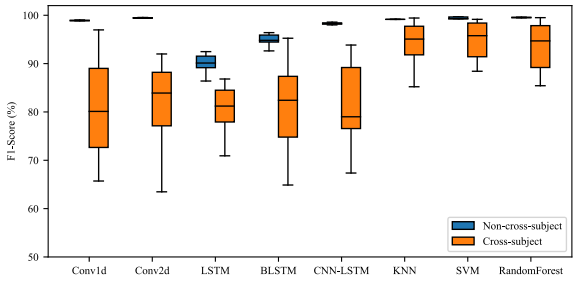
Dataset	Accuracy (%)		F1-score (%)	
	Non-cross-subject	Cross-Subject	Non-cross-subject	Cross-Subject
mHealth	99.24	93.40	99.21	92.33
PAMAP2	97.79	90.61	97.71	89.30
UCIDSADS	98.43	92.62	98.42	91.76

To compare the effect of traditional hand-crafted features and features automatically extracted by the neural network on cross-subject recognition, the feature extraction part is removed in deep learning models and only retained the fully connected layers to form an MLP classifier. According to the criteria defined in section III-B, the activity recognition performance of MLP using hand-crafted features as input under both non-cross-subject and cross-subject conditions is evaluated. The experimental results are shown in Table 8, where MLP using hand-crafted features has achieved better cross-subject recognition results than the five deep learning models on all three datasets, and the average F1-score and accuracy is comparable with SVM and random forest, proving the superiority of traditional hand-crafted features in cross-subject recognition.

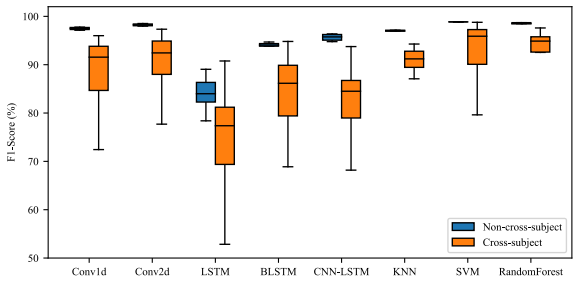
2) TRAINING STRATEGY FOR DECISION-TREE-BASED LEARNING METHODS

One of the key solutions for the cross-subject recognition problem is to maximize the discrimination among different classes and ignore the various distribution of subjects, which is performed by transfer learning in deep neural networks as mentioned in Section II. In this paper, we propose a novel training strategy for the decision-tree-based learning methods under this principle to cope with cross-subject scenarios.

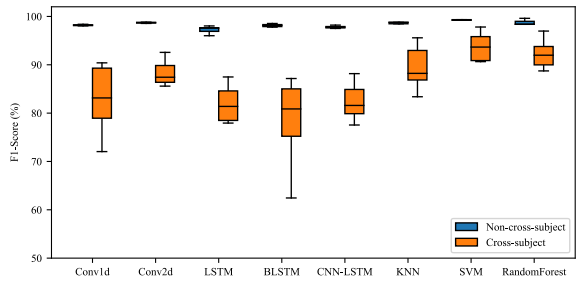
Recall that the Gini impurity, when making decision trees in the random forest, indicates the label diversity of data in the



(a) mHealth

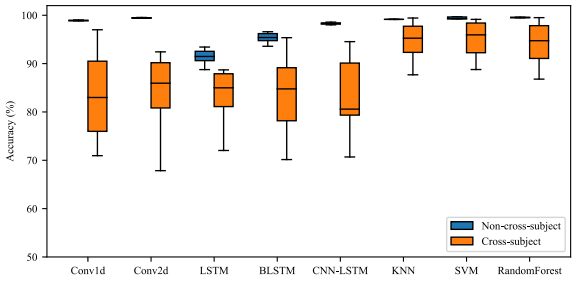


(b) PAMAP2

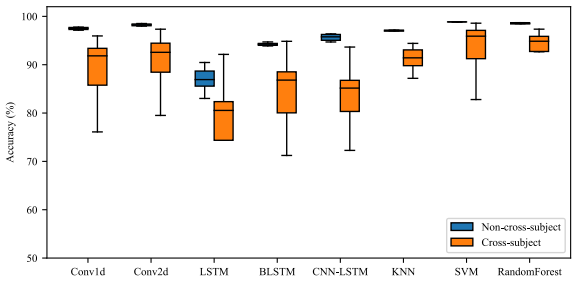


(c) UCIDSADS

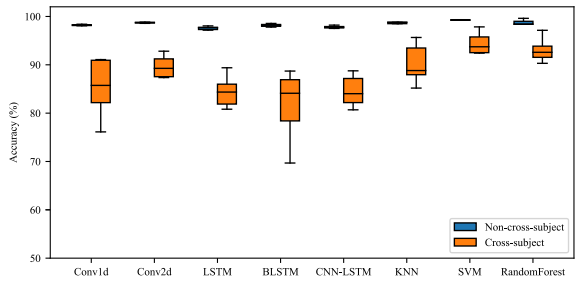
FIGURE 6. The box plot of F1-score of different learning models on the three datasets using non-cross- and cross-subject evaluation criterion.



(a) mHealth



(b) PAMAP2



(c) UCIDSADS

FIGURE 7. The box plot of accuracy of different learning models on the three datasets using non-cross- and cross-subject evaluation criterion.

current node, as shown in (1) where p_i actually means the data proportion of class i. In the traditional training phase, each decision tree is established by greedily selecting the features and corresponding thresholds to minimize the weighted sum of Gini impurity of every left and right child nodes recursively, formulated as:

$$\min\left(\frac{n_l}{n}\left(1-\sum_i p_{i,l}^2\right)+\frac{n_r}{n}\left(1-\sum_i p_{i,r}^2\right)\right) \quad (13)$$

where $n = n_l + n_r$ is the total number of samples from left and right child nodes. The process of node division is actually feature selection, thus ideally features unrelated to subjects characteristic but strongly related to distinguishing activities should be selected to achieve better cross-subject generalization.

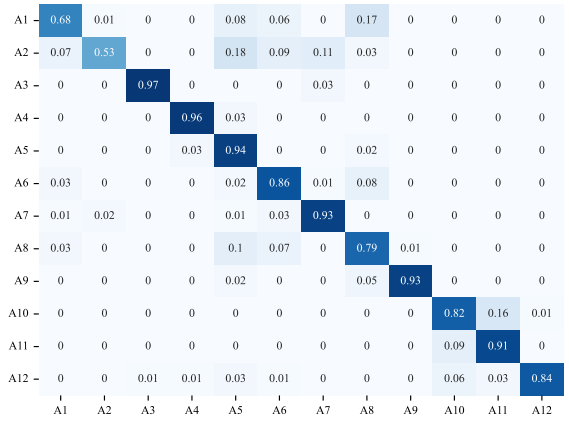
On the other hand, it is necessary to reduce the gini impurity on labels of different activity classes, while keeping the gini impurity on labels of different subjects as much as possible. In this case, the samples are split into nodes regardless of distinguishing the subjects, which can be formulated as a new object as:

$$\max\left(\frac{n_l}{n}\left(1-\sum_j p_{j,l}^2\right)+\frac{n_r}{n}\left(1-\sum_j p_{j,r}^2\right)\right) \quad (14)$$

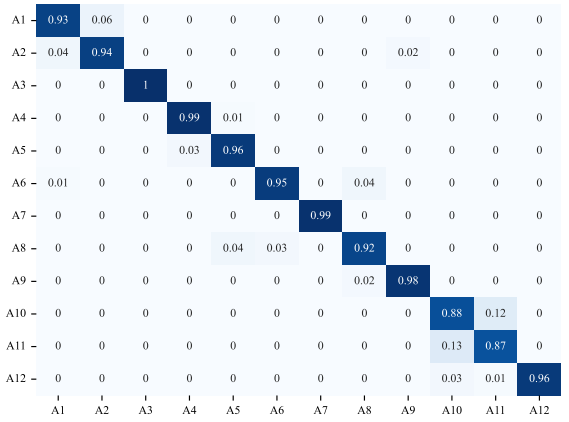
where $p_{j,l}$ and $p_{j,r}$ are the sample proportion of subject j in left and right child nodes, respectively. A parameter $\alpha \in [0, 1)$ is set to represent the importance of gini impurity for subject labels, then the original criterion of finding the best split can be rewritten as maximizing the following formula:

$$\alpha\left(\frac{n_l}{n}\left(1-\sum_j p_{j,l}^2\right)+\frac{n_r}{n}\left(1-\sum_j p_{j,r}^2\right)\right) -(1-\alpha)\left(\frac{n_l}{n}\left(1-\sum_i p_{i,l}^2\right)+\frac{n_r}{n}\left(1-\sum_i p_{i,r}^2\right)\right) \quad (15)$$

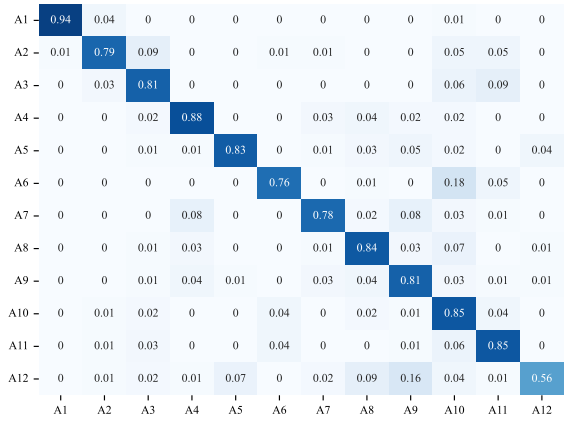
, which degenerates to the original object function (13) when $\alpha = 0$. Similarly, the entropy criterion in (2) can be modified to a subject-independent form as well.



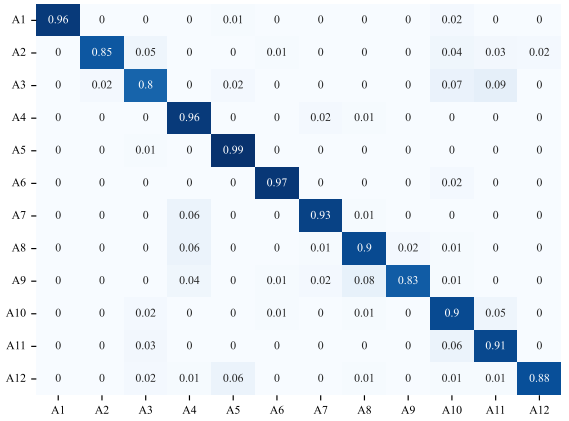
(a) Confusion matrix of DL on mHealth



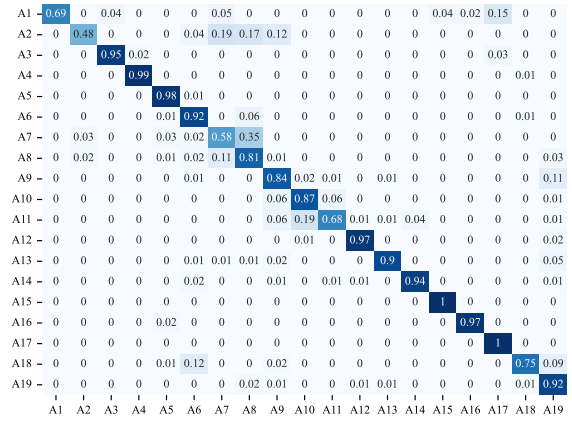
(b) Confusion matrix of TML on mHealth



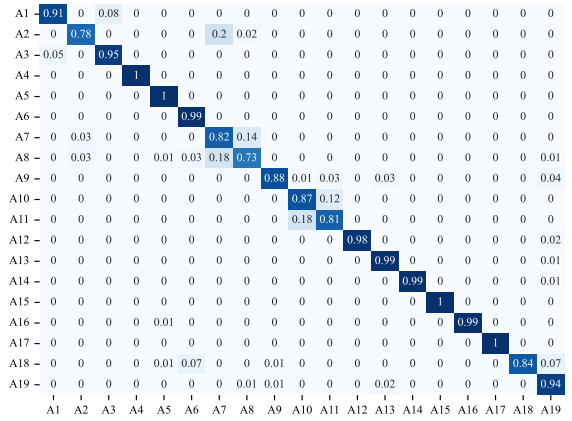
(c) Confusion matrix of DL on PAMAP2



(d) Confusion matrix of TML on PAMAP2



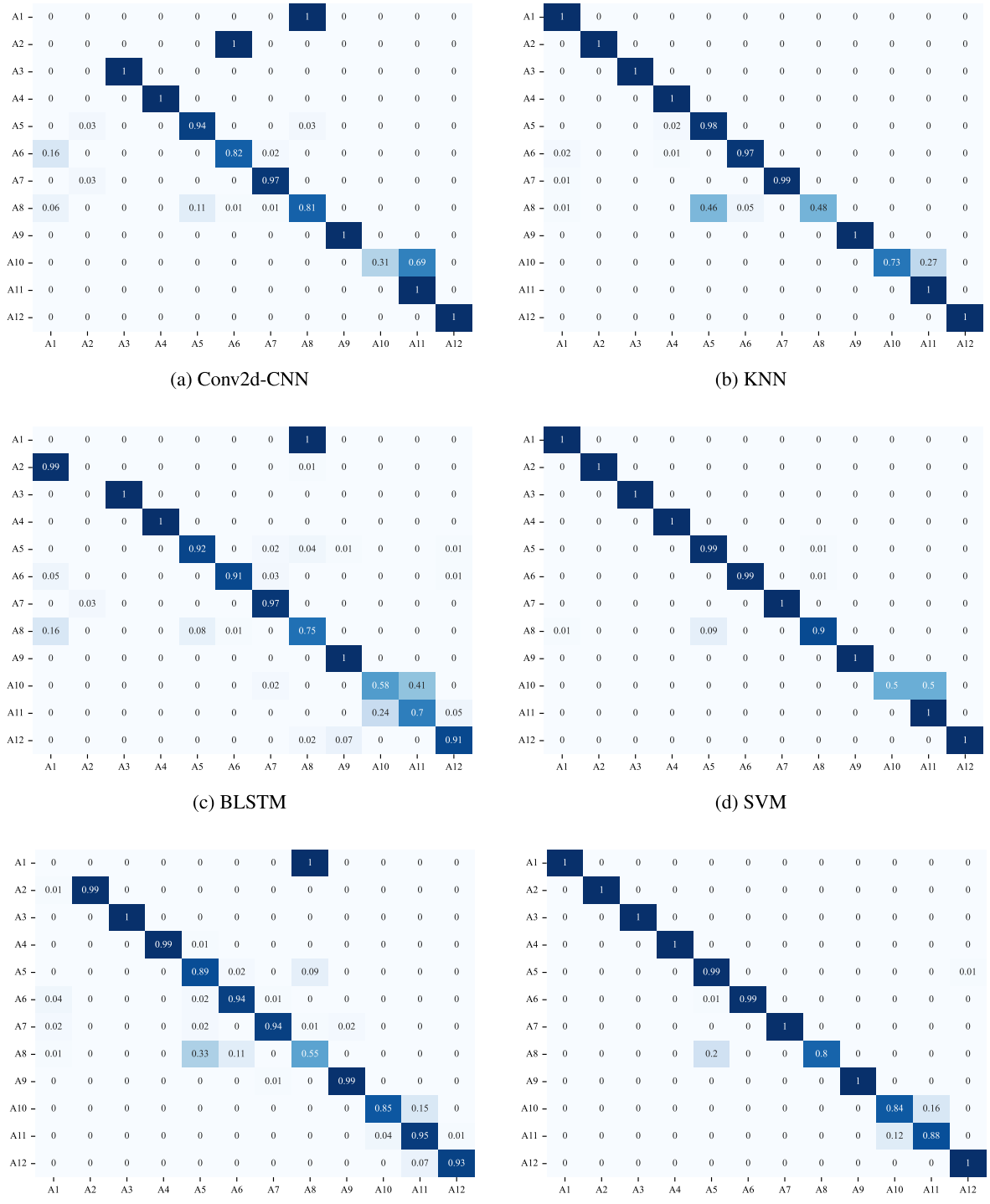
(e) Confusion matrix of DL on UCIDSADS



(f) Confusion matrix of TML on UCIDSADS

FIGURE 8. Average confusion matrix of deep learning and traditional machine learning models on the three dataset under strict LOSO cross-validation.

By varying α from 0.1 to 0.9 with each step of 0.1 into Table 2 as new hyper-parameter grid searching, the random forest model is re-trained and tested under strict cross-subject LOSO described in Section III-B. Table 9 shows the average accuracy and F1-score of random forest with modified training strategy, where the cross-subject performance is better than the all the method as shown in Fig. 6 and 7. The paired t-tests is conducted between the original random forest and the modified one on F1-score to determine the degree of significant difference in terms of the significance level



(a) Conv2d-CNN (b) KNN

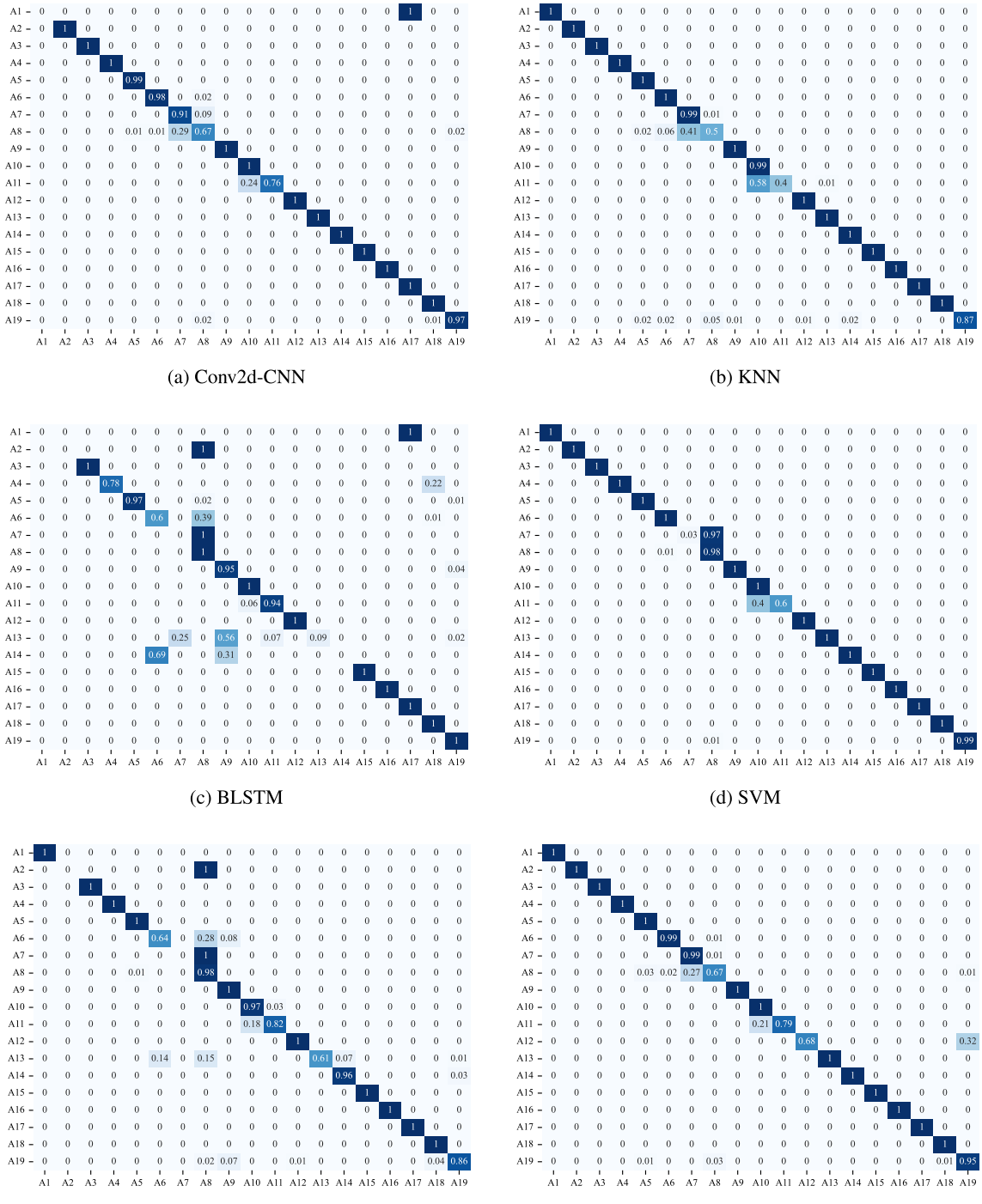
(c) BLSTM (d) SVM

(e) CNN-LSTM (f) Random forest

FIGURE 9. Confusion matrix of subject 1 on mHealth using different learning models.

$p-$value (two tailed). Using a threshold of $p = 0.05$ for the null hypothesis, the result reflects the effectiveness of the proposed training strategy. Note that in PAMAP2 the null hypothesis is nearly failed to reject due to the uneven activities labels distribution as shown in Fig. 2, and the sample numbers for different subjects are also significantly diverse [26], which is hard for the modified random forest to make a balanced tree node splitting between subject labels and activity labels.

We further explore the behavior of the modified object function by varying the number of decision trees in the random forest, using a set of fine-grained α with each step of



(a) Conv2d-CNN

(b) KNN

(c) BLSTM

(d) SVM

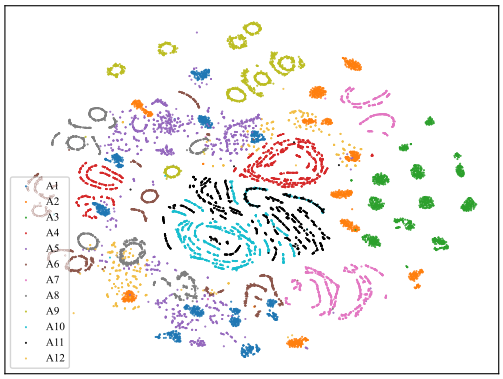
(e) CNN-LSTM

(f) Random forest

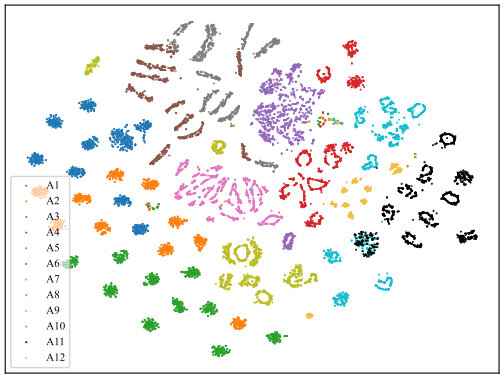
FIGURE 10. Confusion matrix of subject 4 on UCIDSADS using different learning models.

0.0375. Fig. 12 shows the influence of different α values on the average recognition accuracy using different numbers of decision trees. According to the figure, the modified objective function can achieve stable improvement in accuracy with appropriate α, and the trends of improvement introduced by different α are highly consistent in single dataset when the number of trees increases in the random forest. Moreover, the optimal α is correlated with the characteristics of the dataset. For instance, the modified objective function achieves better results on the UCIDSADS with a larger data scale and even distribution of labels, while in PAMAP2 we find a rapid performance degradation, which further explains why the null hypothesis is nearly failed to reject.

Fig. 13 illustrates the comparison of accuracy for each individual subject using the original and modified random forest, with the optimal α value and 50 decision trees. It is

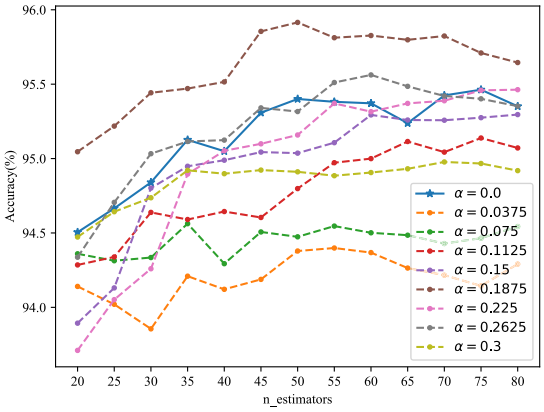


(a) Raw data

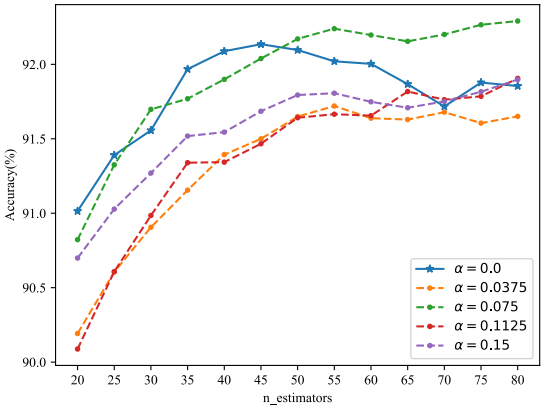


(b) Hand-crafted features

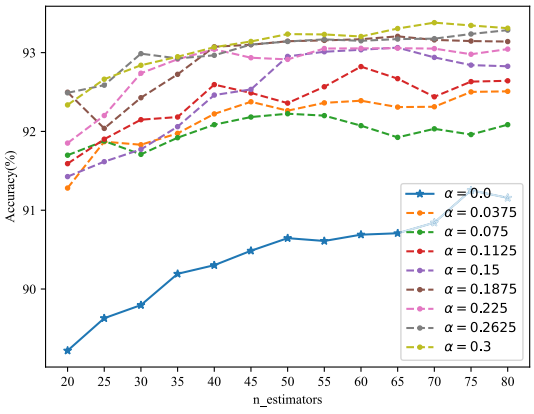
FIGURE 11. The t-SNE projection for raw data and hand-crafted features on mHealth dataset.



(a) mHealth



(b) PAMAP2



(c) UCIDSADS

FIGURE 12. The behavior of the modified object function when changing the number of decision trees in the random forest and the α.

worth noting that in all the datasets, the modified strategy can improve the worst recognition accuracy, as subject 6 in mHealth, subject 1 in PAMAP2, and subject 8 in UCIDSADS. This is because the modified object function tends to assign labels of different individuals to different child leaf nodes evenly, which reduces the accuracy deviation between individuals and further strengthens the cross-subject generalization.

As mentioned in Section II, datasets selection, window length, testing criterion, and other factors directly affect the performance of the HAR model in the experiment, so there is no standard comparison benchmark among different studies. Nevertheless, Table 10 lists the result of state-of-the-art cross-subject HAR studies that considered the same datasets as this paper used, and the random forest with the proposed learning method is also included. Note that "\" means no results are provided on the dataset.

In Table 10, only [60] used the same strict LOSO criterion with training, validation, and testing as mentioned in Section III-B where the hyper-parameters are transparent to the testing set, while in other LOSO-based studies the model are determined by the best result on the testing set. It worth noting that in the domain adaptation research [51], [52], [54], [55], [58], the labeled and unlabeled target samples were used when training the models, while the other studies had only training data for constructing the classifiers.

TABLE 9. The average accuracy and F1-score of random forest with modified training strategy. The $p-$value from paired t-tests on F1-score is also presented.

Dataset	Accuracy (%)		F1-score (%)		p-value
	Non-cross-subject	Cross-Subject	Non-cross-subject	Cross-Subject	
mHealth	99.42	95.09	99.42	94.49	0.024
PAMAP2	98.46	92.21	98.46	91.64	0.038
UCIDSADS	97.85	93.29	97.82	92.70	0.005

TABLE 10. Comparison of the accuracy (a) and F1-score (f) for cross-subject HAR on the three datasets.

Studies	Criterion	mHealth	PAMAP2	UCIDSADS
Unsupervised Domain Adaptation [54]	LOSO	\	\	96.9**a**
Unsupervised Domain Adaptation [55]	\	\	77.68**f**/79.79**a**	75.72**f**/76.67**a**
Unsupervised Domain Adaptation [58]	LOSO	\	13**a**	51**a**
Data Augmentation [49]	LOSO	\	78.6**f**	\
Domain Generalization [60]	Strict LOSO	\	81.6**f**	\
Domain Generalization [17]	LOSO	96.07**f**/96.07**a**	82.13**f**/83.21**a**	91.59**f**/92.14**a**
CNN [41]	LOSO	88.2**f**/85.1**a**	\	\
Supervised Domain Adaptation [52]	LOSO	\	89.6**a**	\
Supervised Domain Adaptation [51]	LOSO	\	\	95.6**a**
Manifold Learning [48]	LOSO	\	\	87.0**a**
Proposed	Strict LOSO	94.49**f**/95.09**a**	91.64**f**/92.21**a**	92.70**f**/93.29**a**

C. DISCUSSION

Except for LSTM, the deep neural network models used in this paper have achieved comparable or better classification accuracy than [22], [67], [68] under non-cross-subject tests. The LSTM model is sensitive to the length of time series. In the experiment, in order to obtain a unified comparison benchmark, we adopt a fixed time window of one-second length for sliding window segmentation on all three datasets with different sampling frequencies, which makes the samples of the three datasets have a different number of sampling points. As shown in Fig. 6 and 7, on the PAMAP2 and mHealth datasets with higher sampling rates (more sampling points), the non-cross-subject classification accuracy of the LSTM model is much lower than that of other models. Therefore, the effect of the number of sampling points on the LSTM model is evaluated as shown in Fig. 14, where the dark-color lines denote the non-cross-subject test while the corresponding light-color lines are the results of the cross-subject test. Overall, using fewer sampling points tends to yield higher cross- and non-cross-subject recognition accuracy, but in most cases the cross-subject test has a 15∼20% drop in F1-score relative to the non-cross-subject test, indicating that the number of sampling points has no decisive influence on the cross-subject recognition performance.

In cross-subject recognition, traditional machine learning methods show higher generalization performance, among which the instance-based KNN method does not require training. However, in practical applications the samples of the training set need to be saved as the classification basis, thus occupying a lot of memory. Compared with SVM, the random forest not only achieves better recognition accuracy under modified training strategy, but also has the advantages of a smaller memory footprint, shorter prediction time, and faster training speed [64]. In general, the random forest model is the best choice for resource-constrained embedding wearable devices.

In this paper, traditional machine learning methods based on hand-crafted features are more suitable for the new subject scenarios in terms of computational complexity and generalization, which however does not mean that deep learning methods are useless in cross-subject recognition. The characteristics of end-to-end training and automatic feature extraction make deep learning models flexible and easy to expand. For example, fine-tuning the trained deep learning models with a small number of labeled samples of the new target subject can quickly reduce the differences in data distribution and obtain a personalized classification model. Table 11 shows the improved cross-subject recognition performance of the Conv2d-CNN model after fine-tuning, where n-shot means the number of samples from the testing subject. The traditional machine learning method is limited by the training method and can not perform fine-tuning on the pre-trained model, thus we re-train the random forest model under the condition of leaking a small number of target samples. As shown in Table 12, on the mHealth dataset with a small data scale, the random forest model with leaked testing samples is slightly better than Conv2d-CNN, while on the UCIDSADS dataset with a larger data scale, the fine-tuned Conv2d-CNN performs better.

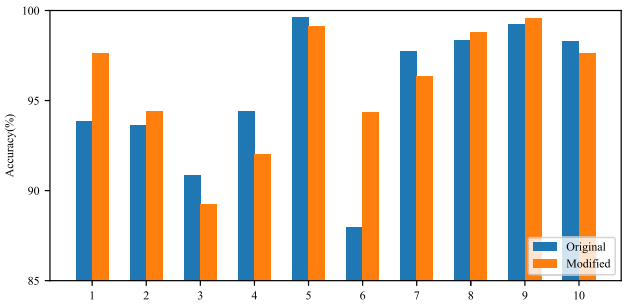
TABLE 11. The improvement of average F1-score (%) for Conv2d-CNN model on three datasets by leaking some testing samples.

	baseline	1-shot	5-shot	10-shot
mHealth	83.60	91.60	96.97	98.15
PAMAP2	90.07	90.20	94.36	95.93
UCIDSADS	88.80	93.52	96.63	96.48

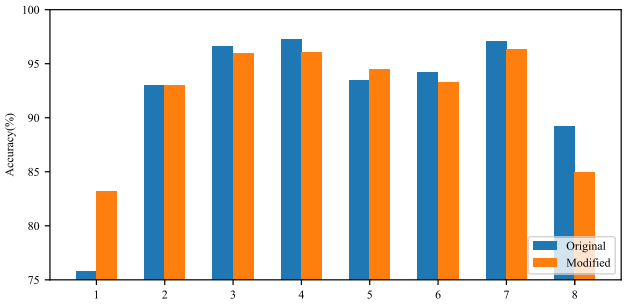
TABLE 12. The improvement of F1-score (%) for random forest model on three datasets by leaking some testing samples.

	baseline	1-shot	5-shot	10-shot
mHealth	94.49	96.63	98.18	98.62
PAMAP2	91.64	93.78	95.05	95.66
UCIDSADS	92.70	94.07	94.23	94.54

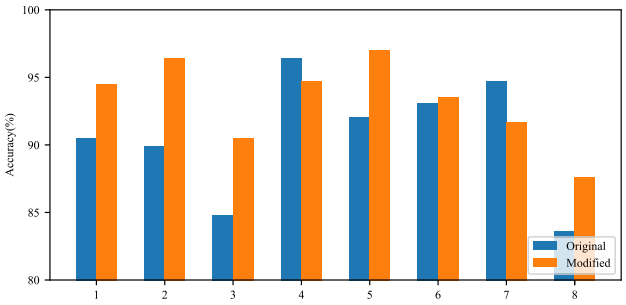
This paper has some limitations. First, the sizes of the datasets we use are small and complete, and they have relatively even distribution on activity labels. We have not covered situations that have huge amount of missing data or significant uneven labels like the last subject in PAMAP2,



(a) mHealth



(b) PAMAP2



(c) UCIDSADS

FIGURE 13. The comparison of accuracy for each individual subject using the original and modified random forest on the three datasets. The x-axis denotes the different subjects.

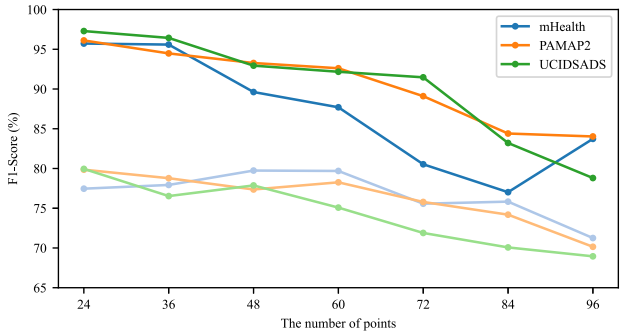


FIGURE 14. The impact of the number of input sampling points for LSTM model on the three datasets.

which might be an advantage for deep learning case. Second, the explanation for the reason that traditional machine learning performs better than deep learning on cross-subject HAR is limited. The decision boundaries for different methods have not been explicitly examined in each LOSO test. Third, the deep learning architectures are inspired by previous studies, and we have not evaluated whether the structure has implicit impact on the result (e.g. the number of convolutional layers in CNN). Finally, although statistical significance, the improvement of modified training process in random forest is small, and we have not proved the methodology on other tasks other than HAR to provide enough evidence for the superiority.

V. CONCLUSION

In this paper, five deep neural network models and three traditional machine learning models are trained and evaluated on three classic HAR datasets: mHealth, PAMAP2, and UCIDSADS. A strict cross-subject LOSO test is deployed to simulate new subject scenarios and evaluate the generalization performance of deep neural networks and traditional machine learning in cross-subject recognition, and the result indicates that all models experience significant performance degradation due to the heterogeneity among subjects, compared to non-cross-subject recognition. In general, the traditional machine learning methods using hand-crafted features achieve better cross-subject recognition than deep learning models on the three datasets, and the analysis proves that the automatic end-to-end feature extraction using deep neural networks is more susceptible to distribution difference between users and prone to learning user-dependent features from training sets. This paper also provides a novel decision-tree-based training strategy, which makes the random forest model achieve best cross-subject HAR performance over all the using learning models, and the competitive results are obtained compared with state-of-the-art cross-subject HAR solutions. In detail, the average F1-score (accuracy) on the three datasets are 94.49% (95.09%), 91.64% (92.21%), and 92.70% (93.29%). Future work will make attempts on other complex datasets and other learning frameworks like AdaBoost, GAN, and VAE to find out the best solution for cross-subject HAR application. The effectiveness of the proposed learning strategy for decision-tree-based methods will be further evaluated on other cross-subject applications like handwriting classification and speech recognition.

ACKNOWLEDGMENT

The authors would like to thank the anonymous referees for their valuable comments and helpful suggestions, and also would like to thank Takashi Mifune from irasutoya.com for providing the public available figure samples presented in Fig. 1.

REFERENCES

[1] H. Ghayvat, S. Mukhopadhyay, B. Shenjie, A. Chouhan, and W. Chen, "Smart home based ambient assisted living: Recognition of anomaly in the activity of daily living for an elderly living alone," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, May 2018, pp. 1–5.

[2] J. Qi, P. Yang, M. Hanneghan, S. Tang, and B. Zhou, "A hybrid hierarchical framework for gym physical activity recognition and measurement using wearable sensors," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 1384–1393, Apr. 2019.

[3] R. S. Antunes, L. A. Seewald, V. F. Rodrigues, C. A. D. Costa, L. Gonzaga Jr., R. R. Righi, A. Maier, B. Eskofier, M. Ollenschläger, F. Naderi, R. Fahrig, S. Bauer, S. Klein, and G. Campanatti, "A survey of sensors in healthcare workflow monitoring," *ACM Comput. Surv.*, vol. 51, no. 2, pp. 42:1–42:37, Apr. 2018.

[4] M. Arif and A. Kattan, "Physical activities monitoring using wearable acceleration sensors attached to the body," *PLoS ONE*, vol. 10, no. 7, Jul. 2015, Art. no. e0130851.

[5] Y.-L. Hsu, S.-C. Yang, H.-C. Chang, and H.-C. Lai, "Human daily and sport activity recognition using a wearable inertial sensor network," *IEEE Access*, vol. 6, pp. 31715–31728, 2018.

[6] B. A. M. Hashim and R. Amutha, "Human activity recognition based on smartphone using fast feature dimensionality reduction technique," *J. Ambient Intell. Humanized Comput.*, vol. 12, no. 2, pp. 2365–2374, Feb. 2021.

[7] I. M. Nasir, M. Raza, J. H. Shah, S.-H. Wang, U. Tariq, and M. A. Khan, "HAREDNet: A deep learning based architecture for autonomous video surveillance by recognizing human actions," *Comput. Electr. Eng.*, vol. 99, Apr. 2022, Art. no. 107805.

[8] M. A. Hasan and M. N. Mishuk, "MEMS IMU based pedestrian indoor navigation for smart glass," *Wireless Pers. Commun.*, vol. 101, no. 1, pp. 287–303, Jul. 2018.

[9] Z. Yang, Y. Pan, Q. Tian, and R. Huan, "Real-time infrastructureless indoor tracking for pedestrian using a smartphone," *IEEE Sensors J.*, vol. 19, no. 22, pp. 10782–10795, Nov. 2019.

[10] M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, and P. J. M. Havinga, "Complex human activity recognition using smartphone and wrist-worn motion sensors," *Sensors*, vol. 16, no. 4, p. 426, 2016.

[11] M. H. Rahmani, R. Berkvens, and M. Weyn, "Chest-worn inertial sensors: A survey of applications and methods," *Sensors*, vol. 21, no. 8, p. 2875, Apr. 2021.

[12] O. Bebek, M. A. Suster, S. Rajgopal, M. J. Fu, X. Huang, M. C. Cavusoglu, D. J. Young, M. Mehregany, A. J. van den Bogert, and C. H. Mastrangelo, "Personal navigation via high-resolution gait-corrected inertial measurement units," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 11, pp. 3018–3027, Nov. 2010.

[13] S. Fan, Y. Jia, and C. Jia, "A feature selection and classification method for activity recognition based on an inertial sensing unit," *Information*, vol. 10, no. 10, p. 290, Sep. 2019.

[14] D. Anguita, A. Ghio, L. Oneto, X. P. Perez, and J. L. R. Ortiz, "A public domain dataset for human activity recognition using smartphones," in *Proc. 21st Int. Eur. Symp. Artif. Neural Netw., Comput. Intell. Mach. Learn.*, 2013, pp. 437–442.

[15] O. Banos, J.-M. Galvez, M. Damas, H. Pomares, and I. Rojas, "Window size impact in human activity recognition," *Sensors*, vol. 14, no. 4, pp. 6474–6499, Apr. 2014.

[16] S. Sankar, P. Srinivasan, and R. Saravanakumar, "Internet of Things based ambient assisted living for elderly people health monitoring," *Res. J. Pharmacy Technol.*, vol. 11, no. 9, pp. 3900–3904, 2018.

[17] L. Bai, L. Yao, X. Wang, S. S. Kanhere, B. Guo, and Z. Yu, "Adversarial multi-view networks for activity recognition," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 4, no. 2, p. 42:1–42:22, Jun. 2020.

[18] T. Zebin, P. J. Scully, and K. B. Ozanyan, "Human activity recognition with inertial sensors using a deep learning approach," in *Proc. IEEE Sensors*, Oct. 2016, pp. 1–3.

[19] W. Jiang and Z. Yin, "Human activity recognition using wearable sensors by deep convolutional neural networks," in *Proc. 23rd Int. Conf. Multimedia (ACM)*, Oct. 2015, pp. 1307–1310.

[20] F. Hernandez, L. F. Suarez, J. Villamizar, and M. Altuve, "Human activity recognition on smartphones using a bidirectional LSTM network," in *Proc. 22th Symp. Image, Signal Process. Artif. Vis. (STSIVA)*, Apr. 2019, pp. 1–5.

[21] F. J. Ordóñez and D. Roggen, "Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. 16, no. 1, p. 115, Jan. 2016.

[22] R. Huan, Z. Zhan, L. Ge, K. Chi, P. Chen, and R. Liang, "A hybrid CNN and BLSTM network for human complex activity recognition with multi-feature fusion," *Multimedia Tools Appl.*, vol. 80, no. 30, pp. 36159–36182, Dec. 2021.

[23] N. Dua, S. N. Singh, and V. B. Semwal, "Multi-input CNN-GRU based human activity recognition using wearable sensors," *Computing*, vol. 103, pp. 1461–1478, Mar. 2021.

[24] P. Kasnesis, C. Z. Patrikakis, and I. S. Venieris, "PerceptionNet: A deep convolutional neural network for late sensor fusion," in *Intelligent Systems and Applications* (Advances in Intelligent Systems and Computing), K. Arai, S. Kapoor, and R. Bhatia, Eds. Cham, Switzerland: Springer, 2019, pp. 101–119.

[25] O. Banos, C. Villalonga, R. Garcia, A. Saez, M. Damas, J. A. Holgado-Terriza, S. Lee, H. Pomares, and I. Rojas, "Design, implementation and validation of a novel open framework for agile development of mobile health applications," *Biomed. Eng. OnLine*, vol. 14, no. 2, p. S6, 2015.

[26] A. Reiss and D. Stricker, "Introducing a new benchmarked dataset for activity monitoring," in *Proc. 16th Int. Symp. Wearable Comput. (ISWC)*, Jun. 2012, pp. 108–109.

[27] A. Reiss and D. Stricker, "Creating and benchmarking a new dataset for physical activity monitoring," in *Proc. PETRA*. New York, NY, USA, Jun. 2012, pp. 1–8.

[28] B. Barshan and M. C. Yüksek, "Recognizing daily and sports activities in two open source machine learning environments using body-worn sensor units," *Comput. J.*, vol. 57, no. 11, pp. 1649–1667, 2013.

[29] M. Ullah, H. Ullah, S. D. Khan, and F. A. Cheikh, "Stacked lstm network for human activity recognition using smartphone data," in *Proc. 8th Eur. Workshop Vis. Inf. Process. (EUVIP)*, Oct. 2019, pp. 175–180.

[30] A. Ferrari, D. Micucci, M. Mobilio, and P. Napoletano, "Trends in human activity recognition using smartphones," *J. Reliable Intell. Environ.*, vol. 7, no. 3, pp. 189–213, Sep. 2021.

[31] A. Jalal, M. Batool, and K. Kim, "Stochastic recognition of physical activity and healthcare using tri-axial inertial wearable sensors," *Appl. Sci.*, vol. 10, no. 20, p. 7122, Oct. 2020.

[32] M. S. A. Arani, D. E. Costa, and E. Shihab, "Human activity recognition: A comparative study to assess the contribution level of accelerometer, ECG, and PPG signals," *Sensors*, vol. 21, no. 21, p. 6997, Oct. 2021.

[33] M. Janidarmian, A. R. Fekr, K. Radecka, and Z. Zilic, "A comprehensive analysis on wearable acceleration sensors in human activity recognition," *Sensors*, vol. 17, no. 3, p. 529, 2017.

[34] P. Casale, O. Pujol, and P. Radeva, "Human activity recognition from accelerometer data using a wearable device," in *Pattern Recognition and Image Analysis* (Lecture Notes in Computer Science), J. Vitrià, J. M. Sanches, and M. Hernández, Eds. Berlin, Germany: Springer, 2011, pp. 289–296.

[35] A. Ayman, O. Attalah, and H. Shaban, "An efficient human activity recognition framework based on wearable IMU wrist sensors," in *Proc. IEEE Int. Conf. Imag. Syst. Techn. (IST)*, Dec. 2019, pp. 1–5.

[36] S. Mekruksavanich and A. Jitpattanakul, "Exercise activity recognition with surface electromyography sensor using machine learning approach," in *Proc. Joint Int. Conf. Digit. Arts, Media Technol. With ECTI Northern Sect. Conf. Electr., Electron., Comput. Telecommun. Eng. (ECTI DAMT NCON)*, Mar. 2020, pp. 75–78.

[37] A. M. Helmi, M. A. A. Al-Qaness, A. Dahou, R. Damaševičius, T. Krilavičius, and M. A. Elaziz, "A novel hybrid gradient-based optimizer and grey wolf optimizer feature selection method for human activity recognition using smartphone sensors," *Entropy*, vol. 23, no. 8, p. 1065, Aug. 2021.

[38] H. Fan and H. Ling, "Parallel tracking and verifying: A framework for real-time and high accuracy visual tracking," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*. Venice, Italy, Oct. 2017, pp. 5487–5495.

[39] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690–6709, Apr. 2019.

[40] Z. Zhang, J. Geiger, J. Pohjalainen, A. E.-D. Mousa, W. Jin, and B. Schuller, "Deep learning for environmentally robust speech recognition: An overview of recent developments," *ACM Trans. Intell. Syst. Technol.*, vol. 9, no. 5, pp. 49:1–49:28, Apr. 2018.

[41] M. Gholamrezaii and S. AlModarresi, "A time-efficient convolutional neural network model in human activity recognition," *Multimedia Tools Appl.*, vol. 80, no. 13, pp. 19361–19376, May 2021.

[42] J. Sena, J. Barreto, C. Caetano, G. Cramer, and W. R. Schwartz, "Human activity recognition based on smartphone and wearable sensors using multiscale DCNN ensemble," *Neurocomputing*, vol. 444, pp. 226–243, Jul. 2021.

[43] P. Zappi, C. Lombriser, T. Stiefmeier, E. Farella, D. Roggen, L. Benini, and G. Tröster, "Activity recognition from on-body sensors: Accuracy-power trade-off by dynamic sensor selection," in *Wireless Sensor Networks* (Lecture Notes in Computer Science), R. Verdone, Ed. Berlin, Germany: Springer, 2008, pp. 17–33.
[44] D. Roggen, A. Calatroni, M. Rossi, T. Holleczek, K. Förster, G. Tröster, P. Lukowicz, D. Bannach, G. Pirkl, A. Ferscha, J. Doppler, C. Holzmann, M. Kurz, G. Holl, R. Chavarriaga, H. Sagha, H. Bayati, M. Creatura, and J. D. R. Millàn, "Collecting complex activity datasets in highly rich networked sensor environments," in *Proc. 7th Int. Conf. IEEE Netw. Sens. Syst. (INSS)*, Jun. 2010, pp. 233–240.
[45] T. Lv, X. Wang, L. Jin, Y. Xiao, and M. Song, "Margin-based deep learning networks for human activity recognition," *Sensors*, vol. 20, no. 7, p. 1871, Mar. 2020.
[46] F. Li, K. Shirahama, M. A. Nisar, L. Köping, and M. Grzegorzek, "Comparison of feature learning methods for human activity recognition using wearable sensors," *Sensors*, vol. 18, no. 2, p. 679, 2018.
[47] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman, "Activity recognition from accelerometer data," *Amer. Assoc. Artif. Intell.*, vol. 5, pp. 1541–1546, Jul. 2005.
[48] R. Saeedi, K. Sasani, S. Norgaard, and A. H. Gebremedhin, "Personalized human activity recognition using wearables: A manifold learning-based knowledge transfer," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 1193–1196.
[49] A. Hoelzemann, N. Sorathiya, and K. Van Laerhoven, "Data augmentation strategies for human activity data using generative adversarial neural networks," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops other Affiliated Events (PerCom Workshops)*, Mar. 2021, pp. 8–13.
[50] F. Cruciani, C. D. Nugent, J. M. Quero, I. Cleland, P. Mccullagh, K. Synnes, and J. Hallberg, "Personalizing activity recognition with a clustering based semi-population approach," *IEEE Access*, vol. 8, pp. 207794–207804, 2020.
[51] C.-Y. Lin and R. Marculescu, "Model personalization for human activity recognition," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PerCom Workshops)*, Mar. 2020, pp. 1–7.
[52] A. Akbari and R. Jafari, "Personalizing activity recognition models through quantifying different types of uncertainty using wearable sensors," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 9, pp. 2530–2541, Sep. 2020.
[53] A. Hosseini, D. Zamanzadeh, L. Valencia, R. Habre, A. A. T. Bui, and M. Sarrafzadeh, "Domain adaptation in children activity recognition," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 1725–1728.
[54] K. Zhang, J. Chen, J. Wang, Y. Leng, C. W. de Silva, and C. Fu, "Gaussian-guided feature alignment for unsupervised cross-subject adaptation," *Pattern Recognit.*, vol. 122, Feb. 2022, Art. no. 108332.
[55] J. Zhao, F. Deng, H. He, and J. Chen, "Local domain adaptation for cross-domain activity recognition," *IEEE Trans. Human-Mach. Syst.*, vol. 51, no. 1, pp. 12–21, Feb. 2021.
[56] S. Hajifar, S. R. Lamooki, L. A. Cavuoto, F. M. Megahed, and H. Sun, "Investigation of heterogeneity sources for occupational task recognition via transfer learning," *Sensors*, vol. 21, no. 19, p. 6677, Oct. 2021.
[57] E. Soleimani and E. Nazerfard, "Cross-subject transfer learning in human activity recognition systems using generative adversarial networks," *Neurocomputing*, vol. 426, pp. 26–34, Feb. 2021.
[58] A. Chakma, A. Z. M. Faridee, M. A. A. H. Khan, and N. Roy, "Activity recognition in wearables using adversarial multi-source domain adaptation," *Smart Health*, vol. 19, Mar. 2021, Art. no. 100174.
[59] Z. Zhou, Y. Zhang, X. Yu, P. Yang, X.-Y. Li, J. Zhao, and H. Zhou, "XHAR: Deep domain adaptation for human activity recognition with smart devices," in *Proc. 17th Annu. IEEE Int. Conf. Sens., Commun., Netw. (SECON)*, Jun. 2020, pp. 1–9.
[60] C. F. S. Leite and Y. Xiao, "Improving cross-subject activity recognition via adversarial learning," *IEEE Access*, vol. 8, pp. 90542–90554, 2020.
[61] S. Wan, L. Qi, X. Xu, C. Tong, and Z. Gu, "Deep learning models for real-time human activity recognition with smartphones," *Mobile Netw. Appl.*, vol. 25, pp. 743–755, Dec. 2019.
[62] C. Hou, "A study on IMU-based human activity recognition using deep learning and traditional machine learning," in *Proc. 5th Int. Conf. Comput. Commun. Syst. (ICCCS)*, May 2020, pp. 225–234.
[63] G. De Leonardis, S. Rosati, G. Balestra, V. Agostini, E. Panero, L. Gastaldi, and M. Knaflitz, "Human activity recognition by wearable sensors : Comparison of different classifiers for real-time applications," in *Proc. IEEE Int. Symp. Med. Meas. Appl. (MeMeA)*, Jun. 2018, pp. 1–6.
[64] S. Angerbauer, A. Palmanshofer, S. Selinger, and M. Kurz, "Comparing human activity recognition models based on complexity and resource usage," *Appl. Sci.*, vol. 11, no. 18, p. 8473, Sep. 2021.
[65] Y. Saez, A. Baldominos, and P. Isasi, "A comparison study of classifier algorithms for cross-person physical activity recognition," *Sensors*, vol. 17, no. 1, p. 66, 2017.
[66] P. Kumar and S. Suresh, "Deep learning models for recognizing the simple human activities using smartphone accelerometer sensor," *IETE J. Res.*, pp. 1–11, Sep. 2021. [Online]. Available: https://www.engineeringvillage.com/app/doc/?docid=cpx_e28b51a17bdf992714M7e841017816328, doi: 10.1080/03772063.2021.1967792.
[67] Z. Qin, Y. Zhang, S. Meng, Z. Qin, and K.-K.-R. Choo, "Imaging and fusing time series for wearable sensor-based human activity recognition," *Inf. Fusion*, vol. 53, pp. 80–87, Jan. 2020.
[68] J. Maitre, K. Bouchard, and S. Gaboury, "Alternative deep learning architectures for feature-level fusion in human activity recognition," *Mobile Netw. Appl.*, vol. 26, no. 5, pp. 2076–2086, Oct. 2021.

ZHE YANG received the B.S. and Ph.D. degrees from the Department of Information Science & Electronic Engineering, Zhejiang University, Hangzhou, China, in 2014 and 2019, respectively. From 2017 to 2018, he visited the Department of Electrical and Computer Engineering, University of California at Santa Barbara, Santa Barbara, CA, USA, as a Visiting Scholar. He is currently a Postdoctoral Researcher with the Department of Information Science & Electronic Engineering, Zhejiang University. His research interests include robotic vision, sensor fusion, machine learning, and mobile embedded systems.

MENGJIE QU received the B.S. degree from the Department of Electronic & Information Engineering, Hangzhou Dianzi University, Hangzhou, China, in 2016. He is currently pursuing the master's degree with the Department of Engineering, Zhejiang University, Hangzhou. His research interests include human activity recognition and machine learning.

YUN PAN (Member, IEEE) received the B.S. degree from the Department of Information Science & Electronic Engineering, Zhejiang University, Hangzhou, China, in 2002, and the Ph.D. degree in electronic engineering from Tsinghua University, Beijing, China, in 2008. He visited the Expertise Centre for Digital Media (EDM), Hasselt University, Belgium, in 2013, and then, from 2013 to 2015, he visited the Department of Electrical and Computer Engineering, University of California at Santa Barbara, Santa Barbara, USA. He is currently a Faculty Member of the College of Information Science & Electronic Engineering, Zhejiang University, as an Associate Professor and a Ph.D. Advisor with the honor of Distinguished Young Scholar. His current research interests include mobile computing, mobile healthcare microsystems, on-chip communication, application-specific heterogeneous architecture design, and smart camera systems.

RUOHONG HUAN received the B.S. and M.S. degrees from the Department of Information Science and Electronic Engineering, Zhejiang University, China, in 2002 and 2005, respectively, and the Ph.D. degree from the Institute of Electronics, Chinese Academy of Sciences, China, in 2008. She is currently an Associate Professor with the College of Computer Science and Technology, Zhejiang University of Technology, China. Her current research interests include image processing and target recognition, video processing, and human behavior recognition.