IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received May 19, 2020, accepted June 3, 2020, date of current version August 4, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3003162

Harmonic Loss Function for Sensor-Based Human Activity Recognition Based on LSTM Recurrent Neural Networks

YUE HU[1,2], XIAO-QING ZHANG[3], LI XU[4], FENG XIAN HE[2,5], ZHAO TIAN[1], WEI SHE[1,4,5], (Member, IEEE), AND WEI LIU[1,4]

[1]School of Software, Zhengzhou University, Zhengzhou 450001, China
[2]Yellow River Institute of Hydraulic Research, Zhengzhou 450003, China
[3]Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China
[4]Collaborative Innovation Center for Internet Healthcare, Zhengzhou University, Zhengzhou 450001, China
[5]Research Center on Levee Safety Disaster Prevention, Zhengzhou 450003, China

Corresponding author: Zhao Tian (tianzhao@zzu.edu.cn)

This work was supported in part by the Open Project Fund of the Research Center on Levee Safety Disaster Prevention under Grant 2018007, in part by the National Key Research and Development Program of China under Grant 2018YFB1201403, and in part by the CERNET Innovation Project under Grant NGII20180702.

ABSTRACT Human activity recognition (HAR) has been a very popular field in both real practice and theoretical research. Over the years, a number of many-vs-one Long Short-Term Memory (LSTM) models have been proposed for the sensor-based HAR problem. However, how to utilize sequence outputs of them to improve the HAR performance has not been studied seriously. To solve this problem, we present a novel loss function named harmonic loss, which is utilized to improve the overall classification performance of HAR based on baseline LSTM networks. First, label replication method is presented to duplicate true labels at each sequence step in many-vs-one LSTM networks, thus each sequence step can generate a local error and a local output. Then, considering the importance of different local errors and inspired by the Ebbinghaus memory curve, the harmonic loss is proposed to give unequal weights to different local errors based on harmonic series equation. Additionally, to improve the overall classification performance of HAR, integrated methods are utilized to exploit the sequence outputs of LSTM models based on harmonic loss and ensemble learning strategy. Finally, based on the LSTM model construction and hyper-parameter setting, extensive experiments are conducted. A series of experimental results demonstrate that our harmonic loss significantly achieves higher macro-F1 and accuracy than strong baselines on two public HAR benchmarks. Compared with previous state-of-art methods, our proposed methods can achieve competitive classification performance.

INDEX TERMS Human activity recognition, label replication, harmonic loss, LSTM.

I. INTRODUCTION

In recent years, human activity recognition (HAR) has been attracting more and more attention from both industrial and academic field [1]–[5]. HAR can assist us to make efficient decisions for future human actions through sensory data and has been applied in various applications, such as health monitoring, human-computer interactions, industrial settings, and smart homes [6]–[8], etc. There are many types of human activities (walking, sitting, eating and running, etc) in real life, and researchers have proposed lots of methods to recognize these human activities, however, these methods can not achieve desirable overall classification results. Hence, it is important and necessary to develop advanced methods to get good recognition results of various human activities.

HAR aims to identify the physical activity performed by a person, which is built on the assumption that human activities can be transferred to specific wireless sensor signals. A lot of methods have been proposed to address HAR problems, and we divided them into two main categories: traditional machine learning methods and deep learning methods. Traditional machine learning methods range from support vector machine (SVM) [9], [10], decision tree (DT) [11], random forest (RF) [12], Gaussian Mixture [13], K-nearest neighbor

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

(KNN) [14], to Naïve Bayes [15]. Though these methods can achieve good accuracy results on public HAR benchmarks, they receive poor results on macro-F1. Recently, with the rapid development of artificial intelligence (AI) technology [16], massive deep learning methods including forward neural networks (FNNs) [17], [18], recurrent neural networks (RNNs) [19]–[22], autoencoder (AE) [23], and convolutional neural networks (CNNs) [18], [24]–[27] have been proposed to recognize different types of human activities. In addition to supervised deep learning methods, in literature [28]–[31], researchers have gradually used deep unsupervised learning methods to solve classification problems and achieved state-of-the-art performance like deep clustering and structured autoencoders. These deep learning methods also have achieved state-of-the-art accuracy results, however, the overall classification performance are poor, e.g., macro-F1.

Furthermore, HAR aims to provide fruitful information for user's activities by leveraging time-series sensory data collected from surrounding environments [32], which can be considered as a streaming data mining problem. Collected sensory data for HAR is chronological, which is also known as stream data. Recurrent neural networks (RNNs), particularly those based on the LSTM network and its various variants, e.g., Gated Recurrent Unit (GRU) [33], can capture long-range dependencies efficiently and mine fruitful information from stream data strongly [34], [35]. Many complicated LSTM models have been proposed and they can achieve high accuracy of HAR. Nevertheless, the macro-F1 results of LSTM models are poor on many public HAR benchmarks too, because researchers mainly focused on developing advanced many-vs-one LSTM models for improving accuracy as the following steps. First, many-vs-one LSTM models are constructed by using many LSTM cells. Then many sequential inputs are passed into many-vs-one LSTM models, only generating the final sequence output and error in the training phase, which cannot make use of all sequential outputs and sequential errors. To the best of our knowledge, how to exploit sequential outputs of many-vs-one LSTM models for enhancing the performance of HAR has not been studied seriously by previous works.

To tackle the above-mentioned limitations of many-vs-one LSTM models, this work mainly focuses on exploiting sequential outputs and errors for achieving better overall classification performance. Firstly, the label replication method is presented to copy true labels at every sequenced step of LSTM models, thus, every sequenced step can generate a local output and local error. Then, getting inspiration from the Ebbinghaus memory curve of humans [36], a novel loss function named harmonic loss (HL) is proposed based on the harmonic series equation. The proposed harmonic loss takes all local errors into consideration by giving unequal weights to each local error rather than only computes the last local error in the training phase. In addition, integrated methods are proposed to utilize sequence outputs of many-vs-one LSTM models based on our harmonic loss and ensemble learning strategy. It can further enhance the overall classification of HAR. The experimental results demonstrate that the proposed harmonic loss can achieve higher macro-F1 than strong baselines significantly and more stable. Compared with previous advanced methods, our proposed methods achieve competitive overall classification performance. The main contributions of this work are summarized as follows:

1) Label replication method is utilized for many-vs-one LSTM models, which enables many-vs-one LSTM models to generate local errors and local outputs at each sequenced step.
2) Getting inspiration from the Ebbinghaus memory curve of humans, we present an improved loss titled harmonic loss, which computes all local errors of LSTM models by giving different weights to each local error based on the harmonic series equation. Compared with traditional losses for LSTM models only compute final sequence error for the gradient descent, our harmonic loss not only computes all sequence errors for the gradient descent but also considers the weights of different sequence errors. Thus, our harmonic loss achieves better overall classification than strong baselines.
3) Based on both ensemble learning strategy and the harmonic loss, integrated methods are introduced to make use of sequence outputs of many-vs-one LSTM models. The results demonstrate that integrated methods surpass previous state-of-art methods.

The rest of this work is organized as follows. Section II introduces the preliminary works of HAR, LSTM models, and loss functions. The framework of sensor-based human activity recognition using LSTM, label replication method, our harmonic loss, integrated methods, and evaluation measures are described in Section III. Section IV presents public HAR benchmarks and baselines. Section V provides a series of experimental results and analysis. Conclusion and future work are provided in Section VI.

II. RELATED WORK

HAR aims to understand and predict people's activities and behaviors, which can enable the intelligent computing systems to assist users proactively based on their requirement through the sensory data [32]. A lot of benchmark datasets have been introduced to meet the need of HAR, which promots the emergence and development of HAR domain. HAR datasets comprise of self data collection and public datasets. Self data collection refers to researchers collected data by themselves and performed relevant research, which requires much efforts and is also tedious to preprocess the collected data [37], [38]. Public datasets refer to researchers performed work on public HAR datasets, such as Opportunity dataset [17], public domain UCI dataset [10], [39], and WISDM dataset [40].

Recurrent neural networks (RNNs) are feed-forward neural networks augmented by the inclusion of edges that span adjacent sequence steps and can process sequential data efficiently in supervised or unsupervised learning tasks. LSTM

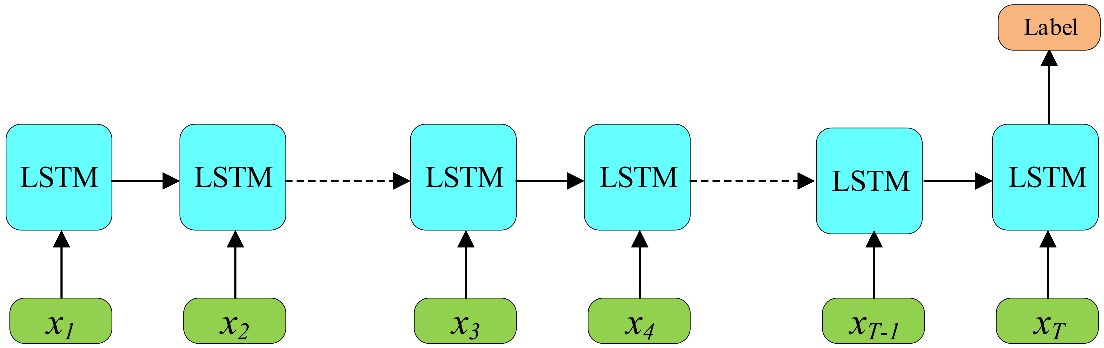


FIGURE 1. Representative many-vs-one LSTM model for HAR. Green rectangles represent input sequence. Blue rectangles represent LSTM blocks (layers). The orange rectangle represents labels.

is one of the most popular RNN architectures, which can alleviate vanishing gradient problem of RNNs. The construction of the LSTM memory cell is still close to the original RNN memory cell, in addition to the forget gates. The forget gates can make the cell remember its most previous values by setting large values. To understand LSTM models easily, Fig. 1 presents the classical LSTM network model, which is used for comparison in the following sections. Massive variants of RNNs have been developed for various learning tasks [41], [42] including HAR [19]–[22], [43]–[47]. Literature [22] presents a deep Res-LSTM model for HAR and achieves good results of HAR. In [43], researchers applied a deep recurrent neural network (DRNN) model and the results showed the DRNN got good performance. Researchers in literature [24], [47] proposed hybrid deep models for HAR through combining the RNN model with the CNN model, and the effectiveness of the hybrid deep models was verified on public HAR datasets.

In this work, the connectivity pattern among stacking LSTM layers follows the LSTM models devised by [22], [48]. We use label replication method to replicate true labels at every sequenced step for LSTM models, hence, each sequence step can generate a local sequence error. The label replication method is similar to Lee *et al.* [49] proposed the optimization objective function for convolutional neural network (CNN) and has been widely used in various learning tasks. Literature [50]–[52] use the label replication method for medical data analysis and achieve competitive performance through comparison to advanced methods. Dai & Le proposed label replication method for character-level document classification in semi-supervised sequence learning tasks. Literatures [53], [54] also use similar methods in natural language processing and achieve good performance.

In literature [55], researchers propose an improved loss named convex loss for many-vs-one LSTM models based on the label replication method and achieve good results. Focal loss [56] is a very popular loss and has been applied to different tasks successfully. Literature [57] develops a complement objective training method (COT) to improve the accuracy of classification by introducing a complement entropy loss (CEL) based on entropy cross loss in the training, the results show that the proposed COT can improve the classification results. Given the improved losses for many-vs-one LSTM models based on the label replication method, which cannot consider the relative importance of different sequence errors. To tackle this end, we proposed an improved loss function considering the relative importance of all sequence errors based on the label replication method.

III. METHODOLOGY

In this section, the representative framework of HAR using LSTM models and label replication method are introduced. Then, our harmonic loss is proposed based on the harmonic series equation to optimize the loss function of LSTM models. Finally, the integrated method are presented to further improve the overall classification performance.

A. LSTM RECURRENT NEURAL NETWORK MODEL FOR HAR

The framework of sensor-based activity recognition using LSTM models as shown in Fig. 2. First, mobile devices are used to collect signal data from various sensors. Then, the collected sensory data are used to train the devised many-vs-one LSTM recurrent neural network models. Finally, the trained LSTM model is used to recognize various types of human activities such as running and walking. In this work, we mainly focus on optimizing the loss function, thus, utilizing many-vs-one LSTM models to recognize different types of human activities on public HAR datasets rather than collect sensor-based activity recognition signal data.

B. LABEL REPLICATION METHOD

Given the collected sensory data are sequential, hence, representative LSTM model is usually used to match and recognize the actions performed by a user, as shown in Fig. 1. It passes all inputs in sequential order, which only generates one output and one error at the final sequence step. Thus, the final error calculated at the final sequence step is the average of the batch loss calculated on each example separately can be represented

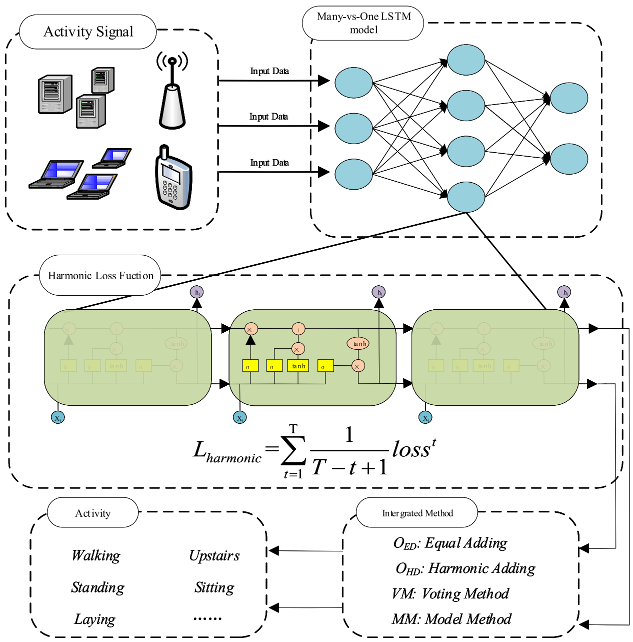


FIGURE 2. Framework of sensor-based activity recognition using LSTM.

as follows:

$$Loss(p,q) = \frac{1}{|M|} \sum_{l=1}^{l=M} -q_l log p_l \tag{1}$$

p and q represent predicted labels and true labels, and M denotes batch size.

One common shortcoming is that classical LSTM models (Fig. 1) must learn to deliver fruitful information across sequence steps, in order to obtain the expected outputs. However, this architecture ignores the effects of local errors of intermediate sequence steps except for last sequence step. To this end, we replicate the static labels at each sequence step as shown in Fig. 3(a). Thus, LSTM models can generate local errors at every sequence step as shown in Fig. 3(b). Researchers have proposed the improved loss function based on label replication method, named convex loss [55], which considers all local sequence errors equally in the training. The convex loss is a convex combination of the final sequence loss and the average of the all sequence losses can be written as follows:

$$L_{convex} = \alpha * \frac{1}{T} \sum_{t=1}^{T} loss^t + (1-\alpha) loss^T \tag{2}$$

where T is the total number of sequence steps. $\alpha \in [0, 1]$ is a hyper-parameter, which determines the relative importance of intermediate predicted outputs. Although, the relative importance of intermediate outputs is unequal in the training phase, Eq. 2 gives the same weights to intermediate predicted outputs. To tackle the above problem, getting inspiration from the Ebbinghaus memory curve, we propose an improved loss function named harmonic loss based on the harmonic series equation and it would be described in the following section.

C. HARMONIC LOSS

To understand our harmonic loss thoroughly, we present the concept of Ebbinghaus memory curve and the harmonic series equation, and followed by detailed construction of harmonic loss.

Memory is a particular process of remembering, maintaining and recognizing for external information. Forgetting is an unavoidable activation of the brain of humans in the process of memory. According to the Ebbinghaus memory curve, the forgetting rule of the human brain has a fixed rule and is imbalanced. The forgetting speed is fast at first but gradually slow later. Analogously, the learning process of many LSTM cells only captures a few key information representations through many sequences and filter out useless information in sequence steps, and the final sequence output receives more key information representations than other sequences. Hence, the relative importance of local errors of different sequence steps is unequal based on the label replication method as shown in Fig. 3(b). Considering similarities between the Ebbinghaus memory curve and the learning process of LSTM cells, we intend to select a function like the Ebbinghaus memory curve to set corresponding weights to different local errors. Based on this idea, we associate the curve of harmonic series equation is similar to the Ebbinghaus memory curve. Hence, the harmonic series equation is selected, which can be represented as following eq. 3 in fundamental math.

$$f(n) = \frac{1}{2} + \frac{2}{3} + \frac{3}{4} + \ldots + \frac{1}{n-1} + \frac{1}{n} \tag{3}$$

Based on the above idea, we propose an improved loss named harmonic loss (HL), which is the sum of all weighted local losses. Moreover, our harmonic loss can be computed as the following equation:

$$L_{harmonic} = \sum_{t=1}^{T} \frac{1}{T-t+1} loss^t \tag{4}$$

where $1/(T - t + 1)$ is the weight coefficient represents different local losses of sequence steps, as shown in Fig. 3 (b). In our harmonic loss, reverse constant coefficients of the harmonic series equation are utilized to set weights for different local losses, that is, the maximum weight is given to final sequence loss and the minimum weight is given to the first sequence loss. Moreover, the harmonic loss scales the proportion of hard classified instances in total losses by summing of all weighted local losses with voting strategy, while previous works mainly proposed static weight hyper-parameters of different predicted values to increase the loss of hard classified instances in total losses. Therefore, the proposed harmonic loss can avoid earlier sequence losses and easily classified instances dominate the gradient in the training.

On the basis of the above analysis, our harmonic loss is more able to predict hard classified examples (labels with few examples) accurately than other previous works and can improve the overall classification performance of HAR. At the prediction phase, the LSTM model with harmonic loss

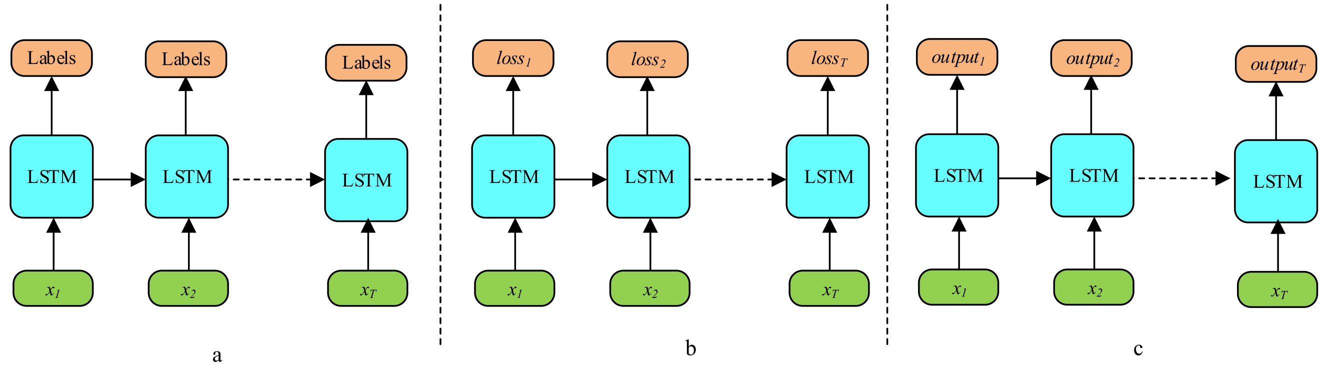


FIGURE 3. (a) LSTM with label replication method, replicating true labels at each sequence step.(b) LSTM model with harmonic loss generates different local losses at each sequence step.(c) Many-vs-one LSTM model generates different sequence outputs at each sequence step with label replication method.

can generate one output at the final sequence step, and also can generate multiple outputs.

D. INTEGRATED METHODS

As previously introduced, representative LSTM models for HAR, as you can see from Fig. 1. It only generates an output at the final sequence step for prediction, yet losing sight of the effects of other sequence outputs for overall classification performance. In the experiments, we find that the predicted results of last several sequence outputs are well in addition to the final sequence prediction result. Inspired by ensemble learning strategy and our harmonic loss, four different integrated methods are proposed to take use of sequence outputs of LSTM models correspondingly. To achieve surpassing classification performance, each sequence step can generate an output for prediction at the prediction phase as shown in Fig. 3(c).

First integrated method named equal adding (ED) method and can be computed as follows:

$$O_{ED} = O_T + O_{T-1} + O_{T-2} + \cdots + O_{T-n+1} \tag{5}$$

where T is the number of sequence steps, and n = 1, 2, 3, ..., T denotes number of sequence outputs are selected for adding, mainly depending on overall classification performance. In the equal adding (ED) method, each sequence output has equal voting rights for final prediction results, which is a shortcoming. Because selected sequence outputs has unequal contributions for prediction in many-vs-one LSTM models.

Given the shortcoming of ED, second integrated method named harmonic adding (HD) method. Considering similarity between our harmonic loss and the Ebbinghaus memory curve, harmonic series equation is used to set different weights for different sequence outputs. HD can be represented by using eq. 6.

$$O_{HD} = \frac{1}{2}O_T + \frac{1}{3}O_{T-1} + \frac{1}{4}O_{T-2} + \cdots + \frac{1}{n+1}O_{T-n+1} \tag{6}$$

In eq. 6, we set the maximum weight to the final sequence output and give smaller weights to intermediate sequence output with the reverse harmonic series equation. The number of sequence outputs is also determined by classification results. For the HD, different outputs have unequal voting rights to decide on the final prediction result, which conduce to improve the overall classification results of HAR based on LSTM models.

Third method is the voting method, where a process is selected from sequence outputs of LSTM models to make the final prediction by a simple majority voting. The voting method makes the decision based on predicted labels rather than the predicted probability distribution. Firstly, several sequence outputs are selected depending on the experiment results. Then combining these sequence outputs through the voting process, offering the chance of correcting potential errors by providing a diverse output space. Furthermore, the voting method has the same shortcoming as the ED method for HAR using many-vs-one LSTM models.

Last integrated method called model ensemble (ME), which adds final sequence outputs of two simple LSTM models (a basic LSTM model and a basic GRU model) together, and then makes the final prediction. ME is one of the most widely used integrated methods for improving classification performance in data Mining competition. Two simple LSTM models utilize harmonic loss as loss function in the training.

E. EVALUATION MEASURES

Accuracy (ACC) is a common evaluation measure to assess the performance of different methods, which is used to qualify the probability for classifying each instance correctly. Eq. (7) present the formula of accuracy:

$$ACC = \frac{\sum_{i=1}^{C} TP_{ii}}{total} = \frac{\sum_{i=1}^{C} TP_{ii}}{\sum_{i=1}^{C} BI_i} = \frac{\sum_{i=1}^{C} TP_{ii}}{\sum_{i=1}^{C} NI_i} \tag{7}$$

where TP_{ii} is the number of correctly classified instances for the inferred label i, NI_i refers to the total number of instances that are classified as label i, and total indicates the total number of test samples.

However, accuracy cannot reflect the overall classification performance of a method well. For example, a method may achieve high accuracy on an HAR task, nevertheless, classification results of some human actions are very poor, which is a common phenomenon on imbalanced HAR datasets. Hence, to assess the overall classification performance of all methods comprehensively, F1 score is also utilized to assess the overall classification performance of all methods. F1 is the harmonic mean of precision and recall, which is an important evaluation measure in the research domain of data mining. F1 can be calculated with Eq. 8:

$$F1 = \frac{2 * precision * recall}{precision + recall}. \tag{8}$$

According to reference [19], macro-F1 and weighted-F1 are able to evaluate the overall classification results better than accuracy. Furthermore, macro-F1 is more important than weighted-F1, which is better to assess overall classification performance of a method thoroughly, especial for label imbalance problem. Fm and Fw can be represented as the following equations:

$$F_m = \frac{2}{|c|} \sum_c \frac{prec_c * recall_c}{prec_c + recall_c} \tag{9}$$

$$F_w = 2 \sum_c \frac{N_c}{N_total} \frac{prec_c * recall_c}{prec_c + recall_c}. \tag{10}$$

where, c denotes the classes of human activities, N_c is the number of instances in class c, and N_{total} is the total number of instances. Therefore, accuracy, macro-F1 and weighted-F1 are used as evaluation measures in this work.

IV. EXPERIMENTS

In this section, two public HAR datasets are firstly described, then the construction of LSTM models, baselines, and parameter settings of integrated methods are presented.

A. DATASETS

To evaluate the effectiveness of our proposed methods thoroughly, we conduct all experiments on two common HAR benchmarks: the public domain UCI dataset (balanced) and the Opportunity dataset (imbalanced). Both benchmarks contain most human activities to reflect real life, and enable to abstract features and labels for modeling.

1) PUBLIC DOMAIN UCI DATASET

The dataset was collected from a group of 30 volunteers with ages ranging from 19 to 48 years. Each volunteer was instructed to perform six daily living activities (standing, sitting, laying down, walking, walking downstairs and upstairs) while wearing a smartphone on the waist. The powerful ambient sensors accelerometer and gyroscope of the smartphone were able to collect tri-axial linear acceleration and tri-axial velocity signals at a sampling rate of 50Hz. Then collected sensor signals were sampled in fixed-width sliding windows of 2.56 second and each window contained 128 time steps. The dataset is randomly partitioned into two independent sets: 70% of the data were selected for generating the training data and the remaining 30% of the data were selected for generating the testing data.

2) OPPORTUNITY DATASET

The Opportunity dataset is another benchmark dataset for evaluating human activity recognition algorithms, which comprises most daily living activities coming from the real environment with collecting diverse sensor data [17]. In this work, "OPPORTUNITY Activity Recognition Challenge" subset is used for HAR, which contains 3 subjects and 6 runs per subject. The Run 4 and Run 5 for Subjects 2 and 3 are selected as the testing data, and the others are selected as the training set. This subset includes 17 different mid-level gesture classes and a "NULL" class. Since collected sensor data is transferred by the wireless network, which may cause missing data, the linear interpolation method is used to fill missing data. For the above two public datasets, we use the same data preprocessing methods to preprocess the datasets as previous works in [22].

B. BASELINES

To verify the effectiveness of the proposed methods, we conduct experiments on two common RNN models. The first one is a simple LSTM RNN model, and another one is GRU RNN model, which is a variant of LSTM. Both LSTM and GRU models are the most used recurrent neural network architectures for modeling sequential data. Each LSTM model has 2 two layers and each layer has 32 neural units in this work, which is the same with previous work [22].

The values of learning rate range from 0.001 to 0.0035, batch size is set for 500, and all methods running iterations ranging from 700 to 1200 in the training. Furthermore, we set the best results in advance and the early stop trick is used in the training for all methods. The proposed harmonic loss is compared with the following baselines based on two above-mentioned LSTM models.

Convex Loss: which is a convex combination of the final loss and the average of all sequence losses [55].

Focal Loss: which scales loss of incorrectly predicted instances with a factor disproportional to the predicted probability [56].

COT: it is a complement objective training method by utilizing cross entropy loss and complement entropy loss together [57].

Vanilla: which uses the original cross entropy loss function without adding extra penalty parameters [58].

L2 Norm: which is a combination of the original cross entropy loss and L2 norm [59].

L1 Norm: which is a combination of the original cross entropy loss and L1 norm [60].

L1&L2 Norm: which uses the original cross entropy loss with adding L1 norm and L2 norm [59].

Static Scaling: which scales the loss of small labels with a constant. This method is simple but effective [61].

TABLE 1. Experimental results of different loss functions on two public datasets.

Datasets-Method	OPPO			UCI		
	Accuracy	Weighted-F1	Macro-F1	Accuracy	Weighted-F1	Macro-F1
LSTM						
Vanilla	88.78%	88.19%	52.80%	91.01%	90.96%	91.10%
L2	88.78%	88.29%	53.16%	91.62%	91.61%	91.50%
L1	88.81%	88.30%	53.37%	91.03%	91.02%	91.11%
L1&L2	88.79%	88.33%	53.41%	91.14%	91.12%	91.16%
Scaling	89.02%	88.51%	54.14%	92.09%	92.00%	92.19%
Focal	88.86%	88.48%	54.09%	91.96%	91.88%	91.99%
Convex	89.15%	88.70%	54.36%	92.23%	92.19%	92.34%
COT	88.81%	88.23%	53.10%	91.14%	91.14%	91.24%
Harmonic	89.69%	89.44%	57.52%	92.98%	92.92%	93.12%
GRU						
Vanilla	88.99%	88.26%	53.46%	91.21%	91.12%	91.24%
L2	88.90%	88.33%	53.51%	91.31%	91.30%	91.27%
L1	88.99%	88.33%	53.48%	91.24%	91.14%	91.24%
L1&L2	88.87%	88.29%	53.47%	91.32%	91.30%	91.44%
Scaling	89.12%	88.54%	55.37%	91.99%	91.89%	92.06%
Focal	88.97%	88.51%	54.25%	91.92%	91.90%	91.90%
Convex	89.20%	88.80%	55.82%	92.06%	92.00%	92.13%
COT	89.12%	88.41%	53.76%	91.56%	91.39%	91.67%
Harmonic	89.78%	89.57%	57.67%	92.85%	92.84%	92.93%

Additionally, experiment settings for four ensemble methods as follows:Equal Adding (ED) method only selects last two sequence outputs for comparison based on experiment results. For harmonic adding (HD) method, last two sequence outputs and all sequence outputs are selected, named HD_2 and HD_all correspondingly. For voting method, last three sequence outputs are selected. The above three integrated methods only use basic LSTM models.The model ensemble is an integration of two baseline LSTM models (simple LSTM and GRU). Furthermore, the best running results of each algorithm are reported for comparison.

V. RESULTS ANALYSIS AND DISCUSSION

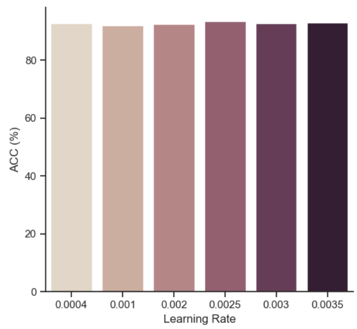
A. HYPER-PARAMETER RESULT ANALYSIS

Fig. 4 presents the accuracies of learning rate and batch size based on the basic LSTM model with harmonic loss over the public UCI dataset. Fig.4(a) shows the results of different learning rates, the basic LSTM achieves the best accuracy when the learning rate is set for 0.0035, but gaps among different learning rate are small. Hence, we set the initial learning rate for 0.001 and change the learning rate dynamically when the network stops improving accuracy after five epochs. In the experiments, we set the range of learning rates from 0.001 to 0.005. Fig.4(b) gives the accuracy of batch size, the basic LSTM achieve outperforms other batch sizes when we set batch size for 500. It is very difficult and challenging to set proper batch size, because it is associated with many factors, such the sampling distribution of each batch size, label distribution, loss function, and learning rate. In the following experiments, we adopt the same batch size and the learning rate for all loss functions.

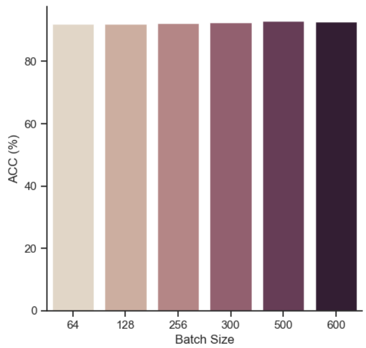
B. PERFORMANCE COMPARISON OF LOSS FUNCTIONS

Table 1 presents experimental results of different loss functions on both public HAR datasets over two LSTM models. Furthermore, to compare the overall classification performance of different loss functions visually, Fig. 5 gives compared results of all loss functions on three evaluation measures.

From Table 1 and Fig. 5, it can be concluded that compared with the vanilla, all other losses show different improvements on three evaluation measures. Especially, the harmonic loss achieves the best classification performance. Moreover, our harmonic loss achieves similar improvements on accuracy and weighted-F1 on both datasets over both models and increases by approximately 0.7% through comparison to other advanced losses. Macro F1 is an important indicator for assessing the overall classification performance of a method as above mentioned. Macro-F1 of all methods is smaller than accuracy and weighted-F1 on the Opportunity dataset as you



(a) Learning rate.



(b) Batch size.

FIGURE 4. Performance comparison of different hyperparameters.

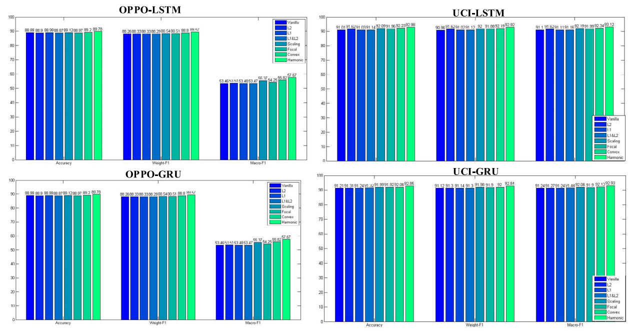


FIGURE 5. Result comparison of loss functions.

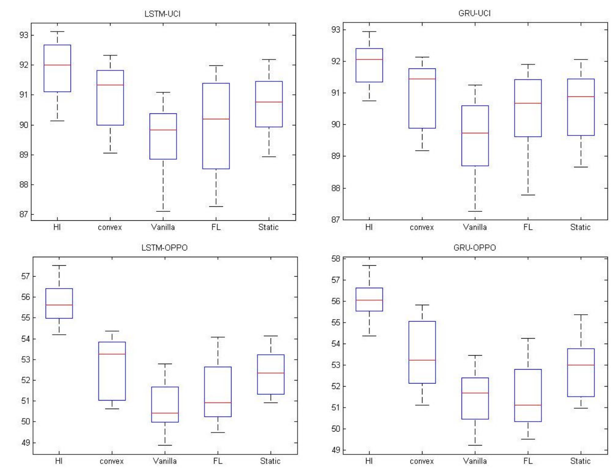


FIGURE 6. Box plots of five different methods.

can see from Table 1 and Fig. 5. Label imbalanced problem of the Opportunity dataset can answer for the results of macro-F1, and the largest class imbalance between two classes is 137-fold. Our harmonic loss achieves the best macro-F1 with only 57.52% and 57.67% on the Opportunity dataset over two LSTM models correspondingly, which increases about 2% than other compared methods, but it still has huge room for improvement in the future. According to the results of harmonic loss, it shows that our harmonic loss not only takes all local errors into consideration but also avoids early local errors dominate the maximum proportion of loss by the harmonic series equation.

Given expeimental results of Table 1 and Fig. 5, Norm technique is unable to achieve competitive performance, because it cannot handle the over-fitting problem efficiently by increasing the sparsity of weights of hidden layers. Focal loss cannot achieve expected results, likely since the dataset is not large, and hyper-parameters of focal loss is not optimum, which are difficult to set. Earlier sequence losses occupy the most proportion of the convex loss, which indicates that convex loss cannot achieve the performance as we expect. Additionally, our harmonic loss is distinct from convex loss, which gives different importance to each sequence loss by setting corresponding weights via harmonic series equation. The LSTM gets better performance than the GRU on the public domain UCI dataset in general, however, the GRU gets better performance than the LSTM on the Opportunity dataset. It shows different LSTM models can not achieve excellent results on all datasets, but the proposed harmonic loss works well.

C. STABILITY ANALYSIS OF LOSS FUNCTIONS

Fig. 6 shows the plot box of our harmonic loss and four advanced loss functions with macro-F1 on both datasets. It can be inferred that all contrastive methods have analogous fluctuation, which also indicates that it is difficult to train LSTM models. Based on the experimental results, our harmonic loss and convex loss has the smallest fluctuation of macro-F1. However, focal loss has the biggest fluctuation of weighted-F1. The mean macro-F1 of harmonic loss is 92.01% on the public UCI dataset, which is higher than other baselines. Furthermore, it can be seen that all methods are more unstable on the Opportunity dataset and the results are poor because data distribution is skewed.

Static scaling sets the importance of different classes statically in the entire training. However, the relative importance of different labels changes with every iteration dynamically in the training, subsequently, static scaling cannot achieve stable performance in the training procedure. Though focal loss uses the modulating factor to change the weights between correctly classified classes and incorrectly classified classes, the tunable focusing parameter of the modulating factor is set manually, which can result in instability. Harmonic loss

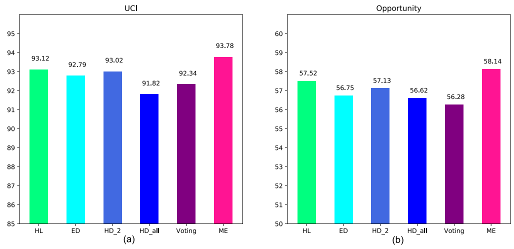


FIGURE 7. Comparison of four integrated methods based on baseline LSTM model.

TABLE 2. Comparison of this work and existing methods.

UCI Dataset		OPPO Dataset	
Model	Accuracy (%)	Model	F1 (%)
Deep Res-LSTM [22]	91.6	Bidir-LSTM [22]	89.2
MCHF-SVM [10]	89.3	1NN [17]	85.0
Baseline LSTM [22]	90.8	Deep Res-LSTM [22]	90.2
HCF+ANN [18]	91.08	CNN [24]	88.3
PCA [60]	87.7	CNN [26]	85.1
CNN [25]	90.0	DBN [26]	78.03
CNN [27]	92.35	SVM [9]	80.4
CNN [27]	90.77	Baseline LSTM [22]	88.2
Harmonic	92.98	Haromnic	89.57
ME	93.79	ME	90.36

and convex loss are capable of changing the importance of different classes in different phases of training via sum predicted errors of sequence steps. However, the earlier sequence losses of convex loss dominate the gradient, which makes it ignore the importance of later sequence losses easily. But our proposed harmonic loss manages to handle it by giving unequal weights, which conduces to the stability and the improvement of classification performance.

D. PERFORMANCE COMPARISON OF INTEGRATED METHODS

Fig. 7 presents the macro-F1 of four output integrated methods on both datasets. HL denotes the LSTM model with harmonic loss. HD_2 and HD_all denote harmonic adding method, and the main difference between them is the number of sequence outputs is unequal. Model ensemble method achieves the best macro-F1 with 93.87% and 58.14% on both public benchmarks, which demonstrates that introducing ensemble learning strategy to exploit the advantages of basic LSTM model is an effective method to get promising overall classification performance. HD_all method gets the worst performance among all methods on the public domain UCI dataset, and voting method achieves the worse performance than other methods on the Opportunity dataset. Harmonic adding methods do not get expected performance, though the performance of HD_2 and HL is similar.

E. PERFORMANCE COMPARISONS OF STATE-OF-ART METHODS

Table 2 presents the HAR results of the proposed methods in this work and previous state-of-art methods. In general, the proposed methods outperform most previous state-of-art methods about 1.2%. The proposed ensemble method achieves the best accuracy with 93.79% on the public UCI dataset and the best weighted-F1 with 90.36% on the Opportunity dataset, which is a simple ensemble of two baseline LSTM models based on the harmonic loss. Our harmonic loss based on the baseline LSTM model gets 92.98% accuracy on the public UCI dataset, and also gets 89.57% weighted-F1 on the Opportunity dataset respectively. It outperforms the baseline LSTM model by approximately 1.5% on two public HAR datasets without devising complex LSTM models. Moreover, compared with advanced variants of LSTM models (Bidir-LSTM and deep Res-LSTM), our harmonic loss enables the baseline LSTM model to achieve competitive performance, which also demonstrates the effectiveness of the proposed harmonic loss. One likely reason to explain that the proposed harmonic loss increases the proportion of easily incorrect classes dynamically in the training phase, which not only takes all local sequence errors into accounts but also considers the relative importance of different local errors. Hence, How to exploit sequential outputs of many-vs-one LSTM models for boosting the performance of HAR is a possible research direction, because it has not been studied seriously.

In the future, we can apply RNN models with harmonic loss to process other stream data such as electrocardiograph (ECG) signals, electroencephalogram (EEG) signals, and clinical time series data, because harmonic loss can make RNN models gain better classification performance. In the future, we will apply complex RNN models and CNN-RNN models to test the performance of the proposed harmonic loss.

VI. CONCLUSION AND FUTURE WORK

To achieve competitive overall classification performance of HAR with simple LSTM models, label replication method, harmonic loss, and integrated methods are studied in this work. We apply the label replication method to replicate true labels at each sequence step of LSTM models. Based on the label replication method, an improved loss function named harmonic loss is introduced, which not only takes all local sequence errors into accounts but also considers the relative importance of different local errors in the training. Then integrated methods based on ensemble learning strategy and harmonic loss are presented and analyzed. Finally, after introducing the experiment environment setup, HAR datasets, the construction of LSTM models, baselines, the parameters of integrated methods are provided, and extensive experiments are conducted. Experiment results demonstrate that, compared with focal loss, convex loss, and static scaling, our harmonic loss can not only achieve higher results of three evaluation measures but also achieve greater stability. Furthermore, our proposed methods also surpass previous state-of-art methods.

In the future, the improved harmonic loss functions and other novel loss functions will be proposed for both CNNs and RNNs to improve the classification performance of different learning tasks.

ACKNOWLEDGMENT

(Yue Hu and Xiao-Qing Zhang contributed equally to this work.)

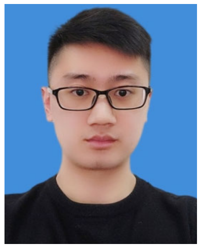
REFERENCES

[1] A. Wang, G. Chen, J. Yang, S. Zhao, and C.-Y. Chang, "A comparative study on human activity recognition using inertial sensors in a smartphone," *IEEE Sensors J.*, vol. 16, no. 11, pp. 4566–4578, Jun. 2016.
[2] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: A survey," *Pattern Recognit. Lett.*, vol. 119, pp. 3–11, Mar. 2019.
[3] H. F. Nweke, Y. W. Teh, M. A. Al-Garadi, and U. R. Alo, "Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges," *Expert Syst. Appl.*, vol. 105, pp. 233–261, Sep. 2018.
[4] S. R. Ramamurthy and N. Roy, "Recent trends in machine learning for human activity recognition—A survey," *Wiley Interdiscipl. Rev., Data Mining Knowl. Discovery*, vol. 8, no. 4, p. e1254, Jul. 2018.
[5] A. D. Antar, M. Ahmed, and M. A. R. Ahad, "Challenges in sensor-based human activity recognition and a comparative analysis of benchmark datasets: A review," in *Proc. Joint 8th Int. Conf. Informat., Electron. Vis. (ICIEV), 3rd Int. Conf. Imag., Vis. Pattern Recognit. (icIVPR)*, May 2019, pp. 134–139.
[6] A. Secerbegovic, A. Gogic, N. Suljanovic, M. Zajc, and A. Mujcic, "Computational balancing between wearable sensor and smartphone towards energy-efficient remote healthcare monitoring," *Adv. Electr. Comput. Eng.*, vol. 18, no. 4, pp. 3–11, 2018.
[7] H. Kalantarian, C. Sideris, B. Mortazavi, N. Alshurafa, and M. Sarrafzadeh, "Dynamic computation offloading for low-power wearable health monitoring systems," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 3, pp. 621–628, Mar. 2017.
[8] X. Zhang, H. Zhao, S. Zhang, and R. Li, "A novel deep neural network model for multi-label chronic disease prediction," *Frontiers Genet.*, vol. 10, p. 351, Apr. 2019. [Online]. Available: https://www.frontiersin.org/article/10.3389/fgene.2019.00351
[9] H. Cao, M. N. Nguyen, C. Phua, S. Krishnaswamy, and X.-L. Li, "An integrated framework for human activity classification," in *Proc. ACM Conf. Ubiquitous Comput. (UbiComp)*, 2012, pp. 331–340.
[10] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. R. Reyes-Ortiz, "Energy efficient smartphone-based activity recognition using fixed-point arithmetic," *J. Universal Comput. Sci.*, vol. 19, no. 9, pp. 1295–1314, 2013.
[11] P. Casale, O. Pujol, and P. Radeva, "Human activity recognition from accelerometer data using a wearable device," in *Proc. Iberian Conf. Pattern Recognit. Image Anal.* Berlin, Germany: Springer, 2011, pp. 289–296.
[12] A. Bayat, M. Pomplun, and D. A. Tran, "A study on human activity recognition using accelerometer data from smartphones," *Procedia Comput. Sci.*, vol. 34, pp. 450–457, Jan. 2014.
[13] Y. Kwon, K. Kang, and C. Bae, "Unsupervised learning for human activity recognition using smartphone sensors," *Expert Syst. Appl.*, vol. 41, no. 14, pp. 6067–6074, Oct. 2014.
[14] S. Kaghyan and H. Sarukhanyan, "Activity recognition using k-nearest neighbor algorithm on smartphone with tri-axial accelerometer," *Int. J. Inform. Models Anal.*, vol. 1, pp. 146–156, Jul. 2012.
[15] M. M. Hassan, M. Z. Uddin, A. Mohamed, and A. Almogren, "A robust human activity recognition system using smartphone sensors and deep learning," *Future Gener. Comput. Syst.*, vol. 81, pp. 307–313, Apr. 2018.
[16] H. Wu, Q. Liu, and X. Liu, "A review on deep learning approaches to image classification and object segmentation," *Comput., Mater. Continua*, vol. 1, no. 2, pp. 1–5, 2018.
[17] H. Sagha, S. T. Digumarti, J. del R. Millan, R. Chavarriaga, A. Calatroni, D. Roggen, and G. Tröster, "Benchmarking classification techniques using the opportunity human activity dataset," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2011, pp. 36–40.
[18] C. A. Ronao and S.-B. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks," *Expert Syst. Appl.*, vol. 59, pp. 235–244, Oct. 2016.
[19] N. Y. Hammerla, S. Halloran, and T. Plötz, "Deep, convolutional, and recurrent models for human activity recognition using wearables," 2016, *arXiv:1604.08880*. [Online]. Available: http://arxiv.org/abs/1604.08880
[20] Y. Guan and T. Plötz, "Ensembles of deep LSTM learners for activity recognition using wearables," *ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 2, p. 11, Mar. 2017.
[21] A. Murad and J.-Y. Pyun, "Deep recurrent neural networks for human activity recognition," *Sensors*, vol. 17, no. 11, p. 2556, Nov. 2017.
[22] Y. Zhao, R. Yang, G. Chevalier, X. Xu, and Z. Zhang, "Deep residual bidir-LSTM for human activity recognition using wearable sensors," *Math. Problems Eng.*, vol. 2018, pp. 1–13, Dec. 2018.
[23] L. Wang, "Recognition of human activities using continuous autoencoders with wearable sensors," *Sensors*, vol. 16, no. 2, p. 189, Feb. 2016.
[24] F. 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.
[25] C. A. Ronaoo and S.-B. Cho, "Evaluation of deep convolutional neural network architectures for human activity recognition with smartphone sensors," *Acad. Presentation Paper Book Korean Soc. Inf. Sci.*, pp. 858–860, 2015.
[26] J. Yang, M. N. Nguyen, P. P. San, X. L. Li, and S. Krishnaswamy, "Deep convolutional neural networks on multichannel time series for human activity recognition," in *Proc. 24th Int. Joint Conf. Artif. Intell.*, 2015, pp. 1–7.
[27] A. Ignatov, "Real-time human activity recognition from accelerometer data using convolutional neural networks," *Appl. Soft Comput.*, vol. 62, pp. 915–922, Jan. 2018.
[28] X. Peng, H. Zhu, J. Feng, C. Shen, H. Zhang, and J. T. Zhou, "Deep clustering with sample-assignment invariance prior," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Dec. 31, 2019, doi: 10.1109/TNNLS.2019.2958324.
[29] X. Peng, J. Feng, S. Xiao, W.-Y. Yau, J. T. Zhou, and S. Yang, "Structured AutoEncoders for subspace clustering," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 5076–5086, Oct. 2018.
[30] M. Caron, P. Bojanowski, A. Joulin, and M. Douze, "Deep clustering for unsupervised learning of visual features," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 132–149.
[31] S. Gidaris, P. Singh, and N. Komodakis, "Unsupervised representation learning by predicting image rotations," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–16. [Online]. Available: https://openreview.net/forum?id=S1v4N2l0-
[32] Z. S. Abdallah, M. M. Gaber, B. Srinivasan, and S. Krishnaswamy, "Activity recognition with evolving data streams: A review," *ACM Comput. Surv.*, vol. 51, no. 4, p. 71, Jul. 2018.
[33] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," 2014, *arXiv:1406.1078*. [Online]. Available: http://arxiv.org/abs/1406.1078
[34] L. Ding, L. Li, J. Han, Y. Fan, and D. Hu, "Detecting domain generation algorithms with Bi-LSTM," *Comput., Mater. Continua*, vol. 61, no. 3, pp. 1258–1303, 2019.
[35] Y. Shen, Y. Li, J. Sun, W. Ding, X. Shi, L. Zhang, X. Shen, and J. He, "Hashtag recommendation using LSTM networks with self-attention," *Comput., Mater. Continua*, vol. 61, no. 3, pp. 1261–1269, 2019.
[36] S. G. Hu, Y. Liu, T. P. Chen, Z. Liu, Q. Yu, L. J. Deng, Y. Yin, and S. Hosaka, "Emulating the ebbinghaus forgetting curve of the human brain with a NiO-based memristor," *Appl. Phys. Lett.*, vol. 103, no. 13, Sep. 2013, Art. no. 133701.
[37] S. Bhattacharya and N. D. Lane, "From smart to deep: Robust activity recognition on smartwatches using deep learning," in *Proc. IEEE Int. Conf. Pervas. Comput. Commun. Workshops (PerCom Workshops)*, Mar. 2016, pp. 1–6.
[38] Y. Chen and Y. Xue, "A deep learning approach to human activity recognition based on single accelerometer," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2015, pp. 1488–1492.
[39] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "A public domain dataset for human activity recognition using smartphones," in *Proc. ESANN*, Bruges, Belgium, 2013, pp. 437–442,
[40] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," *ACM SIGKDD Explor. Newslett.*, vol. 12, no. 2, pp. 74–82, Mar. 2011.
[41] M. MacKay, P. Vicol, J. Ba, and R. B. Grosse, "Reversible recurrent neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 9029–9040.
[42] H. Salehinejad, S. Sankar, J. Barfett, E. Colak, and S. Valaee, "Recent advances in recurrent neural networks," 2017, *arXiv:1801.01078*. [Online]. Available: http://arxiv.org/abs/1801.01078
[43] M. Munoz-Organero, "Outlier detection in wearable sensor data for human activity recognition (HAR) based on DRNNs," *IEEE Access*, vol. 7, pp. 74422–74436, 2019.
[44] R. A. Viswambaran, G. Chen, B. Xue, and M. Nekooei, "Evolutionary design of recurrent neural network architecture for human activity recognition," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Jun. 2019, pp. 554–561.

[45] K. Chen, L. Yao, X. Wang, D. Zhang, T. Gu, Z. Yu, and Z. Yang, ''Interpretable parallel recurrent neural networks with convolutional attentions for multi-modality activity modeling,'' in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–8.
[46] J. Okai, S. Paraschiakos, M. Beekman, A. Knobbe, and C. R. de Sa, ''Building robust models for human activity recognition from raw accelerometers data using gated recurrent units and long short term memory neural networks,'' in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 2486–2491.
[47] J. Zhu, H. Chen, and W. Ye, ''A hybrid CNN–LSTM network for the classification of human activities based on micro-Doppler radar,'' *IEEE Access*, vol. 8, pp. 24713–24720, 2020.
[48] J. Y.-H. Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, and G. Toderici, ''Beyond short snippets: Deep networks for video classification,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 4694–4702.
[49] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu, ''Deeply-supervised nets,'' in *Artificial Intelligence and Statistics*. 2015, pp. 562–570.
[50] E. Choi, M. T. Bahadori, A. Schuetz, W. F. Stewart, and J. Sun, ''Doctor AI: Predicting clinical events via recurrent neural networks,'' in *Proc. Mach. Learn. Healthcare Conf.*, 2016, pp. 301–318.
[51] J. Xu, R. Rahmatizadeh, L. Bölöni, and D. Turgut, ''Real-time prediction of taxi demand using recurrent neural networks,'' *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 8, pp. 2572–2581, Aug. 2018.
[52] A. N. Jagannatha and H. Yu, ''Bidirectional RNN for medical event detection in electronic health records,'' in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, North American Chapter Meeting*, 2016, p. 473.
[53] T. Young, D. Hazarika, S. Poria, and E. Cambria, ''Recent trends in deep learning based natural language processing,'' *IEEE Comput. Intell. Mag.*, vol. 13, no. 3, pp. 55–75, Aug. 2018.
[54] Z. Che, D. Kale, W. Li, M. T. Bahadori, and Y. Liu, ''Deep computational phenotyping,'' in *Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2015, pp. 507–516.
[55] Z. C. Lipton, D. C. Kale, C. Elkan, and R. C. Wetzel, ''Learning to diagnose with LSTM recurrent neural networks,'' in *Proc. 4th Int. Conf. Learn. Represent. (ICLR)*, San Juan, Puerto Rico, May 2016, pp. 1–18. [Online]. Available: http://arxiv.org/abs/1511.03677
[56] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, ''Focal loss for dense object detection,'' in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.
[57] H.-Y. Chen, P.-H. Wang, C.-H. Liu, S.-C. Chang, J.-Y. Pan, Y.-T. Chen, W. Wei, and D.-C. Juan, ''Complement objective training,'' in *Proc. Int. Conf. Learn. Represent.*, 2019, pp. 1–11. [Online]. Available: https://openreview.net/forum?id=HyM7AiA5YX
[58] Z. Zhang and M. Sabuncu, ''Generalized cross entropy loss for training deep neural networks with noisy labels,'' in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 8778–8788.
[59] X. Luo, X. Chang, and X. Ban, ''Regression and classification using extreme learning machine based on L1-norm and L2-norm,'' *Neurocomputing*, vol. 174, pp. 179–186, Jan. 2016.
[60] Y. Wang, D. Li, Y. Du, and Z. Pan, ''Anomaly detection in traffic using L1-norm minimization extreme learning machine,'' *Neurocomputing*, vol. 149, pp. 415–425, Feb. 2015.
[61] H. Lin, Y. Lu, X. Han, and L. Sun, ''Adaptive scaling for sparse detection in information extraction,'' in *Proc. 56th Annu. Meeting Assoc. Comput. Linguistics*, Melbourne, VIC, Australia, 2018, pp. 1033–1043.

YUE HU received the B.S. degree in electronic information engineering and the M.S. degree in computer technology from Zhengzhou University, China, in 2016 and 2019, respectively. His research interests include blochchain, information security, and mechanic learning.

XIAO-QING ZHANG received the B.S. degree in water conservancy and hydropower engineering from South China Agricultural University, China, in 2016, and the M.S. degree in computer technology from Zhengzhou University, China, in 2019. His research interests include machine learning, deep learning, data mining, and medical images.



LI XU is currently pursuing the degree with Zhengzhou University. His research interests include blochchain and information security.

FENG XIAN HE received the Ph.D. degree. He is currently a Professor Senior Engineer and the Chief Engineer of the Department of Civil Engineering Mechanics Yellow River Institute of Hydraulic Research (YRIHR), the Deputy Chairman of the Youth Committee of the Henan Hydropower Engineering Society, and a master's Supervisor with Hohai University.

ZHAO TIAN received the B.S. degree in information and computing science from Huazhong Agricultural University, in 2008, the M.S. degree in computer software and theory from Zhengzhou University, in 2011, and the Ph.D. degree in safety technology and engineering from Beijing Jiaotong University, in 2016. He is currently a Lecturer with the School of Software, Zhengzhou University. His research interests include information security, artificial intelligence, and intelligent transportation.

WEI SHE (Member, IEEE) received the B.S. degree in control engineering from the Air Defence Academy of PLA, in 2000, the M.S. degree in software engineering from Hunan University, in 2008, and the Ph.D. degree in computer software and theory from Zhengzhou University, China, in 2013. He is currently an Associate Professor with the College of Software, Zhengzhou University. His research interests include information security, the energy Internet, and the Internet healthcare.

WEI LIU received the B.S. and M.S. degrees in information engineering from Zhengzhou University, China, in 2003 and 2008, respectively, and the Ph.D. degree from Tohoku University, Japan, in 2013. He is currently an Associate Professor with the College of Software, Zhengzhou University. His research interests include information security, wireless mesh networks, and the Internet healthcare.