Efficient Smartphone-based Human Activity Recognition using Convolutional Neural Network

Surya Dhanraj
Computer Science & Engineering Dept.
National Institute of Technology Patna
Patna, Bihar 800005, India
dhanraj.it18@nitp.ac.in

Suddhasil De
Computer Science & Engineering Dept.
National Institute of Technology Patna
Patna, Bihar 800005, India
suddhasil.de@acm.org

Dinesh Dash
Computer Science & Engineering Dept.
National Institute of Technology Patna
Patna, Bihar 800005, India
dd@nitp.ac.in

***Abstract*—Human activity recognition primarily targets the identification of individual/group activities from acquired data, which can be captured commonly using smartphone based sensory units. Existing sensor based human activity recognition, including smartphone based activity recognition, has suffered from efficient performances to accurately recognize physical activities. In this paper, an efficient human activity recognition system, based on the new simplified convolutional neural network framework, is proposed. The proposed system also improves the overall recognition accuracy of human activities given as one dimensional time series dataset based on accelerometers and gyroscopes of smartphones. In particular, the proposed system accurately recognizes six types of physical activities, viz. walk, walk-upstairs, walk-downstairs, sit, stand and lay. The proposed system has been thoroughly experimented to show that its overall recognition accuracy increases to as high as 93.926%, with significant improvement in training time and testing time performances, which reach to a minimum of 3.4274 seconds and 372.6 milliseconds respectively. Consequently, the proposed system achieves time bound performances of accurate activity recognition, thereby advocating many societal benefits including disaster management.**

***Index Terms*—Human activity recognition, convolution, neural network, smartphone, disaster management.**

I. Introduction

Human activity recognition is an active research domain [1]–[5], with the primary objective of asserting individual/group activities from acquired data. It plays a crucial role in various applications of disaster management [6]–[11], thus becoming an aide in reducing the severity of calamity and minimizing the panicky situation. It also plays significant roles in healthcare applications [12], [13], home-care applications and smart environments [14], security applications [15], video monitoring systems [16], [17], human-computer interaction and robotics for human behavior characterization [18], [19]. Human activity recognition provides information about individual identities, their psyche, and their mental/emotional states. It focuses on identifying the physical activities of one/more individuals from a set of observations gathered by different techniques. In this process, individual behaviours are observed and analyzed in their environments to conclude the activities.

Human activity recognition problem is inherently nondeterministic in nature. It is grouped into two categories based on the nature of sensor data employed [20] —
1) Unimodal activity recognition: such methods represent human activities from data of single modality (like images). They are further categorized as: (i) space-time methods, where human activities are represented as spatio-temporal features, (ii) stochastic methods, in which activities are recognized by statistical models, (iii) rule-based methods, which use set of rules to describe human activities, and (iv) shape based methods, where activities are represented with high level reasoning by modeling human body part movements).
2) Multimodal activity recognition: such methods combine features collected from various sources of different modalities. They are again categorized as: (i) affective methods, where human activities are represented according to human emotional states, (ii) behavioral methods, which aim to recognize non-verbal multi-modal cues, and (iii) social-networking methods, in which individual characteristics and behaviors are modeled in layers of human-to-human interactions in social events.

Human activity recognition system has conventionally used computer vision based mechanisms [21]–[25] for activity detection of individuals/groups by image processing techniques. Recognition of individual/group activities in these mechanisms depend wholly on their voluntary interactions with capturing devices. Examples of vision based activity recognition systems include intelligent video surveillance [26], smart homes [14], [27] etc. But such mechanisms mostly require infrastructural support, for example, installation of cameras in predetermined positions around specific monitoring areas. Additionally, vision based techniques produce complications regarding system complexity, privacy, and pervasiveness [3]. These shortcomings can be overcome in activity recognition systems using wearable sensors.

Sensor based human activity recognition incorporate sensor environments with data mining and machine learning techniques to model vast diversity of human activities [1], [3], [11], [12]. Most common activities analyzed relate to some form of human movement (involving sensors like accelerometers, gyroscopes, GPS etc.), or some parameters of the environment

This work is supported by the Science and Engineering Research Board, a statutory body of the Department of Science and Technology (DST), Government of India, under grants number ECR/2016/002040. Suddhasil De is the corresponding author.

978-1-7281-6052-8/19/$31.00 ©2019 IEEE
DOI 10.1109/ICIT48102.2019.00061

(involving sensors like thermistors, hygrometers, barometers etc.), or some physiological signals (involving sensors like heartbeat sensor, electrocardiographic sensors etc.). These sensors involved in different activity recognition are popularly available in different forms of handheld devices (like smartphones [28]), but can also be used individually. So smartphone based human activity recognition [29] has been investigated in literature.

In [30], smartphone based activity recognition using convolutional neural network (CNN) has been given for indoor localization. In this work, smartphone sensors are used to identify several types of individual activities. However, authors of this work have neither disclosed nor made public (till date) their dataset. In [31], a hierarchical approach (termed GCHAR) for smartphone based activity recognition has been given based on a preprocessed human activity recognition dataset [32], which is based on accelerometers and gyroscopes of smartphones and is made available in public domain. In [33], an experimental comparison of two popular machine learning algorithms (viz. Naive Bayes and K-nearest-neighbor) using accelerometers and gyroscopes of smartphones to identify basic individual activities with the help of an online activity recognizer. This work is also based on the same dataset in [32]. Smartphone based activity recognition with the same dataset in [32] using support vector machine has been given in [34], where accelerometric and gyroscopic measurements of individual activities are analyzed for daily activity monitoring of elderly people. In [35], [36], human activity recognition using deep CNN with the same dataset has been given. In [37], a hierarchical approach of recognizing immobile human activities has been presented using accelerometers and gyroscopes of smartphones. However, authors here also have neither disclosed nor made public (till date) their dataset. Human physical activities are also recognized using only accelerometers of users' smartphones in [38], [39]. In these classification works, the human activity recognition system has suffered from time efficient performances (particularly, training time and testing time) in order to accurately identify human activity. This paper addresses time efficient methodology for accurate human activity recognition.

In this paper, an efficient human activity recognition system (EHARS) is proposed, which also improves the overall recognition accuracy of human activities. The methodology of the EHARS is based on a new simplified CNN architectural framework for handling one dimensional time series data. The data is gathered using two inertial sensors of smartphones, viz. accelerometer and gyroscopes, when the physical activities are performed. The EHARS is proposed to accurately recognize six types of individual activities, viz. walk, walk-upstairs, walk-downstairs, sit, stand and lay. All activities that the EHARS considers belong to the ambulation group. The EHARS has been thoroughly experimented with the UCI's HARUS dataset[1], and the experimental results show that the EHARS enhances the overall recognition accuracy to even 93.926%, with significant improvement in training time and testing time performance reaching to as minimum as 3.4274 seconds and 372.6 milliseconds respectively in several experiments. Thus, the EHARS is used to fulfill time bound requirements of many applications for physical activity recognition, which leads to a plethora of applicability areas for societal benefits.

The rest of the paper is organized as follows. Section II provides the literature background and related works for this paper. Section III presents the proposed methodology, wherein the related activity characteristics, the proposed architecture and the approaches for training are explained. Section IV describes the detailed experimentation carried out to thoroughly evaluate the proposed methodology, while Section V concludes the paper.

II. Related Works

With new advancements in machine learning during the last tenner, human activity recognition problem has come to prominence. Applicability of machine learning in solving the activity recognition problem produces continuous improvements. Recent studies on human activity recognition [1]–[3] show the use of different types of approaches to acquire data related to physical activities in specific environments. Collection, modelling, reasoning, and distribution of context of acquired data play a critical role in this scenario.

Computer vision based human activity recognition mechanisms [14], [21]–[27], [40] use cameras, which are typical ambient sensors fastened in designated positions (such as, entry/exit of different locations), for detection of individual/group activities from images/videos by different image processing techniques. Video monitoring for activity recognition has various applications in scrutiny, monitoring of unwanted radicals and providing security assistance. With more advancements of technology, a depth camera can capture the depth information of an image in real time, which has inspired research on activity recognition from 3D data [41], [42]. Human depth silhouettes (i.e. outlines) based research is aimed to solve human activity recognition using spatio-temporal human body information [43]. However, activity recognition in all such mechanisms depend completely on voluntary interaction of individuals/groups with cameras. These requirements result in the drawbacks of vision based mechanisms, which impose installation of predetermined and fixed infrastructures at specific locations, with high deployment and maintenance costs.

Various constraints of vision based activity recognition have motivated the use of sensors for human activity recognition [1], [3], [11], [12]. Compared to vision based methods, sensors are not limited by coverage area, and they can be used for capturing people's daily activities continuously as they are carried along with end-users wherever they go. These sensors can record users' physiological states such as location changes, moving directions, speed, etc. They are popularly available in different forms of handheld devices (like smartphones [28]),

[1]The HARUS dataset [32] is available at https://archive.ics.uci.edu/ml/datasets/human+activity+recognition+using+smartphones.

but can also be used individually as wearable sensors [44]. A Body Sensor Network is formed by a group of wearable sensors, which are attached to an individual human body in order to acquire anatomical data [45]. Such networks are not only used for human activity recognition [46], but also in many other application domains (like, health-care, fitness, smart cities, Internet of Things etc.). However, with the advance in sensing technology, today's smartphones are equipped with multiple sensors and have become a popular source for sensors-based approach of human activity recognition in literature [29], [47], [48]. Human activity recognition using smartphone sensors are used in many advanced applications and interdisciplinary fields, such as health and fitness care system, biometric system, adaptive and rehabilitation technology for elderly care, indoor/outdoor localization and navigation setup etc. Due to its unsophisticated and economical installation cost, and elementary nature, handheld smartphones are fetching the main manifesto for human activity recognition.

Various machine learning algorithms like CNN [30], [35], [36], [38], long short-term memory (LSTM) network [49], Naive Bayes and K-nearest-neighbor [33], Support Vector Machine (SVM) [34], [39], [50], [51], hierarchical approach [31], [37], [39] etc. have been used to build human activity recognition system. In [30], CNN has been used to identify several individual activities based on their own dataset. In [49], human activity recognition has been carried out using a deep bidirectional LSTM network based on the HARUS dataset. In [31], a hierarchical approach for smartphone based activity recognition has been given based on the HARUS dataset. In [50], smartphone inertial sensor based recognition approach in a smartly controlled environment has been given, where kernel principal component analysis and linear discriminant analysis has been combined with deep belief network and multiclass SVM for activity training and recognition based on own dataset. In [35], [36], human activity recognition using deep CNN with the HARUS dataset has been given. In [33], comparative study of Naive Bayes and K-nearest-neighbor algorithms have been given to identify basic activities from smartphone sensors through online activity recognizer based on the HARUS dataset. In [51], a transition aware recognition system has been used for real time classification (using SVM) of physical activities based on data collected from inertial sensors of smartphones based the HARUS dataset along with 2 more datasets. In [37], hierarchical recognition with the transition identification and context awareness of immobile human activities, as well as decision tree has been given based on their own dataset. Smartphone based activity recognition with the HARUS dataset using SVM has been given in [34] for daily activity monitoring of elderly people. Human physical activities are also recognized based on their respective datasets using only accelerometers of users' smartphones in [38], [39]. In [38], CNN has been used, while in [39], random forest, SVM and other classification techniques are considered hierarchically.

In these classification works, the human activity recognition system has suffered from time efficient performances (particularly, training time and testing time) in order to accurately identify human activity, which has been the focus of the proposed EHARS model in this paper.

III. Proposed Methodology

This section introduces and elaborates the proposed EHARS for better recognition of physical activities. In particular, definition of the human activity recognition problem, the complete architecture of the EHARS, and essential training procedures of the EHARS are explained.

A. Problem statement

Definition 1 (Human Activity Recognition Problem): Given a set $\mathbf{F} = \{F_0, F_1, \ldots, F_{i-1}\}$ of i one dimensional time series data defined within a specific time interval $\mathbf{T} = [\mathcal{T}_\mu, \mathcal{T}_\nu]$, the objective is to identify — (i) set of consecutive and non-overlapping time intervals $\{T_k | T_k \neq \emptyset, k = 0, \ldots, j-1\}$ (where, $\bigcup_{k=0}^{j-1} T_k = \mathbf{T}$) by partitioning $\mathbf{T}$ based on $\mathbf{F}$, and (ii) set of labels representing activity during each interval T_k, where each element of $\mathbf{F}$ corresponds to a specific quantified attribute related to non co-occurrent activities.

The above definition indicates the goal of the human activity recognition problem to detect individual/group of physical activities of end-users based on a sequence of observations regarding human body movements. In this paper, different indoor or outdoor physical activities considered for recognition by the EHARS are given in Table I.

B. Proposed EHARS architecture

The objective of proposing the EHARS is to significantly improve both recognition accuracy and performance of human activities for one dimensional time series data gathered from in-built inertial sensors of smartphones. To this end, the methodology of the proposed EHARS incorporates a new architectural framework of convolutional neural networks, being structured from multiple sub-layers, viz. convolution, dropout, pooling, fully-connected, and output sub-layers.

The two consecutive convolution sub-layers in the EHARS handle temporal convolution operations (denoted as $*$) on one-dimensional input data vector $\mathbf{F}$ (belonging to one dimensional time series dataset) with kernel vector (which is also called filter vector) $\mathbf{G}$. The temporal convolution operation is defined as follows:

$$(\mathbf{F} * \mathbf{G})(t) = \int_{-\infty}^{\infty} \mathbf{F}(\tau)\mathbf{G}(t-\tau)d\tau \qquad (1)$$

TABLE I
Human activities for the EHARS

Activity	Description
walk	moving at regular pace
walk-upstairs	climbing up stairs
walk-downstairs	stepping down stairs
sit	being in seated position
stand	being in upright position
lay	lying down

In this definition, t is a time function, whereas τ is a dummy variable. Alternatively, Equation 1 can be defined as:

$$(\mathbf{F} * \mathbf{G})(t) = \int_{-\infty}^{\infty} \mathbf{F}(t-\tau)\mathbf{G}(\tau)d\tau \tag{2}$$

The pair of convolution sublayers is primarily responsible for applying kernel on an input dataset and extracting features based on temporal characteristics. For this reason, the kernel is also sometimes called the feature detector. For given $\mathbf{F}$ and $\mathbf{G}$ to be supported only in $[0, \infty)$, Equation 1 can be modified as:

$$(\mathbf{F} * \mathbf{G})(t) = \int_{0}^{\infty} \mathbf{F}(\tau)\mathbf{G}(t-\tau)d\tau \tag{3}$$

for $\mathbf{F}, \mathbf{G} : [0, \infty) \to \mathbb{R}$. Equation 3 can be written as:

$$(\mathbf{F} * \mathbf{G})(t) = \lim_{t\to\infty} \sum_{k=0}^{t} \mathbf{F}(\tau)\mathbf{G}(t-\tau) \tag{4}$$

where, t is sliding window size (or amount of time steps) applied on that dataset. A pair of same non-linear activation functions (defined as $\mathrm{ReLU}(x) = \max(0, x)$, where x represents the concerned input values) are also incorporated in both the convolution sub-layers to generate efficient mappings between input vector and output vector for processing at the next step.

The pooling sub-layer in the EHARS follows max-pooling approach of down sampling the output vector that has resulted from convolution sub-layers. This sub-layer is mainly responsible for steadily reducing spatial size of dataset representation (by eliminating several learned features) so as to decrease amount of parameters and computation in the EHARS, thereby integrating only the most essential units. To that end, a pool window is defined with specified size to limit the data values for max-pooling. Additionally, prior to pooling sub-layer, dropout sub-layer in the EHARS is applied as a regularization method to randomly eliminate some neural network input units in order to avoid the effects of overfitting. This sub-layer is pre-determined to slow down the learning rate of convolutional neural network (which is fast learning tool), and to optimize the outcome of the final model. The fully-connected sub-layer in the EHARS is responsible for the reasoning of applied activation functions on input vector through transformations via the process of flattening. Lastly, the output sub-layer in the EHARS allows applying of another non-linear activation function (defined as $\mathrm{Softmax}(y) = \frac{e^y}{\sum_{\forall y} e^y}$, where y represents the concerned input values) on the output of fully-connected sub-layer to produce the desired classification.

C. Training Approach

For the purpose of verifying performance and accuracy of the proposed EHARS, thorough experimentation has been conducted using the UCI's HARUS dataset based on human activity recognition collected from end-users' smartphones, which is available in the public domain. It is a preprocessed human activity recognition dataset, containing multivariate time series data. The dataset has included data of triaxial linear acceleration (gathered from an accelerometer) and triaxial angular velocity (gathered from a gyroscope), both of which are present as inertial sensors in today's smartphones. The data in the HARUS dataset is formatted in a three dimensional format: $\langle$samples, time steps, features$\rangle$, where one sequence is one sample and a batch consists of one or more samples. Moreover, one time step is one point of observation in a sample and one feature is one observation at a time step.

In the EHARS, the training strategy involves the configuration of different hyperparameters for setting the convolution sublayer. In particular, the EHARS methodology has configured the following hyperparameters in both the convolution sublayers: 64 parallel feature maps (also called activation maps), kernel size of 5, pooling size of 2, and batch size of 32 (agreeable batch sizes are usually in power of 2). The feature maps indicate the number of times the input vector is organized or interpreted. For appropriate training, the EHARS is performed using a neural network optimization algorithm (namely, the effective Adam version of stochastic gradient descent) to update network weights frequently based on available training data. The categorical cross entropy loss function is used as multi-class classification approach while training the EHARS for the given number of activity classes mentioned for each sample. Based on the EHARS configuration, its training and testing has been rigorously experimented.

IV. Results and Discussion

For rigorous experimentation of the proposed EHARS, the HARUS dataset has been chosen from the public domain. The HARUS dataset is already in preprocessed form, which involves removal of noise, as well as extraction of frequency domain and time domain feature values. The HARUS dataset collection employed 30 participants of age ranges from 19-48 years carrying Samsung Galaxy SII smartphones at their waist during data gathering. The sensor data are noted for 6 physical activities (viz. walk, walk-upstairs, walk-downstairs, sit, stand and lay) using in-built triaxial accelerometer and triaxial gyroscope at a persistent sampling rate of 50 Hz. Accelerometer generates data of triaxial linear acceleration, while gyroscope generates triaxial angular velocity.

The EHARS experimentations focus on its comprehensive evaluation of performances in comparison to several popular classification algorithms (viz. Random Tree, Bagging, Bayes Net, SVM and Decision Table). These experiments are conducted MS Windows Operating System based computer with 8GB RAM and 3.3 GHz processor. Accuracy, training time and testing time metrics are used to quantitatively denote the performances of the EHARS for comparison with these classification algorithms. Accuracy is the most important metric to exhibit the overall classification performances. It is the ratio of the number of correct predictions to the total number of input samples, i.e.

$$\text{Accuracy} = \frac{\text{number of correct predictions}}{\text{total number of input samples}} \tag{5}$$

Figure 1 has compared the classification accuracy of the EHARS with these classification methods. As shown in Fig-

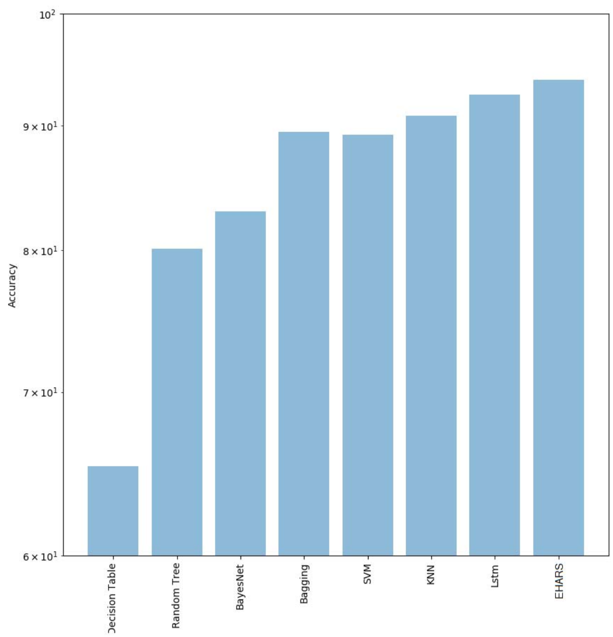

Fig. 1. Comparison of accuracy of the EHARS with other classification algorithms.

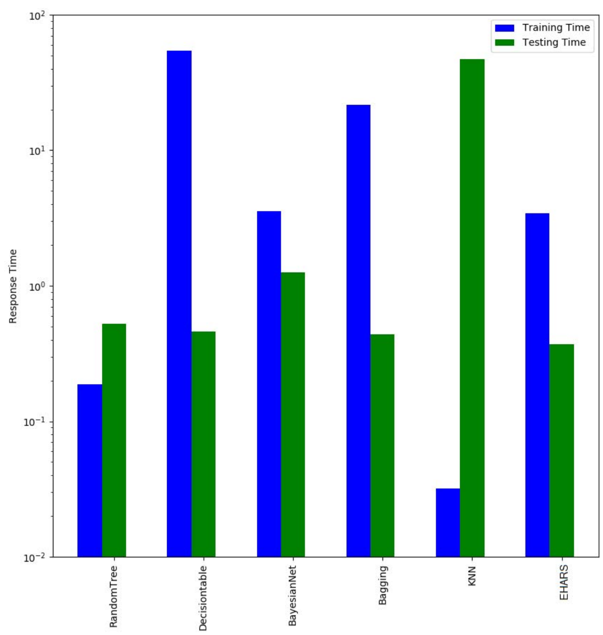

Fig. 2. Comparison of training and testing performances of the EHARS with other classification algorithms.

ure 1, the EHARS obtains much improved overall accuracy, reaching 93.926%. LSTM achieves second accurate accuracy with 92.67%. The other classification algorithms — Random Tree, Bagging, Bayes Net, Decision Table and SVM achieve 80.1154%, 89.4469%, 82.9997%, 65.2867% and 89.23% respectively. Figure 2 shows the time consumed by the EHARS during the training stage and the testing stage, in comparison to other classifiers' training and testing times. As shown in Figure 2, the testing time of the EHARS is minimum in comparison to others, while the training time is comparable to others.

V. Conclusion

The EHARS has been shown as an efficient human activity recognition system with improved overall recognition accuracy, which has incorporated a new simplified CNN architectural framework. The EHARS has been thoroughly experimented with the HARUS one dimensional time series dataset based on accelerometers and gyroscopes of smartphones from public domain. The comparison results have shown the accuracy improvement of the EHARS over the popular activity classification algorithms, as well as much less training time and testing time for producing the outcome in relation to these algorithms. These results have indicated the applicability of the EHARS for real time performances in the accurate activity recognition scenarios having many societal benefits including disaster management.

References

[1] Jindong Wang, Yiqiang Chen, Shuji Hao, Xiaohui Peng, and Lisha Hu. Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters*, 119:3–11, 2019.

[2] Sreenivasan Ramasamy Ramamurthy and Nirmalya Roy. Recent trends in machine learning for human activity recognition—a survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4), 2018.

[3] Óscar D. Lara and Miguel A. Labrador. A survey on human activity recognition using wearable sensors. *IEEE communications Surveys & Tutorials*, 15(3):1192–1209, 2013.

[4] Eunju Kim, Sumi Helal, and Diane Cook. Human activity recognition and pattern discovery. *IEEE Pervasive Computing*, 9(1):48–53, 2010.

[5] Pavan Turaga, Rama Chellappa, Venkatramana S Subrahmanian, and Octavian Udrea. Machine recognition of human activities: A survey. *IEEE Transactions on Circuits and Systems for Video technology*, 18(11):1473–1488, 2008.

[6] Fatai Idowu Sadiq, Ali Selamat, and Roliana Ibrahim. A systematic literature review on activity recognition with context-awareness techniques for mitigation of disasters. *International Journal of Digital Enterprise Technology*, 1(1-2):177–217, 2018.

[7] Patrick Lieser, Alaa Alhamoud, Hosam Nima, Björn Richerzhagen, Sanja Huhle, Doreen Böhnstedt, and Ralf Steinmetz. Situation detection based on activity recognition in disaster scenarios. In *ISCRAM*, 2018.

[8] Shingo Nakajima, Toshiki Yamasaki, Koki Matsumoto, Kazuki Uemura, Tomotaka Wada, and Kazuhiro Ohtsuki. Behavior recognition and disaster detection by the abnormal analysis using svm for eress. In *2018 International Conference on Information Networking (ICOIN)*, pages 646–651. IEEE, 2018.

[9] Fatai Idowu Sadiq, Ali Selamat, and Roliana Ibrahim. Human activity recognition prediction for crowd disaster mitigation. In *Asian Conference on Intelligent Information and Database Systems*, pages 200–210. Springer, 2015.

[10] Fatai Idowu Sadiq, Ali Selamat, Roliana Ibrahim, et al. Performance evaluation of classifiers on activity recognition for disasters mitigation using smartphone sensing. *Jurnal Teknologi*, 77(13), 2015.

[11] Weiwei Yuan, Donghai Guan, Eui-Nam Huh, and Sungyoung Lee. Harness human sensor networks for situational awareness in disaster reliefs: a survey. *IETE Technical Review*, 30(3):240–247, 2013.

[12] Yan Wang, Shuang Cang, and Hongnian Yu. A survey on wearable sensor modality centred human activity recognition in health care. *Expert Systems with Applications*, 137:167–190, 2019.

[13] Henry Friday Nweke, Ying Wah Teh, Ghulam Mujtaba, and Mohammed Ali Al-Garadi. Data fusion and multiple classifier systems for human activity detection and health monitoring: Review and open research directions. *Information Fusion*, 46:147–170, 2019.

[14] U. A. B. U. A. Bakar, Hemant Ghayvat, S. F. Hasanm, and Subhas Chandra Mukhopadhyay. Activity and anomaly detection in smart home: A survey. In Subhas Chandra Mukhopadhyay, editor, *Next Generation Sensors and Systems*, pages 191–220. Springer, 2016.

[15] Jessamyn Dahmen, Brian Thomas, Diane Cook, and Xiaobo Wang. Activity learning as a foundation for security monitoring in smart homes. *Sensors*, 17(4), 2017.

[16] Karishma Pawar and Vahida Attar. Deep learning approaches for video-based anomalous activity detection. *World Wide Web*, 22(2):571–601, 2019.

[17] Andrea Prati, Caifeng Shan, and Kevin I-Kai Wang. Sensors, vision and networks: From video surveillance to activity recognition and health monitoring. *Journal of Ambient Intelligence and Smart Environments*, 11(1):5–22, 2019.

[18] Jianguo Hao, Abdenour Bouzouane, Bruno Bouchard, and Sébastien Gaboury. Activity inference engine for real-time cognitive assistance in smart environments. *Journal of Ambient Intelligence and Humanized Computing*, 9(3):679–698, 2018.

[19] Rajesh Kumar Tripathi, Anand Singh Jalal, and Subhash Chand Agrawal. Suspicious human activity recognition: a review. *Artificial Intelligence Review*, 50(2):283–339, 2018.

[20] He Li, Kaoru Ota, Mianxiong Dong, and Minyi Guo. Learning human activities through wi-fi channel state information with multiple access points. *IEEE Communications Magazine*, 56(5):124–129, 2018.

[21] Tej Singh and Dinesh Kumar Vishwakarma. Human activity recognition in video benchmarks: A survey. In *Advances in Signal Processing and Communication*, pages 247–259. Springer, 2019.

[22] Leonardo Onofri, Paolo Soda, Mykola Pechenizkiy, and Giulio Iannello. A survey on using domain and contextual knowledge for human activity recognition in video streams. *Expert Systems with Applications*, 63:97–111, 2016.

[23] Alejandro Betancourt, Pietro Morerio, Carlo S Regazzoni, and Matthias Rauterberg. The evolution of first person vision methods: A survey. *IEEE Transactions on Circuits and Systems for Video Technology*, 25(5):744–760, 2015.

[24] Jose M Chaquet, Enrique J Carmona, and Antonio Fernández-Caballero. A survey of video datasets for human action and activity recognition. *Computer Vision and Image Understanding*, 117(6):633–659, 2013.

[25] Jake K Aggarwal and Michael S Ryoo. Human activity analysis: A review. *ACM Computing Surveys (CSUR)*, 43(3):16, 2011.

[26] Amira Ben Mabrouk and Ezzeddine Zagrouba. Abnormal behavior recognition for intelligent video surveillance systems: A review. *Expert Systems with Applications*, 91:480–491, 2018.

[27] Asma Benmansour, Abdelhamid Bouchachia, and Mohammed Feham. Multioccupant activity recognition in pervasive smart home environments. *ACM Computing Surveys*, 48(3), 2016.

[28] Özgür Yürür, Chi Harold Liu, Zhengguo Sheng, Victor CM Leung, Wilfrido Moreno, and Kin K Leung. Context-awareness for mobile sensing: A survey and future directions. *IEEE Communications Surveys & Tutorials*, 18(1):68–93, 2014.

[29] Jafet Morales and David Akopian. Physical activity recognition by smartphones, a survey. *Biocybernetics and Biomedical Engineering*, 37(3):388–400, 2017.

[30] Baoding Zhou, Jun Yang, and Qingquan Li. Smartphone-based activity recognition for indoor localization using a convolutional neural network. *Sensors*, 19(3), 2019.

[31] Liang Cao, Yufeng Wang, Bo Zhang, Qun Jin, and Athanasios V Vasilakos. Gchar: An efficient group-based contextaware human activity recognition on smartphone. *Journal of Parallel and Distributed Computing*, 118:67–80, 2018.

[32] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, and Jorge Luis Reyes-Ortiz. A public domain dataset for human activity recognition using smartphones. In *European Symposium on Artificial Neural Networks Computational Intelligence and Machine Learning*, 2013.

[33] Aiguo Wang, Guilin Chen, Jing Yang, Shenghui Zhao, and Chih-Yung Chang. A comparative study on human activity recognition using inertial sensors in a smartphone. *IEEE Sensors Journal*, 16(11):4566–4578, 2016.

[34] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, and Jorge L Reyes-Ortiz. Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine. In *Ambient Assisted Living and Home Care*, pages 216–223. Springer, 2012.

[35] Charissa Ann Ronao and Sung-Bae Cho. Human activity recognition with smartphone sensors using deep learning neural networks. *Expert systems with Applications*, 59:235–244, 2016.

[36] Wenchao Jiang and Zhaozheng Yin. Human activity recognition using wearable sensors by deep convolutional neural networks. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 1307–1310. ACM, 2015.

[37] Nicole A Capela, Edward D Lemaire, and Natalie Baddour. Improving classification of sit, stand, and lie in a smartphone human activity recognition system. In *2015 IEEE International Symposium on Medical Measurements and Applications (MeMeA) Proceedings*, pages 473–478. IEEE, 2015.

[38] Andrey Ignatov. Real-time human activity recognition from accelerometer data using convolutional neural networks. *Applied Soft Computing*, 62:915–922, 2018.

[39] Akram Bayat, Marc Pomplun, and Duc A Tran. A study on human activity recognition using accelerometer data from smartphones. *Procedia Computer Science*, 34:450–457, 2014.

[40] Ronald Poppe. A survey on vision-based human action recognition. *Image and vision computing*, 28(6):976–990, 2010.

[41] Jake K Aggarwal and Lu Xia. Human activity recognition from 3d data: A review. *Pattern Recognition Letters*, 48:70–80, 2014.

[42] Shuiwang Ji, Wei Xu, Ming Yang, and Kai Yu. 3d convolutional neural networks for human action recognition. *IEEE transactions on pattern analysis and machine intelligence*, 35(1):221–231, 2012.

[43] Ahmad Jalal, Yeon-Ho Kim, Yong-Joong Kim, Shaharyar Kamal, and Daijin Kim. Robust human activity recognition from depth video using spatiotemporal multi-fused features. *Pattern recognition*, 61:295–308, 2017.

[44] Arsalan Mosenia, Susmita Sur-Kolay, Anand Raghunathan, and Niraj K Jha. Wearable medical sensor-based system design: A survey. *IEEE Transactions on Multi-Scale Computing Systems*, 3(2):124–138, 2017.

[45] Raffaele Gravina, Parastoo Alinia, Hassan Ghasemzadeh, and Giancarlo Fortino. Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges. *Information Fusion*, 35:68–80, 2017.

[46] Andreas Bulling, Ulf Blanke, and Bernt Schiele. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys (CSUR)*, 46(3):33, 2014.

[47] Ozlem Durmaz Incel, Mustafa Kose, and Cem Ersoy. A review and taxonomy of activity recognition on mobile phones. *BioNanoScience*, 3(2):145–171, 2013.

[48] Tanzeem Choudhury, Gaetano Borriello, Sunny Consolvo, Dirk Haehnel, Beverly Harrison, Bruce Hemingway, Jeffrey Hightower, Karl Koscher, Anthony LaMarca, James A Landay, et al. The mobile sensing platform: An embedded activity recognition system. *IEEE Pervasive Computing*, 7(2):32–41, 2008.

[49] Fabio Hernández, Luis F Suárez, Javier Villamizar, and Miguel Altuve. Human activity recognition on smartphones using a bidirectional lstm network. In *2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA)*, pages 1–5. IEEE, 2019.

[50] Mohammed Mehedi Hassan, Md Zia Uddin, Amr Mohamed, and Ahmad Almogren. A robust human activity recognition system using smartphone sensors and deep learning. *Future Generation Computer Systems*, 81:307–313, 2018.

[51] Jorge-L Reyes-Ortiz, Luca Oneto, Albert Samà, Xavier Parra, and Davide Anguita. Transition-aware human activity recognition using smartphones. *Neurocomputing*, 171:754–767, 2016.