Solving Smartphone-based Human Activity problem using Recurrent Neural Network with CNN features

First Authora,[[1]](#footnote-1)\*, Second Authorb and Third Authorb

a*Journal Production Department, IOS Press, Nieuwe Hemweg 6b, 1013 BG, Amsterdam, The Netherlands*

b*Deparment first, then University or Company name, Insert a complete correspondence (mailing) address, Abbreviate US states, Include country*

**Abstract.** Wearable-based approach and vision-based approach are two of the most common approaches in human activity recognition. However, these approaches are inconvenient and not easy to implement. For instance, there is a privacy issue revolving in the vision-based approach. Hence, smartphone-based human physical activity recognition is a popular alternative. In this paper, we propose a HAR model connecting CNN and Bidir-LSTM to interpret and predict accelerometer data captured using a smartphone for activity recognition. The proposed deep model is able to extract deep features from both spatial and temporal domains of the inertial data. The recognition accuracy of the proposed model is assessed using UCI and WISDM accelerometer data. Empirical results exhibit a promising performance.

Keywords: Keyword one, keyword two, keyword three, keyword four, keyword five

# Introduction

On World Health Organization (WHO) website, Global Health Observatory (GHO) data reveals approximately 23% of adults are not sufficiently active [1]. The level of insufficient physical activity is even worse in those high-income countries, which about 60% adults are insufficiently physically active. Physical inactivity is one of the ten top risk factors for mortality. WHO applauded that there is 20% to 30% increased risk of all-cause mortality to people who are insufficiently physically active, compared to those who have at least two and half hours’ moderate intensity physical activity per week.

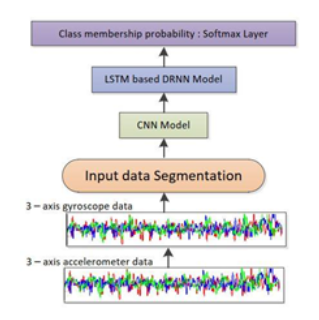
Overwhelming testimony substantiates the belief that insufficiency of physical activity contributes to a host of chronic diseases such as ischemic heart disease, high blood pressure, diabetes, stroke, hypertension, depression and cancers [2][3]. The upsurge of chronic diseases potentially impacts social and economic costs, for example unemployment at the labour market, financial burden etc. [4]. Hence, communities must take action to increase individuals’ physical activity.

Regular recognition and self-monitoring of physical activity potentially encourage habits of adopting a healthy lifestyle such as regular exercise [5][6]. ICT-enabled assisted living or “ambient assisted living” (AAL) systems are solutions for the objective. Intelligent human activity recognition (HAR) system is one of the crucial components in such innovations. There are three kinds of HAR systems: vision-based, wearable sensor-based and smartphone-based [4][7][8][9][10].

# Motivation

Wearable sensor-based approach and vision-based approach are two of the most common approaches in HAR. Both attain promising recognition performance, reaching above 80% accuracy in recognizing human physical activity [7][11][12]. However, these approaches are inconvenient and not easy to implement. For instance, there is a privacy issue revolving in the vision-based approach. Placing a surveillance camera in public places may violate the law and require extensive justification to obtain permits. On the other hand, in wearable sensor-based approach, some people are reluctant to wear the sensor device(s) [13].

Henceforth, physical activity prediction using smartphone is a contemporary research area in HAR domain. Smartphone is a sensor-based ubiquitous piece of technology that is far more than just a communication device. With great technology development, smartphones are packed with high-end hardware and features. A number of sensors is embedded inside smartphones, including motion sensors. The potential of smartphone-based HAR approach is uplifted due to the mobility and simplicity of smartphone usage, providing people something that is very accessible and easy to use. Hence, a smartphone-based human physical activity recognition system based on temporal dynamics deep features is proposed in this paper. An amalgamation of one-dimensional (1D) convolutional neural network and recurrent neural network (RNN) variant, i.e. bi-directional Long Short Term Memory, is proposed to predict human activity based on the inertial data captured from smartphone, illustrated in Fig. 1.



Redraw but based on this fig to fit your model

LSTM based DRNN model is changed to “RNN variant model”

Class membership … changed to classifier

Add one more layer as predicted activity

# Literature Review

As aforementioned, HAR can be categorized into three spheres: (1) vision-based HAR, (2) wearable sensor-based HAR and (3) smartphone-based HAR. Vision-based human activity recognition (coined as VHAR) is a process of categorizing a sequence of image recording with action/ activity class labels [7]. VHAR systems are extensively employed in various applications, especially for public area surveillance, healthcare monitoring as well as human-computer interaction. The proposed VHAR approaches include but not limited to Discrete Fourier Transform-based HAR to extract global representation of activity data [14], stacked Fisher vectors to capture more statistical information from frame images [15], extraction of multi features from body silhouettes and joints information [16] and etc.

However, the privacy issue in VHAR is a major concern from the public. Hence, wearable sensor-based HAR (coined as WHAR) is proposed as an alternative. The main applications of WHAR are in the areas of healthcare, sport training, smart environments etc. Wearable sensors include accelerometer, gyroscope and magnetometer. The literature of WHAR are implementing artificial neural network and smartwatch for performing HAR [17], adoption of Convolutional Neural Network for k-nearest neighourhood-based wearable sensor HAR [18], employment of J48 classifier in HAR for wearable sensors [19] etc.

Inconveniences of wearing, technological barriers such as limitation of current battery technology and culture barriers such as the association of a stigma with the use of medical sensing devices for monitoring limit the potential of WHAR usage. Smartphone is a seemly alternative for collecting motion inertial data signals. Most smartphones are equipped with built-in gyroscope and accelerometer. In the recent years, there are extensive research working on adopting smartphone for HAR [4][8][20][21][22][23][24].

~~In smartphone-based HAR, two popular public available databases Due to the challenge of performing data collection and collecting sufficient data to train and test HAR system, the utilization of public datasets such as Wireless Sensor Data Mining (WISDM) dataset and UC Irvine (UCI) dataset have been a popular alternative among researchers [10], [20], [25]–[28].~~ Kwapisz et al. utilize triaxial acceleration data captured by an Android smartphone to perform human activities [10-check]. The raw triaxial acceleration data is divided into 10-second segments. Then, forty-three statistical features are computed for each segment. The authors evaluate their self-collected dataset, namely Wireless Sensor Data Mining (WISDM) dataset, with various kinds of classifiers. Empirical results show a promising performance.

In [20], a new database is collected with six different activity classes from a group of 30 volunteers carrying the smartphone on their waist. The collected data is triaxial acceleration and angular velocity data. The collected inertial signals are sampled in fixed-width sliding windows of 2.56 seconds with 50% overlap between them. Next, 561 time and frequency domain features are extracted to describe each activity window. Support Vector Machine (SVM) is used to classify the activities.

~~Utilizing the WISDM dataset, Lockhart (2014) has analyzed the learning pattern for various model types which reveal that small amounts of personal data can surpass even the best model. Similarly, with the help of the WISDM dataset, A. D. Ignatov Strijov (2016) have trained and evaluate the data using k-Nearest Neighbour (kNN) algorithms along with neural network method such as using two-layer perceptron with sigmoid activation function trained with backpropagation. This method also registered 96% accuracy in determining the activities carried out.~~

## ~~Deep Learning in HAR~~

Human activities are hierarchical. Hence, deep learning comprising multiple layers of neural networks is explored to reveal features from low to higher levels hierarchically. Deep learning has become a critical research in HAR [25][26][30][31]. Convolutional Neural Network (CNN) or commonly referred as Covnet (LeNet) is a popular deep learning model [35][36]. CNN is proposed to capture the local dependencies and spatial domain of activity signals [40]. The authors utilize multichannel time series data to recognize users’ activity and hand gestures. Since inertial signal is an one-dimensional (1D) data, the traditional CNN has been improved to perform 1D convolution operation on the accelerometer and gyroscope triaxial sensor data [8][25].

~~It is three dimensional, consisting of height, width, and depth. Another version of the convolutional layer is called the fully connected layer (henceforth be known as Dense layer), whereby each neuron receives input from the previous layer, while CNN neurons only receives a specific part of the input from the previous layer. CNN is widely utilized in area such as image identification [37]–[39], however in recent years, it is widely adopted in HAR as well. For instance, [40] used CNN to capture the local dependencies and spatial domain of activity signals. They utilized multichannel time series data to recognize the user’s activity and hand gestures.~~

~~constructed a deep CNN that perform one-dimensional (1D) convolution on the accelerometer and gyroscope triaxial sensor data. Their network achieved 94.79% accuracy on the data. Similarly, [8] also uses a 1D CNN method for HAR on the data gathered using a built-in accelerometer sensor in the smartphone. The triaxial acceleration data are converted into a vector input for the 1D CNN to learn and train. The model achieved an accuracy of 92.71%.~~

~~1D CNN is utilized for a single time series data, while two-dimensional (2D) CNN works best in multiple time series to capture local dependency and spatial domain of multi-modal data. [41] applied 2D CNN on multi-modal data and achieved a classification accuracy of 99.66% based on their model. It achieves significantly high performance with less number of parameters when compared to 1D operation.~~

~~Understanding that Recurrent neural network (RNN) is inspired by Hopfield Net developed by [42]. It was developed to process sequential or continuous data such as time signal and sensor data [31]. RNN utilizes a technique call recurrent, meaning the network will consistently update each node based on the old and new information available. The state of the network will also update accordingly. In short, when the previous state is activated, the current state is predicted by estimating the next state.~~

~~However, as pointed out by [43], and [44], the vanishing or exploding gradient makes training the model more challenging. Exploding gradient refers to a large error accumulated causing a significant change in the neural network while vanishing gradient refers to errors that lead gradient to be vanished. Vanishing gradient problem can attribute to RNN forgetting the first inputs due to memory cells have a short-term memory, meaning the information receives earlier may be lost and this causes the model not able to reflect its current state [28], [45]. This limits its application for HAR, image recognition or computer vision. Thus, an alternative is proposed.~~

~~Long Short-Term Memory (LSTM) is a variant of Recurrent Neural Network (RNN), first proposed by [46]. It can get rid of the vanishing gradient problem in RNN by implementing four gates and enable long-range learning. The four gates are the forget gate, input gate, state gate (or a neuron with a self-recurrent connection) and output gate in each node.~~

Long Short-Term Memory (LSTM), a variant of Recurrent Neural Network (RNN), has been proposed on triaxial accelerometers data for smartphone-based human activity prediction [27]. LSTM adopts past information to predict the outcome of a HAR model. It allows the network to learn when to “forget” previous hidden states and when to update hidden states given new information. However, some information may not be captured since human motion is continuous [28].

Since LSTM only takes in past information, Bi-directional Long Short-Term Memory, coined as BLSTM, is proposed to tackle both past information and future information. In other words, BLSTM is stacked into layers both horizontally and vertically. In the model, a single LSTM node can take in information from the horizontal layer for both past and future information, as well as from a vertical layer which is the lower hidden layer. [28] and [Deep Recurrent Neural Networks for Human Activity Recognition, Abdulmajid, 2017] utilizes BLSTM on HAR using the inertial sensor in the smartphone. Experimental results demonstrate that the models outperform other existing approaches.

~~, achieving an accuracy of 93.79%. The result is comparable to the study conducted by [25] who used a deep learning convolutional neural network (CNN) on raw data and achieve an accuracy of 94.79%. Similar work has also been done by [45] who proposed a Bidir-LSTM network that takes in accelerometer and gyroscope data from a smartphone. They achieve and accuracy of 92.67%.~~

~~Bi-directional Long Short-Term Memory (Bidir-LSTM) is a variant of LSTM. However, different from LSTM who only takes in past information, Bidir-LSTM uses both past information and future information. In other words, Bidir-LSTM is stacked into layers both horizontally and vertically. A single LSTM node can take in information from the horizontal layer for both past and future information and also from a vertical layer which is the lower hidden layer. [28] utilizes Bidir-LSTM on HAR using the inertial sensor in the smartphone, achieving an accuracy of 93.79%. The result is comparable to the study conducted by [25] who used a deep learning convolutional neural network (CNN) on raw data and achieve an accuracy of 94.79%. Similar work has also been done by [45] who proposed a Bidir-LSTM network that takes in accelerometer and gyroscope data from a smartphone. They achieve and accuracy of 92.67%.~~

~~A fusion of CNN and LSTM has been proposed by [47] called Long-term Recurrent Convolutional network (LRCN) model. It can operate well in activity recognition, image captioning and video description. Its work shows that CNN which allow arbiratry-sized input to be map into fixed-sized vector representation can work well with the subsequent LSTM which takes in the outputs of CNN to be passed into a recurrent sequence learning module to be predicted.~~

# Contributions of the work

Upholding the hypotheses of (1) the spatial and temporal information embedded in the inertial signal is crucial to represent activity, and (2) human activities are hierarchical, we propose a temporal dynamics deep learner that extract features from low to higher levels hierarchically from spatial and temporal domains. The main contributions of this work are summarized into threefold:

1. A stacking spatial-temporal deep model is developed to extract low level to higher level features of inertial data for human activity recognition. Piling a convolutional architecture to deep BLSTM models enables both spatial and temporal state dependencies encapsulation to predict human activity.
2. Various machine learning algorithms are explored to evaluate the effectiveness of the deep features extracted by the proposed dynamics deep model. These machine learning algorithms include logistic regression, support vector machine, Naïve Bayes, random forest, multilayer perceptron, k-nearest neighbours etc.
3. An extensive experimental analysis is conducted on two public available datasets, namely WISDM and UCI datasets. Influence of hyperparameter settings towards the recognition performance is also presented on top of addressing the performance comparison with other approaches.



# Proposed Solution

This work proposes a temporal deep learner stacking a hierarchical convolutional architecture with a model that comprehends dynamics pattern of the inertial sequential data to predict human activity. Fig. 2 illustrate the proposed solution. In the architecture, the feature extraction structure comprises of three convolutional layers, one max-pooling layers and one layer of BLSTM. Then, the deep features are further classified using machine learning algorithm. In this work, we only consider acceleration data acquired from accelerometer sensor of the smartphone. In other words, three features of inertial signal are taken into account, that is triaxial (*x*, *y* and *z*-) acceleration data. Utilizing these narrow features, the deep feature extraction structure will extract those underlying rich features before performing classification.

The neural processors of the lower layers attain local features of the inertial signal to signify the elementary motion in physical activity; whilst, higher layer neural processors extract a better abstraction of the motion with higher level features and temporal analysis. Convolutional layers read a sequence data, i.e. one-dimensional signal, using a kernel that reads in small windows at a time and strides across the entire input signal. Each read results an input to be projected onto a feature map, representing the internal interpretation of the input. Since each convolutional layer contains multiple neurons/kernels, multiple feature maps will be constructed after every layer and concatenated, as illustrated in Fig. 3.

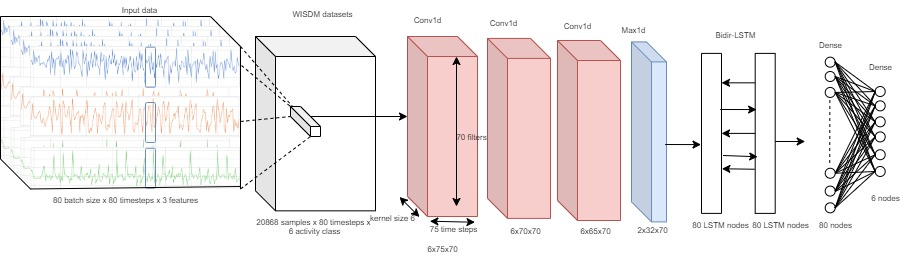
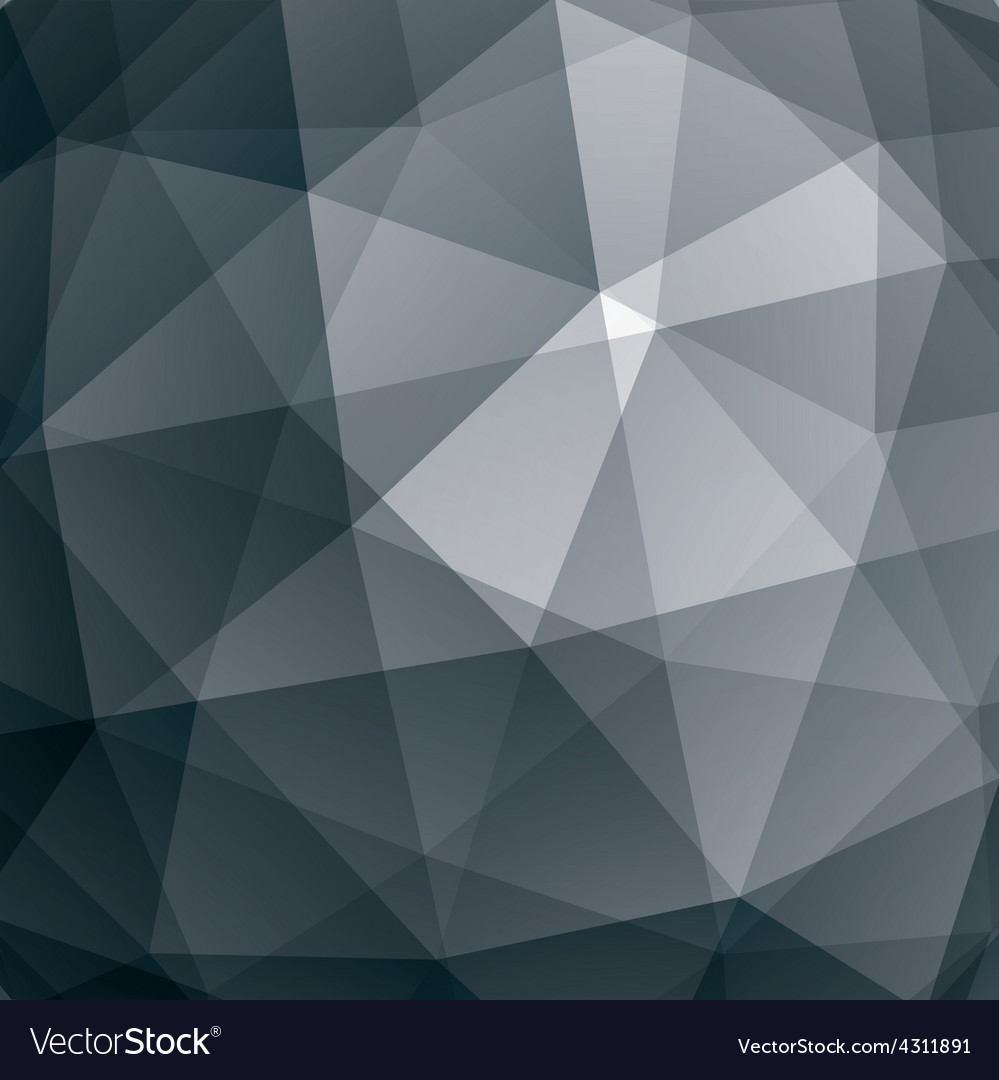
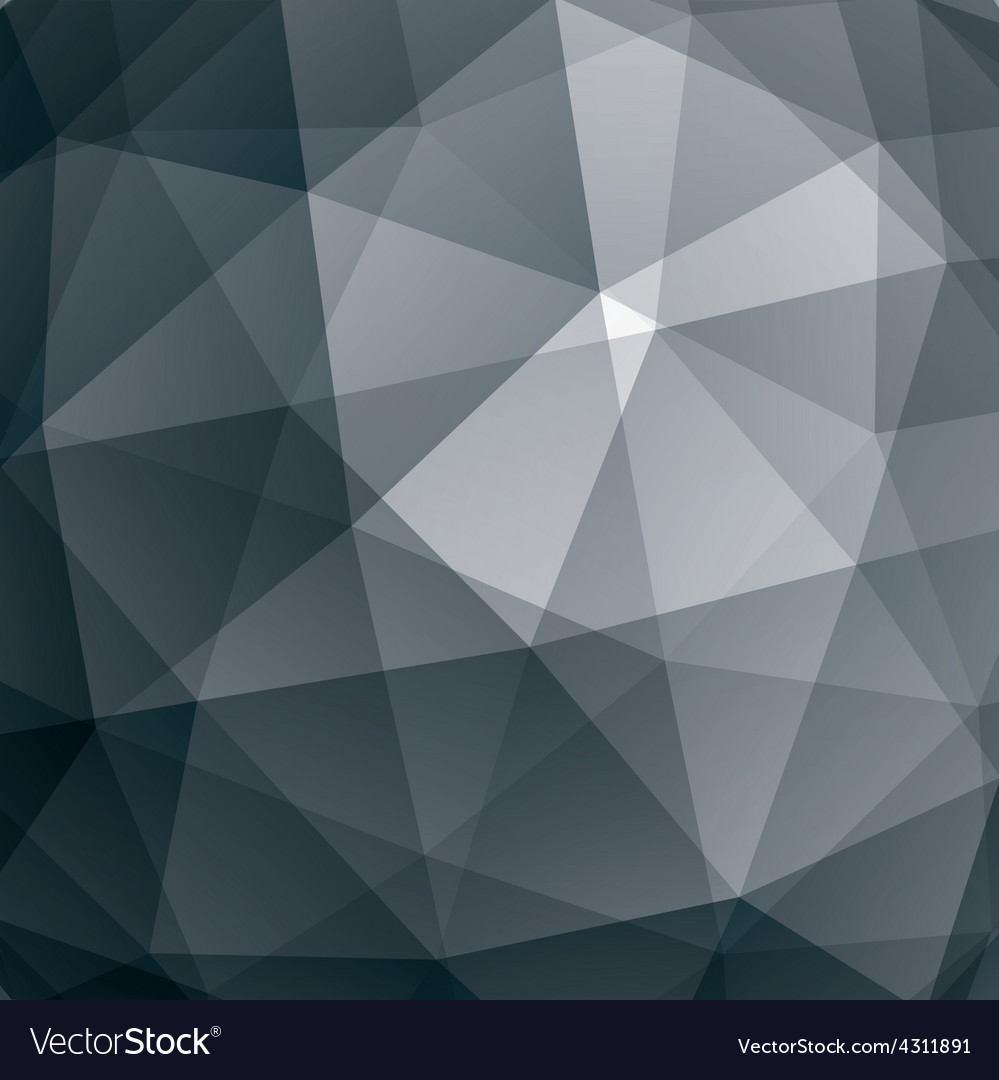
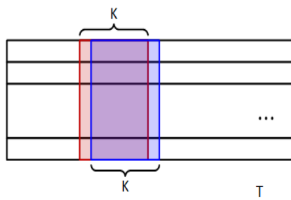
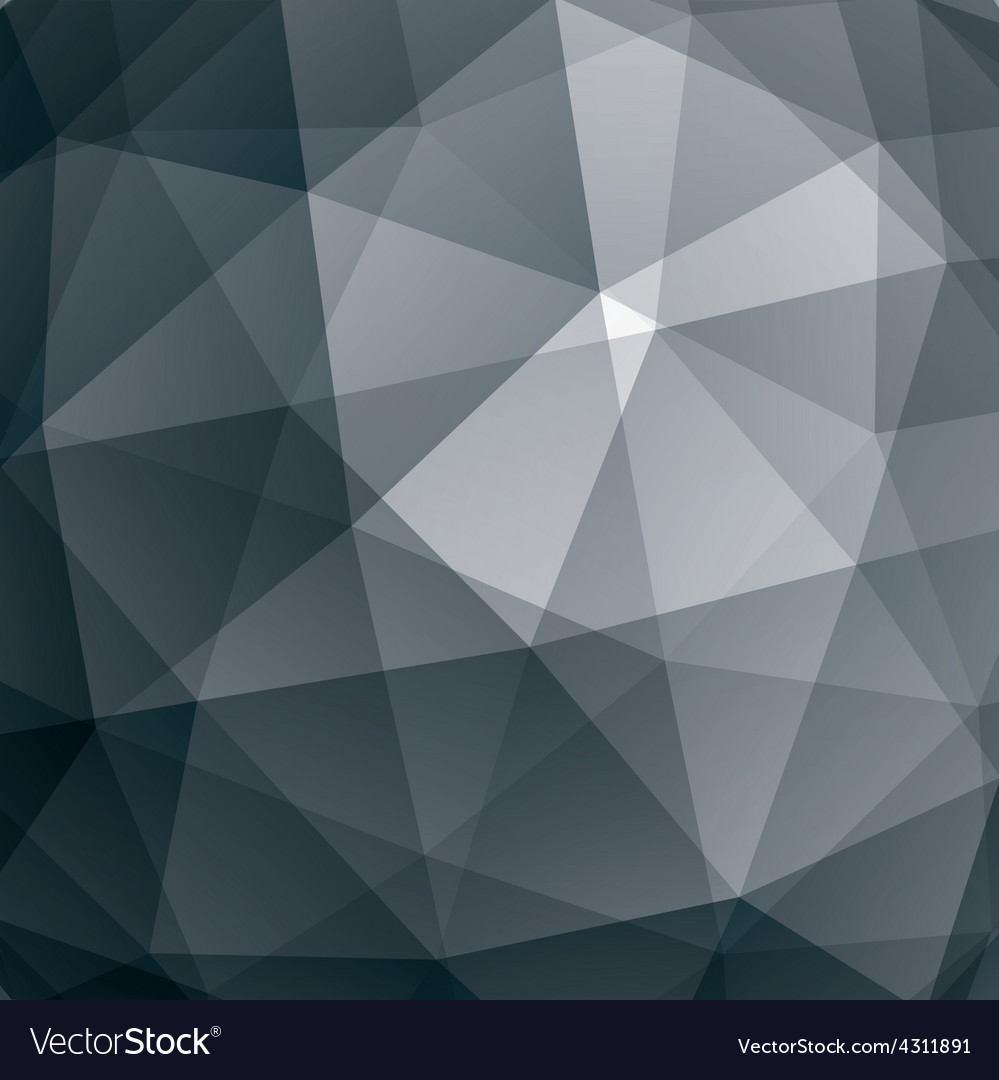


Fig. 2

Neuron A





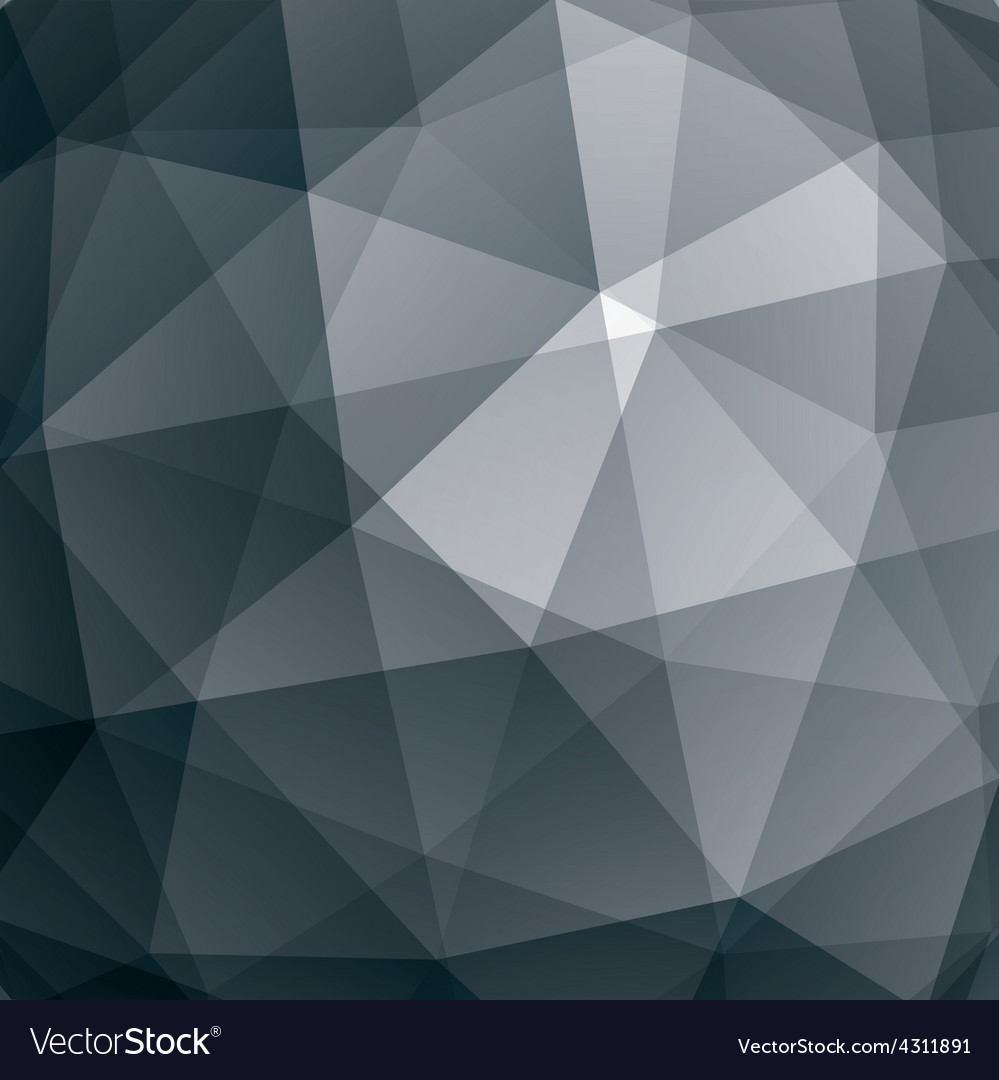
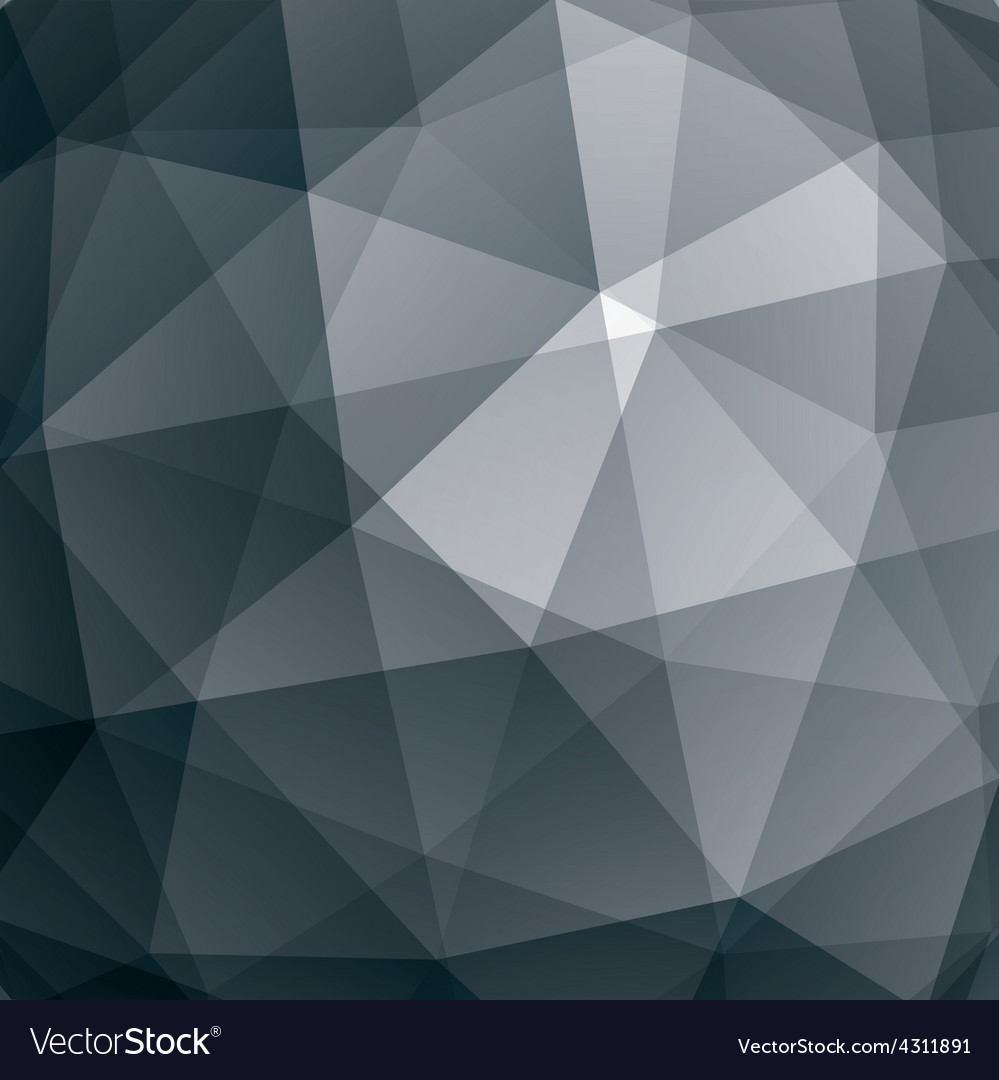


N

concatenation

Convolution across signal

Feature map



Neuron N

Convolution across signal

Feature map

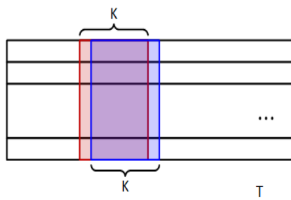


Fig. 3. XXXX

Through the convolution operation, those local dependencies in the inertial data could be apprehended. Hence, the correlation between nearby signal points could be pictured, revealing the structure of the signal pattern. Next, max-pooling layer is implemented to down sample each feature map independently for a summarized version of the captured features. In addition, pooling helps for model’s invariance to local translations of the input. With this property, a slight translational variance of the input data will not affect the values of the pooled output.

xxx

Since CNN model unearths those underlying patterns in input signal, it is able to encapsulate the tiny changes in the motion signal. These changes in sequential form are substantial to characterize activity motion. Hence, RNN variant model is included in the proposed architecture to build time dynamics for the feature map by analyzing the underlying sequential pattern in the spatial-temporal feature map.

In the proposed architecture, bidirectional LSTM (BLSTM) is adopted since it can have better prediction using both past and future information, i.e. utilizing information from the previous and upcoming frames [ref A. Ogawa and T. Hori, ‘‘Error detection and accuracy estimation in 599 automatic speech recognition using deep bidirectional recurrent neural 600 networks,’’ Speech Commun., vol. 89, pp. 70–83, May 2017.]. BLSTM is just like stacking two LSTM on top of each other, illustrated in Fig. 4. One LSTM moves in the forward direction, while the other one moves in the opposite, i.e. backward, direction. Then, the outputs of LSTMs are fused and computed as BLSTM output. After then, the deep features of BLSTM are extracted and fed into machine learning classifier for activity prediction.

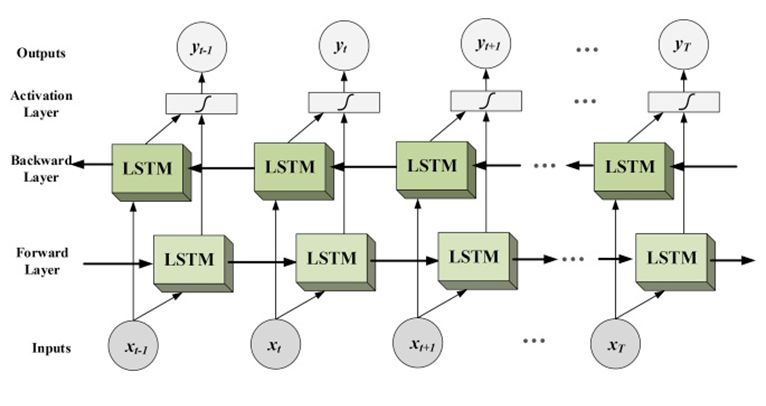


Fig. 4 XXX (picture source is from [ref https://www.i2tutorials.com/technology/deep-dive-into-bidirectional-lstm/]

## Formulation of CNN and BLSTM

In this work, both CNN and Bidir-LSTM are based on [28]’s equations which derived from [35], [42]. The CNN formulation is simplified as followed:

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

Where  denotes CNN output at th layer,  is the activation function such as Relu,  refers to the bias term in CNN, *W* is the weight from the previous layer in CNN, V is the input vector of our accelerometer signal in our datasets.

Max Pooling conduction pooling operation, usually on CNN output. The pooling operation converts the input sets to the highest:

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

Where P is the output of the pooling operation, and  is the CNN output at the  th layer. CNN and max-pooling often place subsequent to each other in deep learning models to learn representations of a set of input.

Taking the summarized version of feature map P from max pooling operation, the LSTM fed this vector into its memory cell and four gates which consist of the forget gate, input gate, input modulation gate, and output gate. First, we look into the forget gate, at any given timestep :

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

Where  is the forget gate output at n layer at timestep  and  is the sigmoid function, denotes the weight of the connection at forget gate, is the LSTM output from the previous layer,denotes the input vector of the dataset, and  denotes the bias term for the forget gate. The next gate is the input gate which is as follows:

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

Where  is the input gate output at n layer at timestep  and  is the sigmoid function,  denotes the weight of the connection at the input gate,  is the LSTM output from the previous layer, denotes the input vector of the dataset, and  denotes the bias term for the input gate. Similar to the input gate and forget gate, the output gate exhibits similar equations:

|  |  |  |
| --- | --- | --- |
|  |  | (5) |

Where  denotes the output gate output at n layer at timestep and  is the sigmoid function, denotes the weight of the connection at the output gate,   is the LSTM output from the previous layer,  denotes the input vector of the dataset, and  denotes the bias term for the output gate. Next is the input modulation gate which is a function of the input vector and the previous state output:

|  |  |  |
| --- | --- | --- |
|  |  | (6) |

Where  denotes the input modulation gate output at n layer at timestep with  function,  denotes the weight of the connection at the input modulation gate which is based on the previous state,  is the LSTM output from the previous layer,  denotes the input vector of the dataset, and  denotes the bias term for the input modulation gate based on previous state gate.

The state gate or memory cell consists of two terms: the previous memory cell state which is modulated by forget gate , and input modulation gate which is modulated by input gate  at timestep  :

|  |  |  |
| --- | --- | --- |
|  |  | (7) |

Ergo, the LSTM outputs the combination of the output gate and state gate, with the following equation:

|  |  |  |
| --- | --- | --- |
|  |  | (8) |

Where denotes the LSTM output at n layer at timestep  ,  refers to the output gate’s output at n layer with  function,  denotes the state gate’s output at n layer as well. The LSTM node output two values to the next LSTM nodes. The subsequent LSTM node will use these two information accordingly to update their state and eventually the update status of the whole network. This gives the LSTM network takes the ability to take into past information.

Bidir-LSTM has both forward sequences  and backward sequences  in the hidden layer. At time t, the hidden layer and the input layer can be defined as follows:

|  |  |  |
| --- | --- | --- |
|  |  | (9) |
|  |  | (10) |
|  | ) | (11) |

Where  denotes the forward sequence, denotes the backward sequence of LSTM operation,  refers to the Bidir-LSTM output, A refers to the activation function used in the network, W refers to the weight of the connection and B refers to the bias term. These forward and backward sequences allow the LSTM nodes to takes in previous and subsequent information to update its state, this will subsequently update the state of the whole network. This gives the Bidir-LSTM’s properties to use past and future information to effectively output a deeper representation of a set of data inputs.

# Experiment

## Experiment Setup

# Evaluation

# References

[1] “WHO | Prevalence of insufficient physical activity,” *WHO*, 2018.

[2] F. W. Booth, C. K. Roberts, and M. J. Laye, “Lack of exercise is a major cause of chronic diseases,” *Compr. Physiol.*, vol. 2, no. 2, pp. 1143–1211, Apr. 2012.

[3] G. Cattadori, C. Segurini, A. Picozzi, L. Padeletti, and C. Anzà, “Exercise and heart failure: an update,” *ESC Heart Failure*, vol. 5, no. 2. Wiley-Blackwell, pp. 222–232, Apr-2018.

[4] R. A. Voicu, C. Dobre, L. Bajenaru, and R. I. Ciobanu, “Human physical activity recognition using smartphone sensors,” *Sensors (Switzerland)*, vol. 19, no. 3, Feb. 2019.

[5] R. A. Carels, L. A. Darby, S. Rydin, O. M. Douglass, H. M. Cacciapaglia, and W. H. O’Brien, “The relationship between self-monitoring, outcome expectancies, difficulties with eating and exercise, and physical activity and weight loss treatment outcomes,” *Ann. Behav. Med.*, vol. 30, no. 3, pp. 182–190, Dec. 2005.

[6] M. M. MacPherson, K. J. Merry, S. R. Locke, and M. E. Jung, “Effects of Mobile Health Prompts on Self-Monitoring and Exercise Behaviors Following a Diabetes Prevention Program: Secondary Analysis From a Randomized Controlled Trial.,” *JMIR mHealth uHealth*, vol. 7, no. 9, p. e12956, Sep. 2019.

[7] R. Poppe, “A survey on vision-based human action recognition,” *Image Vis. Comput.*, vol. 28, no. 6, pp. 976–990, 2010.

[8] S. M. Lee, S. M., Cho, H., & Yoon, “Human Activity Recognition From Accelerometer Data Using Convolutional Neural Network,” *IEEE Int. Conf. Big Data Smart Comput. (BigComp).*, vol. 62, pp. 131–134, 2017.

[9] K. Lee and M. P. Kwan, “Physical activity classification in free-living conditions using smartphone accelerometer data and exploration of predicted results,” *Comput. Environ. Urban Syst.*, vol. 67, no. September 2017, pp. 124–131, 2018.

[10] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, “Activity recognition using cell phone accelerometers,” *ACM SIGKDD Explor. Newsl.*, 2011.

[11] L. Gao, A. K. Bourke, and J. Nelson, “Evaluation of accelerometer based multi-sensor versus single-sensor activity recognition systems,” *Med. Eng. Phys.*, vol. 36, no. 6, pp. 779–785, 2014.

[12] C. Li, “DigitalCommons@USU Wearable Computing: Accelerometer-Based Human Activity Classification Using Decision Tree,” 2017.

[13] D. Ledger and D. McCaffrey, “Inside wearables: how the science of human behavior change offers the secret to long-term engagement,” 2014.

[14] S. Kumari and S. K. Mitra, “Human action recognition using DFT,” in *Proceedings - 2011 3rd National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics, NCVPRIPG 2011*, 2011.

[15] X. Peng, C. Zou, Y. Qiao, and Q. Peng, “Action recognition with stacked fisher vectors,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2014.

[16] A. Jalal, S. Kamal, and D. Kim, “A Depth Video-based Human Detection and Activity Recognition using Multi-features and Embedded Hidden Markov Models for Health Care Monitoring Systems,” *Int. J. Interact. Multimed. Artif. Intell.*, vol. 4, no. 4, p. 54, 2017.

[17] M. C. Kwon and S. Choi, “Recognition of Daily Human Activity Using an Artificial Neural Network and Smartwatch,” *Wirel. Commun. Mob. Comput.*, 2018.

[18] S. Sani, N. Wiratunga, and S. Massie, “Learning Deep Features for kNN-Based Human Activity Recognition,” 2017.

[19] K. W. Ching, M. M. Singh, and Z. F. Zaaba, “Human activity recognition (HAR) for wearable sensors with classification techniques,” *Adv. Sci. Lett.*, vol. 23, no. 5, pp. 4206–4210, May 2017.

[20] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, “A public domain dataset for human activity recognition using smartphones,” in *ESANN 2013 proceedings, 21st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 2013.

[21] 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, 2014.

[22] T. Brezmes, J. L. Gorricho, and J. Cotrina, “Activity recognition from accelerometer data on a mobile phone,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5518 LNCS, no. PART 2, pp. 796–799, 2009.

[23] A. Ignatov, “Real-time human activity recognition from accelerometer data using Convolutional Neural Networks,” *Appl. Soft Comput. J.*, vol. 62, pp. 915–922, 2018.

[24] J. W. Lockhart, “The Benefits of Personalized Data Mining Approaches to Human Activity Recognition with Smartphone Sensor Data,” p. 46, 2014.

[25] 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, 2016.

[26] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, “Deep learning for sensor-based activity recognition: A survey,” *Pattern Recognit. Lett.*, 2019.

[27] Y. Chen, K. Zhong, J. Zhang, Q. Sun, and X. Zhao, “LSTM Networks for Mobile Human Activity Recognition,” no. Icaita, pp. 50–53, 2016.

[28] S. Yu and L. Qin, “Human activity recognition with smartphone inertial sensors using bidir-LSTM networks,” *Proc. - 2018 3rd Int. Conf. Mech. Control Comput. Eng. ICMCCE 2018*, pp. 219–224, 2018.

[29] A. D. Ignatov and V. V. Strijov, “Human activity recognition using quasiperiodic time series collected from a single tri-axial accelerometer,” *Multimed. Tools Appl.*, vol. 75, no. 12, pp. 7257–7270, 2016.

[30] X. Shi, Y. Li, F. Zhou, and L. Liu, “Human Activity Recognition Based on Deep Learning Method,” *2018 Int. Conf. Radar, RADAR 2018*, 2018.

[31] H. Friday Nweke, T. Ying Wah, and U. Alo, “Deep Learning Algorithms for Human Activity Recognition using Mobile and Wearable Sensor Networks: State of the Art and Research Challenges Mobile Cloud Computing View project Novel Deep Learning Architecture for Physical Activities assessment, mental Res,” vol. 105, pp. 233–261, 2018.

[32] J. Louradour and C. Kermorvant, “Curriculum learning for handwritten text line recognition,” in *Proceedings - 11th IAPR International Workshop on Document Analysis Systems, DAS 2014*, 2014.

[33] L. Mou, P. Ghamisi, and X. X. Zhu, “Deep recurrent neural networks for hyperspectral image classification,” *IEEE Trans. Geosci. Remote Sens.*, 2017.

[34] S. Ji, W. Xu, M. Yang, and K. Yu, “3D Convolutional neural networks for human action recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, 2013.

[35] Y. LeCun, P. Haffner, L. Bottou, and Y. Bengio, “Object recognition with gradient-based learning,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 1999.

[36] Y. LeCun *et al.*, “Backpropagation Applied to Handwritten Zip Code Recognition,” *Neural Comput.*, 1989.

[37] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015.

[38] C. Dong, C. C. Loy, K. He, and X. Tang, “Image Super-Resolution Using Deep Convolutional Networks,” *IEEE Trans. Pattern Anal. Mach. Intell.*, 2016.

[39] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Commun. ACM*, 2017.

[40] M. Zeng *et al.*, “Convolutional Neural Networks for human activity recognition using mobile sensors Article,” pp. 381–388, 2014.

[41] S. Ha and S. Choi, “Convolutional neural networks for human activity recognition using multiple accelerometer and gyroscope sensors,” in *Proceedings of the International Joint Conference on Neural Networks*, 2016.

[42] J. J. Hopfield, “Neural networks and physical systems with emergent collective computational abilities (associative memory/parallel processing/categorization/content-addressable memory/fail-soft devices),” *Biophysics (Oxf).*, 1982.

[43] R. Pascanu, T. Mikolov, and Y. Bengio, “Understanding the exploding gradient problem,” *Proc. 30th Int. Conf. Mach. Learn.*, 2012.

[44] Y. Guan and T. Plötz, “Ensembles of Deep LSTM Learners for Activity Recognition using Wearables,” *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, 2017.

[45] F. Hernández, L. F. Suárez, J. Villamizar, and M. Altuve, “Human Activity Recognition on Smartphones Using a Bidirectional LSTM Network,” *2019 22nd Symp. Image, Signal Process. Artif. Vision, STSIVA 2019 - Conf. Proc.*, 2019.

[46] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Comput.*, 1997.

[47] J. Donahue *et al.*, “Long-term recurrent convolutional networks for visual recognition and description,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2015.

1. \*Corresponding author. E-mail: editorial@iospress.nl. Check if the checkbox in menu *Tools/Options/Compatibility/Lay out footnotes like Word 6.x/95/97* is selected if you make a footnote for the corresponding author. [↑](#footnote-ref-1)