Comparative Study of Machine Learning and Deep Learning Architecture for Human Activity Recognition Using Accelerometer Data

Sarbagya Ratna Shakya, Chaoyang Zhang, and Zhaoxian Zhou

***Abstract*—Human activity recognition (HAR) has been a popular fields of research in recent times. Many approaches have been implemented in literature with the aim of recognizing and analyzing human activity. Classical machine learning approaches use hand-crafted feature extraction and are based on classification technique, however of late, deep learning approaches have shown greater success in recognition accuracy with increased performance. With the current, wide popularity of mobile phones and various sensors such as accelerometers, gyroscopes, and cameras that are already installed on mobile phones, the activity recognition using the accumulating data from mobile phones has been a significant area of research in HAR. In this paper, we investigate the HAR based on the data collected through the accelerometer sensor of mobile devices. We employ different machine learning (ML) classifiers, algorithms, and deep learning (DL) models across different benchmark datasets. The experimental results from this study provide a comparative performance analysis based on accuracy, performance, and the costs of different ML algorithms and DL algorithms, based on recurrent neural network (RNN) and convolutional neural network (CNN) models for activity recognition.**

***Index Terms*—ML, DL, CNN, RNN.**

I. Introduction

Due to wide applications in many fields such as human-computer interaction, elder care [1], medicine [2], video surveillance and security applications, HAR has become one of the most active research areas in the field of computers in recent years. The ability of a computer to recognize human activity can be used to detect possible security threats covering large areas through video surveillance. The activities of elderly people can be monitored, and with the proper feedback system, sudden adverse effects such as sudden health problems, can be recognized early and properly mitigated before they become worse. Also, HAR can be used for patients having a mental disorder or disease such as Parkinson's Disease Dementia [3] to regularly monitor their activities and detect if any abnormal activities occur.

Basically, HAR can be divided into three basic types: video-based HAR [4], sensor-based[5], and radio-based[1] HAR method. In video-based HAR, a sequence of images or videos captured from the camera is used to analyze the activity performed by the human. In sensor-based HAR, sensors such as accelerometers, gyroscopes that are placed on various parts of the body are used to collect the data generated by the movement of the human. In the radio-based method, the human body moves in a radio field and the channel fading due to signal attenuation and multipath effect is used for HAR.

Among many approaches, ML methods [6] and DL approaches [7], [8] can be used to recognize the daily activity of the human by using sensor data with high accuracy. By using 3D acceleration data from the accelerometer sensors that are placed on multiple parts of the body, movement is detected, and activity is recognized. The various sensors placed on the human body have different sensitivities toward different activities [9]. The presence of sensors all over their bodies while performing their regular activities can be uncomfortable to the users; therefore, the use of smartphones in HAR is gaining popularity of late.

In this paper, we compare the HAR of different algorithms, including ML classifiers such as random forest (RF), decision tree (DT), K-nearest neighbor (KNN), and DL classifiers such as RNN and CNN to accelerometer data collected from single as well as multiple devices.

The remainder of this paper is organized as follows: Section II provides a literature review of the related work on sensor-based HAR using deep and ML algorithms. Sections III and IV describe the datasets and methods that have been used for the study. Section V presents the experimental results and analysis and in Section VI, conclusions and future work are discussed.

II. Related Work

In this section, we mainly review the sensor-based recognition process using accelerometer data. Because of its low cost, ability to measure the acceleration in three orthogonal axes, and low power consumption, the use of accelerometers is increasing in HAR. Researchers[5] have used multiple inertial sensors at different body parts such as the chest, right thigh, and left ankle. They have highlighted the performance analyses for different supervised and unsupervised classification techniques. The selection of the sensors, the placement position, and the number of sensors to use are still major issues [9] in HAR. The research in [10] has investigated the position of the accelerometer sensor by collecting data from six tri-axial accelerometers that were placed at different parts of the body such as the chest, wrist, lower back, hip, thigh, and foot. The study includes six

Manuscript received September 10, 2018; revised October 25, 2018.

S. R. Shakya, C. Zhang, Z. Zhou are with the School of Computing Sciences and Computer Engineering, University of Southern Mississippi, Hattiesburg, MS 39406 USA (email: sarbagya.shakya@usm.edu, Chaoyang.zhang@usm.edu, zhaoxian.zhou@usm.edu).

doi: 10.18178/ijmlc.2018.8.6.748

activities: walking, running on a motorized treadmill, sitting, lying, standing, and walking up and downstairs. The study has shown that putting sensors on the hip provides better accuracy than at other locations of the body. Not only the position of the sensors but the determination of the number of sensors to be used, have drawn attention toward HAR. In [11] the authors have shown that gyroscopes can have higher accuracy than accelerometer data and the choice of sensors depends on the activities performed. They also suggested using multiple sensors in HAR.

The most used classification approaches for activity recognition using the accelerometer sensor have been the ML method [12], [13] and more recently, DL methods [7], [8]. In [14], the authors used phone-based accelerometer data to identify the physical activity in daily life collected from 29 users. Using three ML classifiers, they achieved high accuracies for common activities such as walking and jogging, but relatively low accuracies on activity such as climbing stairs. In [12], the authors utilized the ensemble of the different ML classifiers with the average of probabilities combination rule and have achieved better results than those from other single state-of-art classifiers.

In recent years, with the advancement and achievements of deep learning in many fields such as speech recognition and natural language processing, DL approaches have been implemented in HAR. Hybrid approaches to DL and hidden Markov models [15] have been implemented, and have achieved better recognition accuracy by using tri-axial accelerometers than have nonhybrid methods. In [16], DL methods based on restricted Boltzmann machines have been used and have outperformed a wide range of common models and have also shown the acceptable use of resource consumption, although in constrained devices such as smartwatches. A single DL network as well as an ensemble of DL networks such as long short-term memory(LSTM) networks has been implemented to analyze the recognizing capabilities for real-world unbalanced data [17]. While most conventional methods use hand-crafted feature extraction, DL performs automatic high-level feature extraction [7] and this has made it a widely adopted approach in recent times.

III. Dataset

In our experiment, we selected two commonly used benchmark datasets for HAR using the accelerometer data collected from smartphones. The first dataset Wireless Sensor Data Mining (WISDM) dataset contains the raw data collected from the smartphone carried on the waist. The second dataset consists of the data collected from five different accelerometers on five different positions on the body. Both datasets consist of daily motion activities such as walking, sitting, standing, going upstairs, and going downstairs. We used these datasets to demonstrate our study in different nature of datasets, the balanced and unbalanced dataset, having similar activities. The detailed description of each dataset is presented.

A. ACTi Tracker Dataset (WISDM Dataset)

ACTi tracker Dataset [18] (Dataset 1) contains the data collected in controlled laboratory settings from the real-world usage of smartphones released by the WISDM Lab in 2013. The dataset contains the raw data collected from the accelerometers of the cellphones attached to the waists of the volunteers. The dataset contains 2,980,765 pieces of labeled data with six attributes: walking (42.1%), jogging (14.7%), stairs (1.9%), sitting (22.3%), standing (9.7%), and lying down (9.3%). The data were collected at a sampling rate of 20Hz (one sample every 50 ms).

B. Sensor Activity Recognition Dataset (Shoaib SA)

In this dataset [19] (Dataset 2), the data were collected in the university building for seven physical activities: walking, sitting, standing, jogging, biking, walking upstairs, and walking downstairs. The activities were performed for 3-4 minutes by ten male participants between 25 and 30 years of age. In the experiment, five smartphones (Samsung Galaxy SII i9100) were placed on five different positions on the volunteers. The smartphones were placed on the common positions of mobile devices in daily life such as right jeans' pocket, left jeans' pocket, upper arm, right wrist, and on the belt using a belt clipper. The dataset contains data collected from the accelerometers, linear accelerometers, gyroscopes, and magnetometer sensors of the mobile phones at the rate of 50 samples per second. In our experiment, we used only the data collected from the accelerometers for the five different smartphone positions. The dataset is balanced and has an equal number of observations for each class.

IV. Methodology

The raw data of the dataset that have been collected from the three-dimensional axis values of the accelerometer sensors are normalized and then analyzed through the ML classifiers and DL algorithms. The output from the model is analyzed based on overall accuracy, precision, and recall values. Since we have only used the raw data directly from the sensors, the hand-crafted feature extraction and feature selection methods which are most common in other research are not implemented for ML in our models. The overall methodology is shown in Fig. 1.

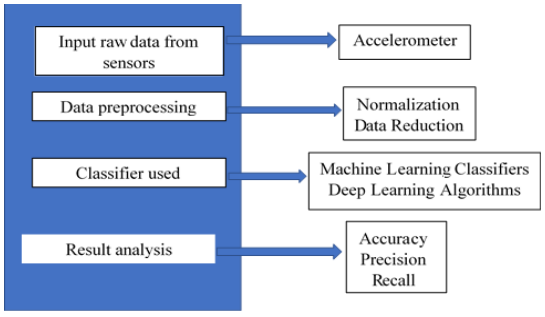


Fig. 1. Overall methodology.

A. ML Algorithm

We used raw data that were collected from the accelerometers in the smartphones. The data comprised 80% of training data and 20% testing data. The different ML classifiers were modeled using the training data, and performance was evaluated using the testing data. The experiment was carried out using various ML algorithms with

different parameters. The classifiers that showed the best results were RF with the 10 estimators, DT, and KNN with 5 neighbors. In this paper, the results and comparison between these three classifiers are presented.

B. DL Algorithm

In our study, we implemented DL architecture and investigated the recognition performances in the CNN and RNN architectures. We have implemented a CNN method that has already shown remarkable success in image classification [20] for the time-series data in HAR. The 1D sensor data are transformed to 2D data and reshaped to the 4D tensor format for 2D convolution, in which the height of the input data shape is pre-defined as 1. For the RNN method, we implemented the model based on LSTM cells. We have used the negative log-likelihood cost function using Adam optimizers.

1) CNN method

TABLE I: SUMMARY OF THE CNN MODEL LAYERS, OUTPUT SHAPE AND NUMBER OF PARAMETERS

layers	Output shape	# parameters
conv2d_1 (Conv2D)	(None, 1, 100, 16)	64
max_pooling2d_1	(None, 1, 50, 16)	0
conv2d_2 (Conv2D)	(None, 1, 50, 64)	1088
max_pooling2d_2	(None, 1, 25, 64)	0
dropout_1 (Dropout)	(None, 1, 25, 64)	0
conv2d_3 (Conv2D)	(None, 1, 14, 256)	196864
max_pooling2d_3	(None, 1, 7, 256)	0
conv2d_4 (Conv2D)	(None, 1, 7, 512)	131584
max_pooling2d_4	(None, 1, 4, 512)	0
dropout_2 (Dropout)	(None, 1, 4, 512)	0
global_average_pooling	(None, 512)	0
dense_1 (Dense)	(None, 50)	25650
dense_2 (Dense)	(None, #of class)	306

In the CNN model, the data from the accelerometer are first divided into time-series segments which are the same size as the window. The size of the window we selected was 100. A 50% overlapping window technique, in which the data is divided into several segments, was used. A batch of segments, each segment sized 1×100, was stored in a 4D tensor similar to an image data. Each segment had half of the overlapping data from its previous segment. The depth axis comprises the three-dimensional sensor values of the data. These segments are divided as 80% training data and 20% testing data. These training data are divided again into 80% training and 20% validation data. With these training data, the deep model is trained with a learning rate of 0.001. The trained model was tested with the validation data in each epoch. A checkpoint was created in which the model was saved if the performance improved with each epoch in the validation loss. At the end of the training, we had the latest optimized model that shows high performance in the validation dataset, which is then used for testing against the testing data. This method implements similar logic as the early stopping technique. The data are fed into different layers, where each layer has a batch size of 100. As shown in Table I, the CNN model contains a total of four convolutional layers: a max pool layer which reduces the data sample by half, a global average pooling layer, a dropout layer for regularization with a probability constant of 0.5, and a dense layer. We used the ReLU activation function in our model to transform our output data. Also, to validate our results, we performed the five-fold cross-validation method on the entire dataset. The results from both approaches were similar.

2) RNN model

In the RNN model we used the multiple layers of LSTM and the dropout wrapper layer, which were then appended together. The output of the first layer would be the input to the next layer. Because LSTM provides the possibility of learning the long dependencies of the data, we used it in our model. All these layers are stacked by using multiple RNN cells. First, we constructed a fixed length of training batches to train the model. The number of batch size was 25 with a learning rate of 0.3 when the total number of epochs equaled 100. After each epoch, we calculated the mean average and mean loss in each epoch.

V. EXPERIMENT RESULTS

We selected two publicly available datasets for activity recognition. We chose two different nature datasets; one dataset had an unbalanced data distribution, collected using a single sensor placed at the waist region, whereas the other was a balanced dataset collected using multiple sensors on various parts of the body. Both sets had data collected from the tri-axial accelerometer sensor.

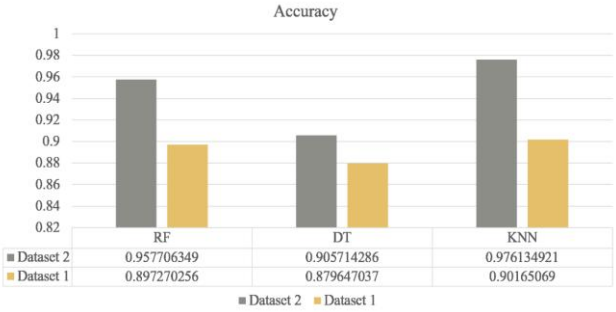


Fig. 2. Overall accuracy.

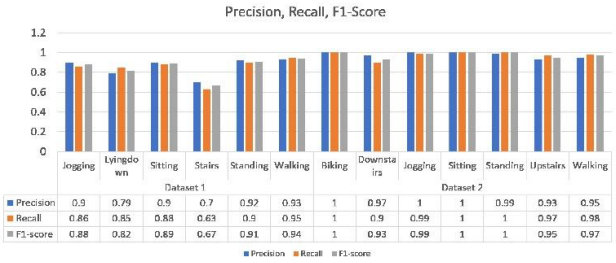


Fig. 3. Precision, Recall, and F1 score for KNN.

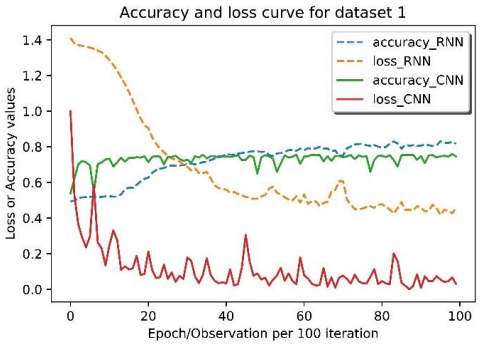


Fig. 4. Accuracy/Loss result for dataset 1.

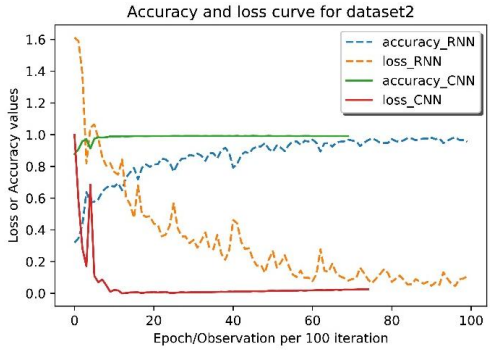


Fig. 5. Accuracy/Loss result for dataset 2.

Fig. 2 shows an overall accuracy comparison of two datasets for three ML classifiers. Comparing both datasets for the same ML classifier reveals that the accuracy is higher for dataset 2 than for dataset 1. This can occur because the data was collected from multiple sensors placed on various parts of the body while the activity was being performed. This shows that if the model is fed with more information during training and designing, overall performance can be affected. Also, the accuracy is higher for the KNN method in both datasets, nearly 91% in dataset 1 and about 98% in dataset 2 than for other ML classifiers. Since dataset 1 is a highly unbalance dataset with a greater variation in the number of data for different activities, the result is further analyzed using other performance metrics such as precision, recall and F1 score. Fig. 3 shows the precision, recall, and F1-score of both the datasets for KNN classifier.

In dataset 1, with the exception for climbing stairs, the performance has high precision, recall, and F1-score, whereas the score is higher and nearly consistent across all the activities in dataset 2. This can be due to the highly imbalanced nature of the dataset 1 where the percentage of data for stairs is about 1.9%, whereas for walking it is about 42.1%. We have used only the raw data without any hand-crafted features extractions for ML classifiers.

Fig. 4 and Fig. 5 show the accuracy and loss curve of both, the RNN and CNN models for the training data for dataset 1 and dataset 2, respectively. For both the datasets, the CNN and RNN models are trained using the training data with the same model hyperparameters in the same environment. The output shown in the figures are the best results, in terms of accuracy, among the different outputs obtained by tuning the distinct set of hyperparameters in both models. The figures show that for the RNN model, the mean accuracy of the total batch for each epoch increases and is 81.74% and 95.65% for dataset 1 and dataset 2 respectively. For the CNN model, the accuracy of testing data is as high as 92.22% in dataset 1 and 99.12% in dataset 2. The mean loss decreases continuously for each epoch or iteration indicating the learning nature of the model. The accuracy table for the datasets and DL models is given in Table II.

TABLE II: ACCURACY TABLE FOR DATASET 1 AND DATASET 2 USING THE RNN AND CNN MODEL

Method	Dataset 1	Dataset 2
RNN	81.74%	95.65%
CNN	92.22%	99.12%

From Table II, it can be observed that using DL architecture such as CNN and RNN for the balanced dataset the result shows a higher improvement than that in an unbalanced dataset with the same model structure and the same set of hyperparameters. Also, the CNN model provides a better prediction than the RNN model for time-series sensor data with the given set of parameters.

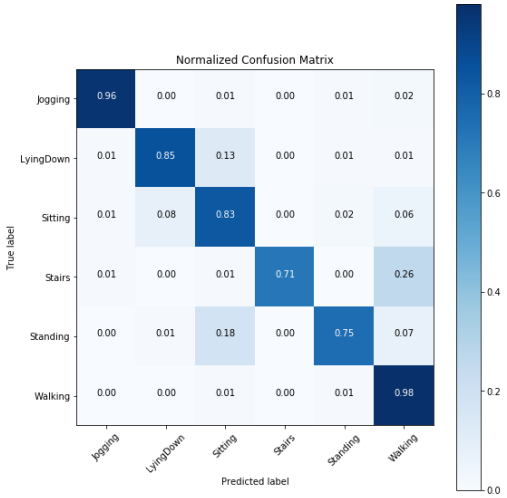


Fig. 6. Normalize confusion matrix for predicted output in testing data of actitracker dataset.

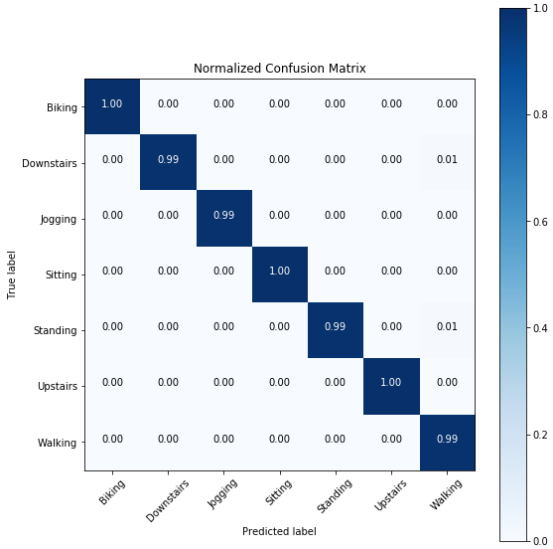


Fig. 7. Normalize confusion matrix for predicted output in testing data of Shoaib dataset.

Fig. 6 and Fig. 7 show the normalized confusion matrix for predicted output in the testing data for dataset 1 and 2 for the CNN model. In Fig. 6, one can see that activities such as jogging and walking are classified more accurately than activities such as climbing stairs which are misclassified as walking. Activities of a similar nature are misclassified, e.g., climbing stairs is misclassified as walking, and standing is misclassified as sitting. This could be a result of variation in the distribution of data for all classes. This can also be illustrated in Fig. 7 where the prediction accuracy is distributed across all the classes. Because of the uniform nature of data distribution among all classes, and because of a balanced nature, similar activities could be classified more accurately.

Table III displays the precision and recall values for different activities using dataset 1 and 2. Comparing

precision and recall for both datasets for similar activities reveals that dataset 2 provides a better result than dataset 1, which also shows that balanced nature of dataset and using multiple sensors can boost activity recognition performance.

TABLE III: PRECISION AND RECALL TABLE FOR BOTH THE DATASET USING CNN METHOD

Dataset 1			Dataset 2		
Activity	Precision	Recall	Activity	Precision	Recall
Jogging	0.97	0.96	Biking	1	1
Lying	0.80	0.85	Downstairs	0.99	0.99
Down					
Sitting	0.84	0.83	Jogging	0.99	0.99
Stairs	0.97	0.71	Sitting	1	1
Standing	0.89	0.75	Standing	1	0.99
Walking	0.93	0.98	Upstairs	1	1
			Walking	0.98	0.99
Avg/total	0.90	0.90		1	1

TABLE IV: ACCURACY TABLE USING K-FOLD CROSS VALIDATION

	KNN		*CNN*	
Fold	*Dataset 1*	*Dataset 2*	*Dataset 1*	*Dataset 2*
1	90.15%	97.62%	84.37%	99.52%
2	90.23 %	97.64%	88.26%	99.52%
3	90.19%	97.68%	86.39%	98.41%
4	90.16%	97.67%	89.11%	99.36%
5	90.24%	97.63%	87.39%	98.96%
Average	90.19%	97.65%	87.31%	99.16%

To further validate our result, we analyzed the performance in the entire dataset using five-fold cross validation. Table IV shows the comparison of the accuracy in each fold of the five-fold cross validation and the average accuracy of dataset 1 and 2 for the KNN and CNN methods. The result is comparable with the result from the previous approaches.

VI. CONCLUSION

We compared HAR performance in two datasets, WISDM and Shoaib, for determining daily activities such as cycling, walking, and standing. On comparing the results of the ML classifiers, we found that we could achieve comparable results with most of the classifiers using only the raw data from the accelerometer sensors. This has reduced the effort for determining the hand-crafted features from raw data for ML classifiers. KNN provides the best results in both dataset with a high performance compared with other classical ML classifiers such as RF and DT. We also compared the results for the RNN and CNN methods in both datasets. Our model shows learning capacity and produces significant recognition accuracy and better results in the balanced dataset than if the same model were used in an unbalanced dataset. The DL recognition models can have a higher accuracy with the data collected from multiple accelerometer sensors placed at different parts of the body than with single accelerometer sensor data. Additionally, DL architecture such as CNN exhibits a higher performance than traditional ML classifiers with no hand-crafted feature selection. In our study, we have used only the accelerometer data collected from the mobile devices whereas data from other sensors can be used simultaneously. Furthermore, sampling methods can be implemented in an unbalanced dataset to study the effect on the performance result of HAR.

REFERENCES

[1] S. Wang and G. Zhou, "A review on radio based activity recognition," *Digit. Commun. Networks*, vol. 1, no. 1, pp. 20–29, 2015.

[2] H. Gjoreski, J. Bizjak, M. Gjoreski, and M. Gams, "Comparing deep a classical machine learning methods for human activity recognition using wrist accelerometer," in *Proc. the 25th International Conference on Artificial Intelligence*, 2016, pp. 1–7.

[3] M. Bachlin *et al.*, "Wearable Assistant for Parkinson's Disease Patients With the Freezing of Gait Symptom," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 2, pp. 436–446, Mar. 2010.

[4] S.-R. Ke, H. L. U. Thuc, Y.-J. Lee, J.-N. Hwang, J.-H. Yoo, and K.-H. Choi, "A review on video-based human activity recognition," *Computers*, vol. 2, no. 2, pp. 88–131, 2013.

[5] F. Attal, S. Mohammed, M. Dedabrishvili, F. Chamroukhi, L. Oukhellou, and Y. Amirat, "Physical human activity recognition using wearable sensors," *Sensors*, vol. 15, no. 12, pp. 31314–31338, 2015.

[6] A. Mannini and A. M. Sabatini, "Machine learning methods for classifying human physical activity from onbody accelerometers," *Sensors*, vol. 10, pp. 1154–1175, 2010.

[7] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: A survey," *CoRR*, vol. abs/1707.0, 2017.

[8] N. Y. Hammerla, S. Halloran, and T. Plotz, "Deep, convolutional, and recurrent models for human activity recognition using wearables," in *Proc. the 25th International Conference on Artificial Intelligence*, 2016, pp. 1533–1540.

[9] H. Yu, S. Cang, and Y. Wang, "A review of sensor selection, sensor devices and sensor deployment for wearable sensor-based human activity recognition systems," in *Proc. 2016 10th International Conference on Software, Knowledge, Information Management Applications (SKIMA)*, 2016, pp. 250–257.

[10] I. Cleland *et al.*, "Optimal Placement of Accelerometers for the Detection of Everyday Activities," *Sensors*, vol. 13, no. 7, pp. 9183–9200, 2013.

[11] Y. Chen and C. Shen, "Performance analysis of smartphone-sensor behavior for human activity recognition," *IEEE Access*, vol. 5, pp. 3095–3110, 2017.

[12] C. Catal, S. Tufekci, E. Pirmit, and G. Kocabag, "On the use of ensemble of classifiers for accelerometer-based activity recognition," *Appl. Soft Comput.*, vol. 37, no. Supplement C, pp. 1018–1022, 2015.

[13] A. Mannini and A. M. Sabatini, "Machine learning methods for classifying human activity from onbody accelerometers," *Sensors*, vol. 10, pp. 1154–1175, 2010.

[14] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," *ACM SIGKDD Explor. Newsl.*, vol. 12, pp. 74–82, 2011.

[15] M. Abu Alsheikh, A. Selim, D. Niyato, L. Doyle, S. Lin, and H.-P. Tan, "Deep activity recognition models with triaxial accelerometers," *ArXiv e-prints*, Nov. 2015.

[16] Z. Liu, M. Wu, K. Zhu, and L. Zhang, "SenSafe: A smartphone-based traffic safety framework by sensing vehicle and pedestrian behaviors," *Mob. Inf. Syst.*, vol. 2016, Sep. 2016.

[17] Y. Guan and T. Ploetz, "Ensembles of deep lstm learners for activity recognition using wearables," *CoRR*, vol. abs/1703.0, 2017.

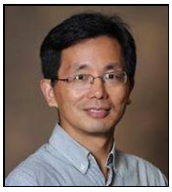
[18] J. W. Lockhart, G. M. Weiss, J. C. Xue, S. T. Gallagher, A. B. Grosner, and T. T. Pulickal, "Design considerations for the WISDM smart phone-based sensor mining architecture," in *Proc. the Fifth International Workshop on Knowledge Discovery from Sensor Data*, 2011, pp. 25–33.

[19] M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, and P. J. M. Havinga, "Fusion of smartphone motion sensors for physical activity recognition," *Sensors*, vol. 14, no. 6, pp. 10146–10176, 2014.

[20] P. Wang, W. Li, Z. Gao, J. Zhang, C. Tang, and P. Ogunbona, "Deep convolutional neural networks for action recognition using depth map sequences," *CoRR*, vol. abs/1501.0, 2015.

Sarbagya Ratna Shakya received the B. Eng. in electronics engineering from National College of Engineering, Tribhuvan University of Nepal in 2009; M. Eng. in computer engineering from Nepal College of Information Technology, Pokhara University of Nepal in 2014. Currently he is a PhD student in School of Computing, University of Southern Mississippi

(USM) since 2015. His current research interests include machine learning, deep learning and high-performance computing.



Chaoyang Zhang received his MS degree in computer science and PhD degree in computational analysis and modeling from Louisiana Tech University in 2001. He is currently a Professor of Computer Science in the School of Computing at the University of Southern Mississippi. He has published more than seventy papers in academic journals and conference proceedings. His research interests include data mining, machine learning, bioinformatics, image processing and high-performance computing.

Zhaoxian Zhou received the B. Eng. from the University of Science and Technology of China in 1991; M. Eng. from the National University of Singapore in 1999 and the PhD degree from the University of New Mexico in 2005. All His degrees are in electrical engineering. From 1991 to 1997, he was an electrical engineer in China Research Institute of Radio wave Propagation. He joined the University of Southern Mississippi in 2005. He has published more than fifty papers in academic journals and conference proceedings. His current research interests include computational science and electrical engineering.