An Investigation on Deep Learning-Based Activity Recognition Using IMUs and Stretch Sensors

1st Nguyen Thi Hoai Thu
School of Electronic and Electrical Engineering
Kyungpook National University
Daegu, Republic of Korea
thunguyen@knu.ac.kr

2nd Dong Seog Han
School of Electronic and Electrical Engineering
Kyungpook National University
Daegu, Republic of Korea
dshan@knu.ac.kr

***Abstract*—With the advancement and ubiquitousness of wearable devices, wearable sensor-based human activity recognition (HAR) has become a prominent research area in the healthcare domain and human-computer interaction. Inertial measurement unit (IMU) which can provide a wide range of information such as acceleration, angular velocity has become one of the most commonly used sensors in HAR. Recently, with the growing demand for soft and flexible wearable devices, mountable stretch sensors have become a new promising modality in wearable sensor-based HAR. In this paper, we propose a deep learning-based multi-modality HAR framework which consists of three IMUs and two fabric stretch sensors in order to evaluate the potential of stretch sensors independently and in combination with IMU sensors for the activity recognition task. Three different deep learning algorithms: long short-term memory (LSTM), convolutional neural network (CNN) and hybrid CNN-LSTM are deployed to the sensor data for automatically extracting deep features and performing activity classification. The impact of sensor type on recognition accuracy of different activities is also examined in this study. A dataset collected from the proposed framework, namely iSPL IMU-Stretch and a public dataset called w-HAR are used for experiments and performance evaluation.**

***Index Terms*—activity recognition, IMUs, stretch sensors, wearable sensors, deep learning**

I. Introduction

Wearable sensor-based human activity recognition has become an emerging research topic that uses body-worn motion sensors to understand human behaviors, detecting abnormal activities (e.g., fall, gait disorder), hence it helps to encourage people to have healthier lifestyles and provides quick response to emergency situations. Recent advances in micro electro-mechanical system (MEMS) have enabled these sensors to be shrunken to extremely small sizes and widely embedded into smart wearable devices such as smartphones and smartwatches. Therefore, in view of ubiquitousness and user privacy preserving, this wearable sensor-based approach is preferred in healthcare applications over other approaches that use radio frequency signals or vision sensors.

One of the most commonly used wearable sensors in HAR is the inertial measurement unit (IMU) which can contain a variety of sensors such as motion sensor (accelerator), rotation sensor (gyroscope) and magnetometer. Bächlin *et al.* [1] use two sensors, which provide 3-D acceleration and attach them to the Parkinson disease patients' leg (*i.e.*, shank and thigh) for detecting freezing of gait events. Authors in [2] and [3] use data collected from accelerator and gyroscope embedded in smartphones to recognize a wide range of daily living activities and detect fall events. Although a growing body of literature has successfully applied the inertial sensors to HAR, these body-worn sensors are made of inflexible and rigid materials that can be obtrusive, interfere with the user's natural movements, and lead to discomfort wearing experience [4]. This limitation has urged researchers to take further action on developing the next generation of wearable sensors.

In recent years, there has been considerable interest in designing soft, textile wearable sensors and they have shown a high potential in a variety of applications from the healthcare domain to human-computer interaction. Prominent among these sensors is the fabric stretch sensor also known as strain sensor. The sensor can be stitched to the clothes or directly attached to human skin to measure a wide range of motion from articulation bending-straightening to respiration and heartbeat. Chander *et al.* [5] have used a stretchable soft robotic sensor to detect the kinematics of the ankle joints during slip and trip perturbations. Bhat *et al.* [6] share a dataset that contains data collected from a wearable accelerometer and stretch sensor. Handcrafted feature extraction methods such as discrete wavelet transform and fast Fourier transform are applied to the collected sensor data before a neural network-based classifier is deployed for activity recognition. A full-body sensing system consisting of IMU, knitted piezoresistive fabric strain sensors, EMG electrodes, goniometers and force sensors is proposed by Klaassen *et al.* [7] for monitoring stroke patients in a home environment.

In this paper, a study on the potential of the stretch sensors in combination with IMU sensor in wearable sensor-based human activity recognition task is conducted with the use of deep learning algorithms. An investigation on different sensor combinations is carried out to evaluate the sensor combination effectiveness and impact of sensor position on activity recognition accuracy. The rest of the paper is organized as follows. Section II introduces the overall framework of multi-modality HAR and gives details on the data collection

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2021R1A6A1A03043144).

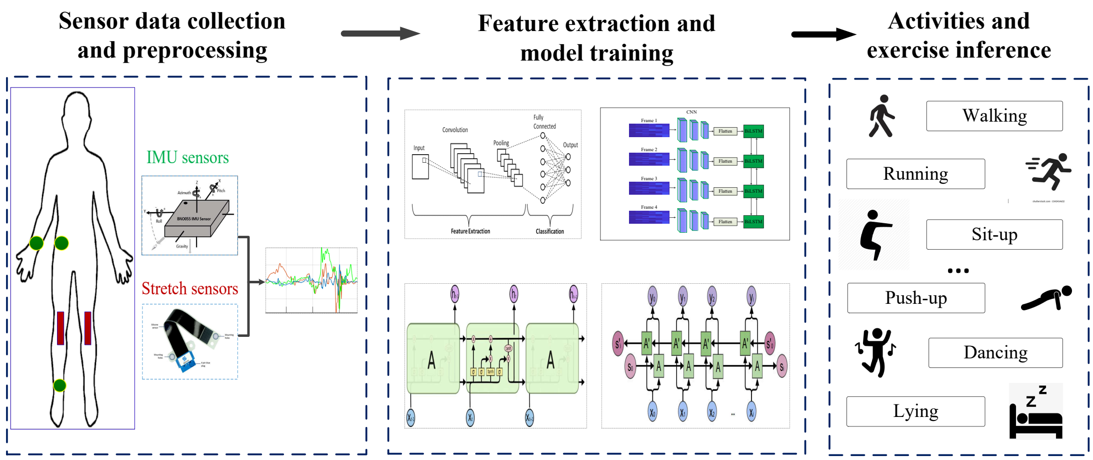


Fig. 1. Overall framework of the deep learning-based human activity recognition system using multi-modal wearable sensors.

method and structure of different deep learning-based HAR models. A set of experiments and results are presented in Section III with some discussion. Finally, our conclusions and some notes on future research are drawn in Section IV.

II. Methodology

The deep learning-based multi-modality HAR framework considered in this study contains three main components: i) data collection and preprocessing, ii) feature extraction and model training, and iii) activities and exercise inference. Firstly, data collected from wearable sensors (*i.e.*, IMUs and stretch sensors) is transmitted to the server and pre-processed here. Secondly, the training dataset which contains preprocessed data with labels is then input into a deep learning model for the training process and automatic feature extraction. Finally, the trained model is used on sensor data in the testing phase or real-life scenarios for inferring human activities. The overall structure of this multi-modality HAR framework is illustrated in Fig. 1.

A. Data Collection and Pre-processing

In this framework, three programmable WiFi 9-axis absolute orientation sensors are used and mounted at the user's wrist, waist and ankle. 3-axial acceleration, angular velocity and linear acceleration information are collected from each sensor, thus, there are 9 attributes at each time step. The fabric stretch sensors used in this study are flexible capacitors. Each stretch sensor is constructed from 5 layers: two outer layers are ground electrodes followed by two dielectric layers and the middle layer is a signal electrode. A coaxial cable is connected to the signal electrode and ground electrodes to collect the capacitance data. The capacitance will increase and decrease when the sensor extends and contracts. Because of this working principle, we stitched the two stretch sensors at the knees of the user garment in order to catch the bending and stretching activities of the user's legs.

Data collected from the IMU sensors is transmitted to the server using the message queuing telemetry transport (MQTT) protocol and the Mosquitto broker [8] while the stretch sensor data is collected by using the Bluetooth low energy (BLE) technology. For a simple sensor data combination and to reduce the complexity, both types of sensors use the same sampling frequency of 25 Hz.

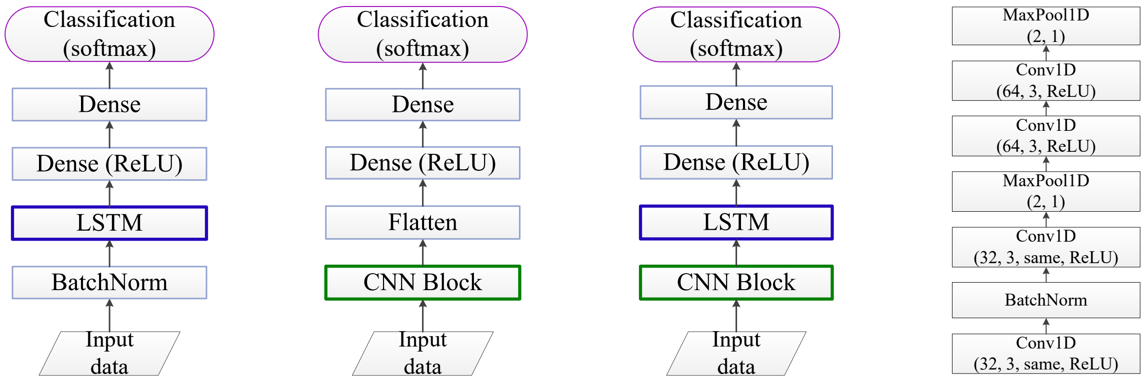
Data loss that happened during the wireless transmission is handled by using a linear interpolation method. Data collected from all the sensors are then synchronized and split into small windows with a fixed size of 2 seconds (50 data samples). An overlap of 50% between every two adjacent windows is applied to avoid missing data and increase the recognition granularity. Single window data of an IMU sensor can be described as

$$\mathbf{W}_{\mathrm{IMU}} = \left(\mathbf{a}_x \ \mathbf{a}_y \ \mathbf{a}_z \ \mathbf{g}_x \ \mathbf{g}_y \ \mathbf{g}_z \ \mathbf{l}_x \ \mathbf{l}_y \ \mathbf{l}_z\right) \in \mathbb{R}^{N \times K} \tag{1}$$

where L is the window length, $\mathbf{a}_x, \mathbf{a}_y, \mathbf{a}_z, \mathbf{g}_x, \mathbf{g}_y, \mathbf{g}_z, \mathbf{l}_x, \mathbf{l}_y$, and $\mathbf{l}_z$ are column vectors contain L data samples of 3-axial acceleration, 3-axial angular velocity and 3-axial linear acceleration, respectively. Because stretch sensor provides only the stretching-shrinking information so its window data can be expressed as a column vector $\mathbf{w}_s \in \mathbb{R}^{L\times 1}$. In this study, L is set to 50. When data of more than one sensor is used, the windows of all sensors are concatenated by joining all the column vectors, and finally a window data $\mathbf{W}_{\mathrm{fusion}} \in \mathbb{R}^{L\times N}$ is obtained, where N is the total number of sensor attributes.

B. Deep Learning Models

With the capability of automatically extracting deep features during training and the fast growth in computing power and data source, deep learning has been widely applied to HAR applications. Therefore, our study on the effectiveness of IMU and stretch sensors in HAR is carried out with a special focus on the deep learning approach. Three deep learning networks which are long short-term memory (LSTM), 1-dimensional convolutional neural network (1D-CNN) and hybrid CNN-LSTM are deployed into the HAR model with different sets of input data. Detailed structures of the three deep learning-based HAR models are described in Fig. 2.



(a) LSTM-based HAR model

(b) CNN-based HAR model

(c) CNN-LSTM-based HAR model

(d) Detailed structure of the CNN block

Fig. 2. Detailed structure of three deep learning models considered in the study.

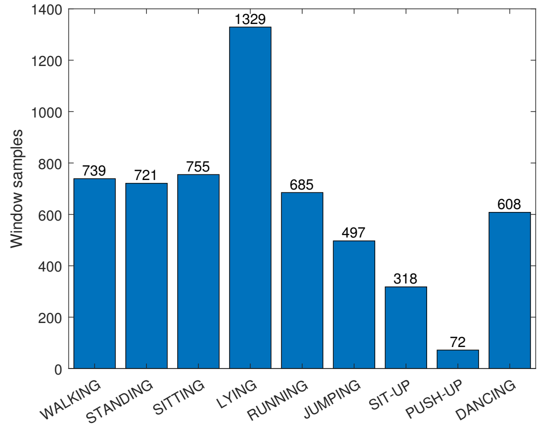
1) LSTM-based HAR: Long short-term memory network (LSTM) [9], with the advantage in being able to connect the information in the past and present, has been widely applied to a variety of problems that relate to sequential data such as speech recognition and neural machine translation. In wearable sensor-based HAR, data is collected through time and managed as a set of time series, therefore an LSTM-based deep learning model is implemented for classifying human activity. Details of this LSTM-based HAR model is shown in Fig. 2(a).

A batch normalization layer [10] is directly applied to the input data as a regularization method for reducing internal covariate shift and obtaining a faster convergence. The normalized sensor data is then forwarded to an LSTM layer where features are extracted along the time direction recurrently. Output hidden state of the final time step is fed into a dense layer (a.k.a fully-connected layer) with a ReLU activation function [11] for further analysis. Finally, another dense layer and a softmax activation function are deployed to normalize the outputs to a probability distribution over the predicted activities. During training, these normalized outputs are used to calculate the categorical cross-entropy loss, while in the testing phase, the activity with the highest probability is considered as the final prediction result.

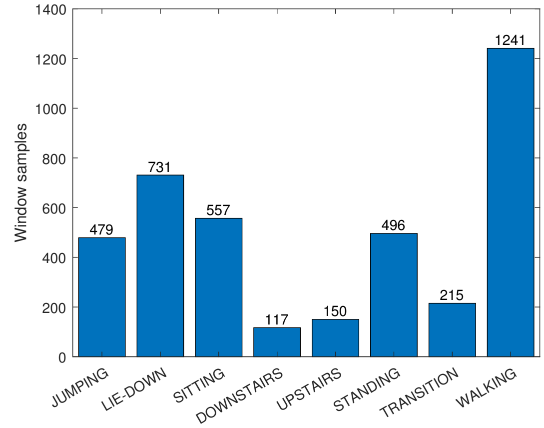
2) CNN-based HAR: LeNet [12], a convolutional neural network and its variants have been successfully applied to not only imagery data in computer vision but also to time series in other fields such as natural language processing and time series forecasting. The operating of sliding several convolution kernels throughout the input feature maps helps CNNs to reduce the number of parameters by using a parameter sharing scheme and utilizing parallel computing techniques. In this research, a 1-dimensional CNN model whose kernels slide along the time direction of the window sensor data $\mathbf{W}$ of length L and width N to extract local temporal features. N, number of attributes at a time step is treated as the depth of the input data, thus, kernels of the first convolutional layer have the depth equal to N.

The CNN-based HAR model contains a CNN block followed by a set of flatten and dense layers for activity classification as illustrated in Fig. 2(c). The CNN block is constructed from multiple convolutional, batch normalization and max-pooling layers. More details on this CNN block is described in Fig. 2(d). A description (32, 3, same, ReLU) of a convolutional layer can be understood as followed: there are 32 kernels whose length equals to 3; zero values are padded to the head and tail of the input in order to make the output has the same length with the input; a ReLU activation function is applied to the output.

3) Hybrid CNN-LSTM-based HAR: Recent research on HAR has proven that models consisting of CNN and LSTM can achieve better performance than single DL-based models [13]. Despite the power of LSTM in processing sequential data, LSTM-based HAR models cannot utilize parallel computing techniques as it has to process input sequence time step by time step. Furthermore, long sequences can easily lead to gradient vanishing and exploding problems during the training process with the use of backpropagation through time. Therefore, in the proposed hybrid CNN-LSTM-based HAR model, instead of applying LSTM directly to the window data $\mathbf{W}$, a 1D-CNN block is deployed to distill the input sequence, extract local temporal features and represent them in the depth dimension. Output feature map of this block has a size of $(L' \times N')$, where L' is the reduced length and N' is the new depth and equals to the number of kernels of the last convolutional layer. This new sequence of features is then fed into the LSTM layer for extracting long-term features. Finally, a group of dense layers with ReLU and softmax activation functions are used for further feature extraction and activity classification.



(a) iSPL IMU-Stretch Dataset



(b) w-HAR Dataset

Fig. 3. Window data distribution on different activities of the two datasets.

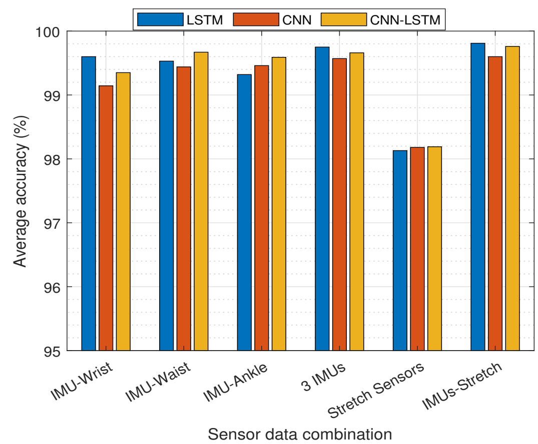
III. Experiment Results

A. Datasets

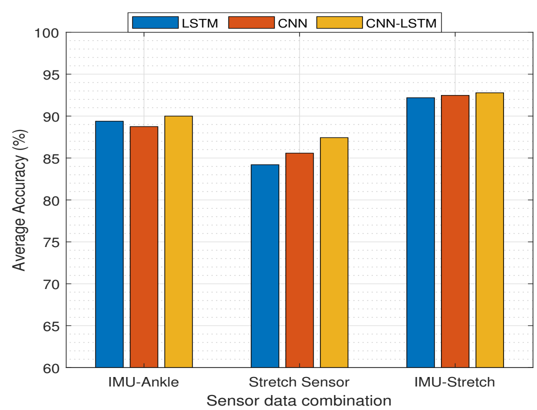
With the use of the proposed multi-modal HAR framework, we collected a dataset, namely iSPL IMU-Stretch, which contains data of 9 different daily-life activities and exercises: walking, standing, sitting, lying, running, jumping, sit-up, push-up and dancing. In addition, a public dataset called w-HAR [6] is also used for a comprehensive investigation on the two types of sensors. The w-HAR dataset was collected from an IMU attached at the ankle and a stretch sensor mounted at the knee. Eight activities including jumping, lying down, sitting, walking downstairs, walking upstairs, standing, walking and transition are collected from 22 users. Acceleration and angular velocity obtained from the IMU sensor are recorded at a sampling frequency of 250 Hz while stretch data is recorded at two frequencies of 100 Hz and 25 Hz. In this study, only the trials whose stretch sensor's sampling frequency is 25 Hz are used. The IMU sensor data are, then downsampled from 250 Hz to 25 Hz for reducing the computational cost. After the preprocessing step, there are 5724 and 3986 windows are obtained from iSPL IMU-Stretch and w-HAR datasets, respectively. The frequency distribution of all the activities in the two datasets is shown in Fig. 3.

B. Experiment Settings

The window data is randomly divided into training and testing sets with a proportion of 70% and 30%, respectively. All the deep learning models are implemented using the deep learning API Keras and are trained from scratch using the machine learning platform TensorFlow. The categorical cross-entropy loss function and adaptive moment optimizer (Adam) are used for the optimization process. For a fast convergence, the learning rate is initially set to 0.001 and reduced by half whenever the loss has stopped decreasing for 10 epochs. Each configuration (*i.e.*, model type and input data type) is trained and tested 10 times, and the average accuracy and F1 score are used as performance metrics. The iSPL IMU-Stretch dataset and implementation codes are available at https://github.com/thunguyenth/HAR_IMU_Stretch.



(a) iSPL IMU-Stretch Dataset



(b) w-HAR Dataset

Fig. 4. Average classification accuracy of different deep learning models on different combinations of sensor data.

TABLE I
F1 SCORE (%) OF DIFFERENT TYPES OF SENSOR DATA ON DIFFERENT ACTIVITIES USING HYBRID CNN-LSTM MODEL

	iSPL IMU-Stretch						w-HAR		
	IMU-Wrist	IMU-Waist	IMU-Ankle	3 IMUs	Stretch sensors	IMU-Stretchs	IMU-Ankle	Stretch-Knee	IMU-Stretch
Walking	99.57	99.79	99.82	99.84	**99.29**	**99.93**	**94.33**	**95.03**	94.81
Standing	99.14	99.80	99.69	99.60	**97.42**	**99.82**	81.69	**72.55**	**91.55**
Sitting	**99.25**	99.72	99.68	99.50	**99.84**	99.64	**85.19**	90.15	**94.13**
Lying	99.92	99.96	99.98	99.95	**99.08**	**100**	**99.39**	**84.56**	99.30
Running	99.40	99.23	99.40	99.62	**98.63**	**99.67**	-	-	-
Jumping	99.73	**99.83**	**99.54**	99.77	99.60	99.80	93.90	**90.78**	**94.52**
Sit-up	99.42	99.53	**99.06**	99.18	**100**	99.36	-	-	-
Push-up	94.28	98.82	**99.40**	98.46	**52.36**	99.01	-	-	-
Dancing	98.52	99.18	98.71	99.45	**96.60**	**99.48**	-	-	-
Walking downstairs	-	-	-	-	-	-	88.98	**90.18**	**88.15**
Walking upstairs	-	-	-	-	-	-	**91.21**	**93.17**	91.23
Transition	-	-	-	-	-	-	**57.36**	**62.93**	57.65

*Note: Values in bold and red color indicate the minimum F1 score of each activity in each dataset while values in bold and green color indicate the maximum values.

C. Results and Discussion

Different combinations of sensor data are considered in both datasets: a) single IMU sensor (wrist, waist, ankle), multiple IMU sensors, multiple stretch sensors and combined IMU-stretch sensors in the iSPL IMU-stretch dataset; b) single IMU sensor (ankle), single stretch sensor and combined IMU-stretch sensors in w-HAR dataset. The average accuracy of three DL-based HAR models on these sensor data combinations are shown in Fig. 4.

In terms of deep learning models, the average accuracies are similar for the three models in the iSPL IMU-Stretch dataset. Overall, the hybrid CNN-LSTM model obtains the most stable performance in all the data combinations of the two datasets. Although the LSTM model achieves the highest accuracy in most of the IMU sensor data, especially in the iSPL IMU-Stretch dataset, it gets the worst performance in stretch sensor data with 3% of accuracy lower than the CNN-LSTM model in the w-HAR dataset. This instability of the LSTM model can be caused when the model is incapable of retaining long-term information in univariate time series (*i.e.*, single stretch sensor data in the w-HAR dataset). In contrast, in the hybrid CNN-LSTM model, since the univariate stretch sensor data is distilled and changed to multivariate time series by the CNN block, the LSTM layer is able to preserve the long-term dependencies.

In terms of sensor data, data collected from single IMUs have higher accuracy than the data of stretch sensor with a gap of 1.5% and 5% in the iSPL IMU-Stretch and w-HAR dataset, respectively when the LSTM-based HAR is used. This is because the stretch sensor provides only 1-dimensional data which is the stretching degree of the knees while the IMU sensor provides a wide range of information (*i.e.*, 3-axial acceleration and angular velocity). Combining data of IMU sensor and stretch sensor helps improve the system performance by approximately 0.5% and 2% of accuracy in the two datasets.

To have a detailed look at the sensor efficiency, F1 score of different sensor data combinations on different activities across all the datasets are computed and described in Table I. Some interesting findings have been found from this result. First, even stretch sensor has the lowest classification accuracy compared to IMU sensor in all three deep learning model according to the results in Fig. 4, it still achieves the highest F1 score values in sitting, sit-up (in the iSPL dataset) and walking, walking downstairs, walking upstairs and transition (in the w-HAR dataset). Thus, when being attached at the knee, stretch sensor has high potential in differentiating activities that have similar patterns from the lower limbs. Second, by only using a simple sensor fusion method, the combination of 3 IMUs in iSLP IMU-Stretch dataset does not gain any highest F1 score. Similarly, the data combination of all the sensors does not gain the height F1 score at all activities in two datasets even though it has the most information compared to other data combinations. In particular, stretch sensor data in w-HAR dataset obtains the higher number of activity-wise F1-score

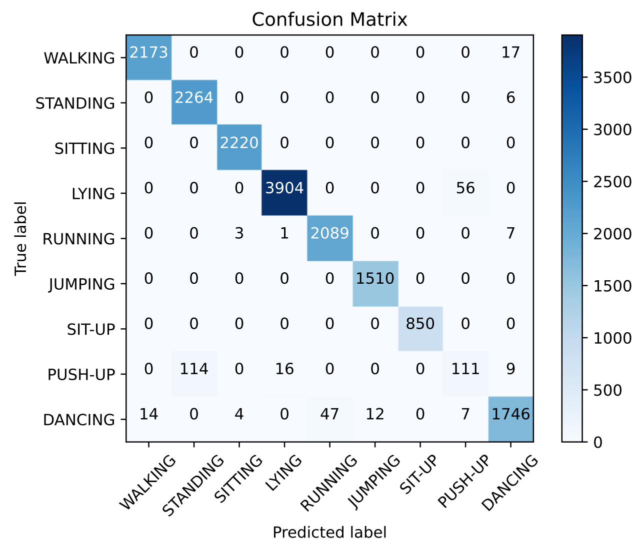


Fig. 5. Confusion matrix of the CNN-LSTM model on the stretch sensor of the iSPL IMU-Stretch dataset.

than the combination data. This finding suggests that by simply concatenating data of all the sensors might dull some important information that exists in the data of single sensors. Thus, this urges the need for developing a sensor fusion algorithm that can ignore noise data of a sensor but pay more attention to essential parts of other sensors. The synchronization among multiple sensors during data collection can also be a reason for this sensor fusion inefficiency.

Furthermore, problem of intra-class variance and inter-class similarity in human activity recognition is investigated. Firstly, the big gap in performance between the two datasets is the consequence of intra-class variance as the iSPL IMU-Stretch dataset is collected from one subject while the w-HAR is collected from 22 subjects. With the fact that each person has their own way to carry out an activity due to different factors such as age and height, data samples can be highly different from each other even though they are from the same activity. Secondly, a confusion matrix of the hybrid CNN-LSTM model on the stretch sensors of the iSPL IMU-Stretch dataset is demonstrated in Fig. 5 for inter-class similarity. It can be seen that push-up, standing, lying are easily misclassified to each other because the sensor can capture only the stretching and bending of the knees and the legs are stationary in a stretching state when these activities are carried out. Therefore, more research on tackling these intra-class variances and inter-class similarities in HAR is necessary.

IV. Conclusion

In this paper, several testing scenarios on different sensor data combinations and various deep learning-based models were performed to thoroughly study the effectiveness of IMU and stretch sensors in human activity recognition. With a wide range of sensor information, the multi-dimensional IMU sensor has outperformed the one-dimensional stretch sensor. However, the investigation results have also corroborated the potential of applying fabric, soft stretch sensors in capturing human motion. While this paper is a preliminary assessment of the practicality of soft, wearable sensors in human motion capture, we are expecting the future research in which multi-dimensional stretch sensors are developed and applied to promising applications of home healthcare such as diagnostic and therapeutics. Furthermore, on the basis of the promising findings presented in this paper, work on developing an optimal sensor fusion method is continuing and will be presented in future papers.

References

[1] M. Bächlin, M. Plotnik, D. Roggen, I. Maidan, J. M. Hausdorff, N. Giladi, and G. Troster, "Wearable assistant for parkinson's disease patients with the freezing of gait symptom," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 436–446, 2010.

[2] C. Chatzaki, M. Pediaditis, G. Vavoulas, and M. Tsiknakis, "Human daily activity and fall recognition using a smartphone's acceleration sensor," in *International Conference on Information and Communication Technologies for Ageing Well and e-Health*. Springer, 2016, pp. 100–118.

[3] J.-L. Reyes-Ortiz, L. Oneto, A. Samà, X. Parra, and D. Anguita, "Transition-aware human activity recognition using smartphones," *Neurocomputing*, vol. 171, pp. 754–767, 2016.

[4] Y. Ling, T. An, L. W. Yap, B. Zhu, S. Gong, and W. Cheng, "Disruptive, soft, wearable sensors," *Advanced Materials*, vol. 32, no. 18, p. 1904664, 2020.

[5] H. Chander, E. Stewart, D. Saucier, P. Nguyen, T. Luczak, J. E. Ball, A. C. Knight, B. K. Smith, R. Prabhu *et al.*, "Closing the wearable gap—part iii: use of stretch sensors in detecting ankle joint kinematics during unexpected and expected slip and trip perturbations," *Electronics*, vol. 8, no. 10, p. 1083, 2019.

[6] G. Bhat, N. Tran, H. Shill, and U. Y. Ogras, "w-HAR: An activity recognition dataset and framework using low-power wearable devices," *Sensors (Switzerland)*, vol. 20, no. 18, pp. 1–26, 2020.

[7] B. Klaassen, B. J. van Beijnum, M. Weusthof, D. Hofs, F. van Meulen, E. Droog, H. Luinge, L. Slot, A. Tognetti, F. Lorussi, R. Paradiso, J. Held, A. Luft, J. Reenalda, C. Nikamp, J. Buurke, H. Hermens, and P. Veltink, "A full body sensing system for monitoring stroke patients in a home environment," *Communications in Computer and Information Science*, vol. 511, pp. 378–393, 2015.

[8] R. A. Light, "Mosquitto: server and client implementation of the mqtt protocol," *Journal of Open Source Software*, vol. 2, no. 13, p. 265, 2017.

[9] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[10] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 37. PMLR, 2015, pp. 448–456.

[11] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ser. ICML'10. Madison, WI, USA: Omnipress, 2010, p. 807–814.

[12] Y. LeCun, B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard, and L. Jackel, "Handwritten digit recognition with a back-propagation network," in *Advances in Neural Information Processing Systems*, D. Touretzky, Ed., vol. 2. Morgan-Kaufmann, 1990, pp. 396–404.

[13] N. T. H. Thu and D. S. Han, "HiHAR: A hierarchical hybrid deep learning architecture for wearable sensor-based human activity recognition," *IEEE Access*, vol. 9, pp. 145 271–145 281, 2021.