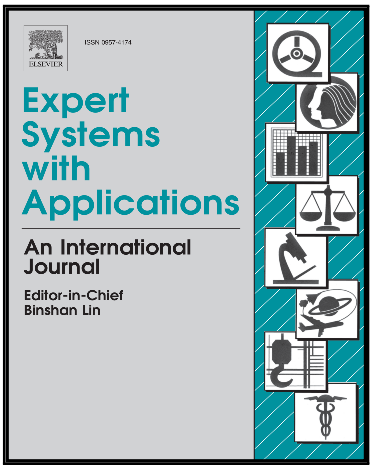
Accepted Manuscript



Deep Learning Algorithms for Human Activity Recognition using Mobile and Wearable Sensor Networks: State of the Art and Research Challenges

Henry Friday Nweke , Ying Wah Teh , Mohammed Ali Al-garadi , Uzoma Rita Alo

PII: S0957-4174(18)30213-6
DOI: 10.1016/j.eswa.2018.03.056
Reference: ESWA 11903

To appear in: *Expert Systems With Applications*

Received date: 29 May 2017
Revised date: 26 March 2018
Accepted date: 27 March 2018

Please cite this article as: Henry Friday Nweke , Ying Wah Teh , Mohammed Ali Al-garadi , Uzoma Rita Alo , Deep Learning Algorithms for Human Activity Recognition using Mobile and Wearable Sensor Networks: State of the Art and Research Challenges, *Expert Systems With Applications* (2018), doi: 10.1016/j.eswa.2018.03.056

Highlights

- Review of deep learning methods for sensor based human activity recognition
- Categorize the studies into generative, discriminative and hybrid methods
- Present training, evaluation procedures and Common datasets
- Outline open research issues presented as future directions

Deep Learning Algorithms for Human Activity Recognition using Mobile and Wearable Sensor Networks: State of the Art and Research Challenges

Henry Friday Nweke[a,b], Ying Wah Teh[a,], Mohammed Ali Al-garadi[a], Uzoma Rita Alo[b]*

[a] *Department of Information Systems, Faculty of Computer Science and Information Technology, University of Malaya, 50603, Kuala Lumpur, Malaysia.*
[b] *Computer Science Department, Ebonyi State University, P.M.B 053, Abakaliki, Ebonyi State, Nigeria*
Email addresses: *henrynweke@siswa.um.edu.my* (H.F. Nweke), *tehyw@um.edu.my* (Y.W. Teh), mohammedali@siswa.um.edu.my (M.A. Al-garadi), auzomarita@yahoo.com (U.R.Alo)
*Corresponding Author: *tehyw@um.edu.my* (Y.W. Teh)

Abstract

Human activity recognition systems are developed as part of a framework to enable continuous monitoring of human behaviours in the area of ambient assisted living, sports injury detection, elderly care, rehabilitation, and entertainment and surveillance in smart home environments. The extraction of relevant features is the most challenging part of the mobile and wearable sensor-based human activity recognition pipeline. Feature extraction influences the algorithm performance and reduces computation time and complexity. However, current human activity recognition relies on handcrafted features that are incapable of handling complex activities especially with the current influx of multimodal and high dimensional sensor data. With the emergence of deep learning and increased computation powers, deep learning and artificial intelligence methods are being adopted for automatic feature learning in diverse areas like health, image classification, and recently, for feature extraction and classification of simple and complex human activity recognition in mobile and wearable sensors. Furthermore, the fusion of mobile or wearable sensors and deep learning methods for feature learning provide diversity, offers higher generalisation, and tackles challenging issues in human activity recognition. The focus of this review is to provide in-depth summaries of deep learning methods for mobile and wearable sensor-based human activity recognition. The review presents the methods, uniqueness, advantages and their limitations. We not only categorise the studies into generative, discriminative and hybrid methods but also highlight their important advantages. Furthermore, the review presents classification and evaluation procedures and discusses publicly available datasets for mobile sensor human activity recognition. Finally, we outline and explain some challenges to open research problems that require further research and improvements.

Keywords: Deep Learning, Mobile and Wearable sensors, Human Activity Recognition, Feature Representation, Review

1. Introduction

Human activity recognition is an important area of research in ubiquitous computing, human behaviour analysis and human-computer interaction. Research in these areas employ different machine learning algorithms to recognise simple and complex activities such as walking, running, cooking, etc. Particularly, recognition of daily activities is essential for maintaining healthy lifestyle, patient rehabilitation and activity shifts among the elderly citizens that can help to detect and diagnose serious illnesses. Therefore, human activity recognition framework provides mechanism to detect both postural and ambulatory activities, body movements and actions of users using different multimodal data generated by variety of sensors(Cao, Wang, Zhang, Jin, & Vasilakos, 2017; Ordonez & Roggen, 2016). Previous studies in human activity recognition can be broadly categorised based on diverse devices, sensor modalities and data utilised for detection of activity details. These include video based, wearable and mobile phone sensors, social network sensors and wireless signals. Video-based sensors are utilised to capture images, video or surveillance camera features to recognise daily activity (Cichy, Khosla, Pantazis, Torralba, & Oliva, 2016; Onofri, Soda, Pechenizkiy, & Iannello, 2016). With the introduction of mobile phones and other wearable sensors, inertial sensor data (S. Bhattacharya & Lane, 2016; Bulling, Blanke, & Schiele, 2014b) are collected using mobile or wearable embedded sensors placed at different body positions in order to infer human activities details and transportation modes. Alternatively, the use of social network methods (Y. Jia, et al., 2016) that exploit appropriate users' information from multiple social network sources to understand user behaviour and interest have also been proposed recently. In addition, wireless signal based human activity recognition (Savazzi, Rampa, Vicentini, & Giussani, 2016) takes advantages of signal propagated by the wireless devices to categorise human activity. However, the use of sensor data generated using smartphones and other wearable devices have dominated the research landscape in human motion analysis, activity monitoring and detection due to their obvious advantages over other sensor modalities (Cornacchia, Ozcan, Zheng, & Velipasalar, 2017).

Generally, mobile phones and wearable based sensors for human activity identification are driven by their ubiquity, unobtrusiveness, cheap installation procedure and ease of usability. Mobile phones have become part of our daily life and can be found in every homes and carried everywhere we go. In this context, mobile phones and wearable sensors are popular alternative methods of inferring activity details. For instance, while the video sensor extract features such as the Histogram of Oriented Gradient (HOG), Spatio-temporal interest Point (STIP) and Region of Interest (ROI), mobile sensors utilise statistical and frequency based features to recognise activity details. Statistical features provide less computation time and complexity(Figo, Diniz, Ferreira, & Cardoso, 2010). Furthermore, vision based techniques intrude on user privacy, require fixed location implementations and capture non-target information (Yang, Nguyen, San, Li, & Krishnaswamy, 2015). In addition, video sensors based human activity recognition are affected by lighting variability leading to decrease in performances due to visual disturbanes(Lukun Wang, 2016). On the other hand, mobile and wearable sensor-based methods provide better advantages for real-time implementation of human activity recognition systems. Moreover, mobile phone and wearable devices are not location dependents, cost effective, easy to deploy and do not pose any health hazard caused by radiation (Alsheikh, et al., 2015) unlike wireless signals based method. Considering the obvious advantages of mobile and wearable sensor based implementation of human activity, number of studies have been proposed by leveraging on the data generated using these devices(J. Morales & Akopian, 2017).

The explosion of smartphones era embedded with multi-sensor systems that enable researchers to collect human physiological signal for monitoring of activity of daily living, have made human motion analysis integral part of our daily life. Smartphones provide access to wide range of sensor such as accelerometer, gyroscope, magnetometer, Bluetooth, Wi-Fi, microphones, proximity and light sensor and cellular radio sensors that can be exploited to infer activity details. Sensors such as accelerometer, gyroscope, magnetometer, heart rate, GPS can be deployed for coarse grain and context activity recognition, user location and social interaction between users. Motion sensors (Accelerometers, gyroscope magnetometer) provide important information that facilitate recognition and monitoring of users' movement such as walking, standing or running. Similarly, proximity and light sensors embedded in mobile devices to enhance user experiences can also be deployed to determine whether the user is in light or dark places(O. Incel, 2015). Other sensors such as barometers, thermometers, air humidity and pedometers have also been applied to maintain healthy status of elderly citizens and for assisted living(J. Gong, Cui, Xiao, & Wang, 2012). For instance, the pedometer found in the Samsung Galaxy smartphones and exercises tracking wearable devices are essential for step counts, heart rate and pulse monitoring. These are effective for important health conditions identifications which may interfere with user activities (Kanaris, Kokkinis, Liotta, & Stavrou, 2017; Natarajasivan & Govindarajan, 2016; Zouba, Bremond, & Thonnat, 2009).

In human activity recognition, data collection with varieties of sensors installed in mobile phone and wearable devices is preceded by other data analytic phases such as pre-processing, data segmentation, extraction of salient and discriminative features, and finally classification of activity details. Pre-processing involves the removal and representation of the raw sensor data. Different methods such as nonlinear, low pass and high pass filter, and Laplacian and Gaussian filter have been utilised for pre-processing. The segmentation procedure divides the signal into different window sizes to extract useful features. Generally, sensor data segmentation is achieved using methods ranging from sliding windows, events or energy based activities(Bulling, Blanke, & Schiele, 2014a). Next, relevant feature vectors are extracted from the segmented data to determine lower set of features to minimise classification errors and reduce computation time. In addition, the extracted features are often further reduced through feature selection methods to the most discriminative features for recognition tasks. Feature vectors for human activity recognition can be broadly categorised into statistical and structural features (Bulling, et al., 2014a; Figo, Diniz, Ferreira, Jo, et al., 2010). Statistical features (mean, median, time domain, frequency domain, standard deviation, etc.) extract quantitative properties of sensor data while structural features use the relationship among the mobile sensor data for feature extraction. Likewise, dimensionality reduction reduces the dimension of the extracted features to decrease the computational time. The dimensionality reductions widely used in human activity recognition are principal component analysis (PCA), linear discriminate analysis (LDA) and empirical cumulative distribution functions (ECDF) (Abidine, Fergani, Fergani, & Oussalah, 2016). The activity recognition and classification phases help to map extracted features into sets of activities using machine learning or pattern recognition methods (Bulling, et al., 2014b). Large varieties of machine learning techniques have played prominent roles in inferring activity details. These include the Support Vector Machine (Anguita, Ghio, Oneto, Parra, & Reyes-Ortiz, 2012; Kim & Ling, 2009), Hidden Markov Model(Safi, Mohammed, Attal, Khalil, & Amirat, 2016), Decision Tree, K-Nearest Neighbour (KNN) (Shoaib, Bosch, Incel, Scholten, & Havinga, 2016) and Gaussian Mixture Model (Rodriguez, Orrite, Medrano, & Makris, 2016). Studies by (Bulling, et al., 2014b; O. D. Incel, Kose, & Ersoy, 2013; Pires, Garcia, Pombo, & Flórez-Revuelta, 2016) provide excellent information on the human activity recognition process using handcrafted features with mobile and wearable sensor data.

Recently, to overcome the challenges associated with single sensor modalities and increase generalization, many studies have proposed information fusion strategies that combine multiple sensors modalities or classifiers to increase robustness, reliabilities, derive confidence measures among different classifiers and reduce the complexity of recognition system (Pires, et al., 2016). Information fusion in human activity recognition are necessitated by increase in sensor of different modalities (Gravina, Alinia, Ghasemzadeh, & Fortino, 2017). Information fusion techniques are prevalent in both handcrafted features and automatic feature learning using deep learning (Habib, Makhoul, Darazi, & Couturier, 2016; Shoaib, Bosch, Incel, Scholten, & Havinga, 2014; C. Zhu & Sheng, 2009; Zouba, et al., 2009). In this review, recent works on information fusion for human activity recognition using automatic feature representation were also analysed.

Of all the different phases of human activity recognition framework, feature extraction is the most important stage(Domingos, 2012). This is because of the correlation between performances of activity recognition system and extraction of relevant and discriminative feature vectors. Therefore, extensive works have been done on how to improve human activity recognition system through extraction of expert-driven features (Figo, Diniz, Ferreira, Jo, et al., 2010). However, expert-driven features extraction methods depend on the knowledge of the experts or guess and applicability of the feature vectors in the problem domains. Even though, conventional handcrafted features learning methods are easy to understand and have been widely utilised for activity recognition, feature vectors extracted using such techniques are tasks or applications dependent, and cannot be transferred to similar activity tasks. Furthermore, hand-engineered features cannot represent the salient characteristics of complex activities, and involve time-consuming feature selection techniques to select the optimal features (Yang, et al., 2015). Also, there are no universal procedures for selecting appropriate features leading to many studies resort to heuristic means using feature engineering knowledge approach. In the nutshell, the major challenges of conventional handcrafted features for mobile and wearable sensor based human activity recognition are summarised below:

- Feature representation techniques in current human activity recognition approaches for mobile and wearable sensors use carefully engineered feature extraction and selections methods that are manually extracted using expert domain knowledge. However, such feature extraction approach are task or applications dependent and cannot be transferred to activity of similar patterns. Furthermore, carefully engineered features vectors are challenging to model complex activity details and involve time consuming feature selections(C. A. Ronao & S.-B. Cho, 2016; Yang, et al., 2015);
- There are no universal procedures for selecting appropriate features but many studies resort to extensive heuristic knowledge to develop and select appropriate tasks for a given human activity recognition system (Zdravevski, et al., 2017);
- Moreover, the current statistical features such as time or frequency domain features for human activity recognition are unable to model and support the dynamic nature of the current seamless and ubiquitous collection of mobile and wearable senor streams(M. Hasan & Roy-Chowdhury, 2015);
- Also, human activity recognition using expert driven features require large amount of labelled training sensor data to obtain accurate recognition performance. The experimental protocol to collect large amount of labelled training data require extensive infrastructural setup that are time consuming. On the contrary, unla-

belled data are easy to obtain leveraging Internet of Things (IoT), smart homes and mobile crowdsourcing from transportation modes (Song-Mi, Sang Min, & Heeryon, 2017);

- Other challenges of handcrafted features are the issues bothering on intra-class variability and inter-class similarities (Bulling, et al., 2014b). In this case, same activities may be performed differently by different individuals or different activities appear to have same pattern of executions. Developing generic expert driven features that can accurately model these issues are challenging;
- Furthermore, human activities are hierarchical and inherently translational in nature with ambiguity in temporal segmentation of sub-activities that constitute the main activity. Therefore, capturing spatial and temporal variation of activities are important for accurate detection of complex activity details (Kautz, et al., 2017);
- To achieve diversity and robust features for human activity recognition performance generalisation across heterogeneous domain, approaches such as multimodal fusion and decision fusion are utilised. However, there still exist, uncertainties on the best fusion techniques to achieve higher generalisation with reduced computation time for mobile and wearable sensor implementation.

To solve the above problems, studies have delved into techniques that involve automatic features extraction with less human efforts (LeCun, Bengio, & Hinton, 2015) using deep learning techniques. Deep learning, a new branch of machine learning that models high-level features in data, has become an important trend in human activity recognition. Deep learning comprises multiple layers of neural networks that represent features from low to high levels hierarchically. It has become a critical research area in image and object recognition, natural language processing, machine translation and environmental monitoring (Y. Guo, et al., 2016). More recently, various deep learning methods have been proposed for mobile and wearable sensor based human activity recognition. These methods include restricted Boltzmann machine, autoencoder, sparse coding, convolutional neural network and recurrent neural network. These deep learning methods can be stacked into different layers to form deep learning models that provide enhanced system performance, flexibility, robustness and remove the need to depend on conventional handcrafted features. The essence of this study is to review different human activity recognition and health monitoring systems in mobile and wearable sensors that utilise deep neural network for feature representations. We provide an extensive review of the recent developments in the field of human activity recognition for mobile and wearable sensors using deep learning. Specifically, we present comprehensive review of deep learning methods; taxonomy of the recent studies in deep learning based activity recognition, their advantages, training procedure and popular deep learning software frameworks. Based on the reviewed papers, open research issues were derived, and future research directions are suggested.

Deep learning and human activity recognition or activity of daily living as a separate research areas have been progressive areas for years. A good number of surveys and reviews have been published. However, these reviews either focus on deep learning and their applications or activity recognition using conventional features learning methods. Furthermore, these reviews have become outdated and require urgent research to analyse the high volume of papers published in the area lately. In deep learning methods, reviews by (Angermueller, Parnamaa, Parts, & Stegle, 2016; Benuwa, Zhan, Ghansah, Wornyo, & Kataka, 2016; Dolmans, Loyens, Marcq, & Gijbels, 2016; Gawehn, Hiss, & Schneider, 2016; LeCun, et al., 2015; W. Liu, Ma, Qi, Zhao, & Chen, 2017; W. Liu, et al., 2016; Mamoshina, Vieira, Putin, & Zhavoronkov, 2016; Ravì, Wong, Deligianni, et al., 2017; Schmidhuber, 2015) provide comprehensive knowledge of the development and historical perspective. While studies such as (Ahmad, Saeed, Saleem, & Kamboh, 2016; Attal, et al., 2015; Bulling, et al., 2014b; Cornacchia, et al., 2017; R. Gravina, et al., 2017; O. D. Incel, et al., 2013; Kumari, Mathew, & Syal, 2017; Onofri, et al., 2016; Pires, et al., 2016; Turaga, Chellappa, Subrahmanian, & Udrea, 2008) discussed the human activity and action recognition based on handcrafted features, sensor fusion techniques to increase the robustness of recognition algorithms and developmental trends on wearable sensors for the collection of activity data. Others presented the use of handcrafted and deep learning based features for human activity recognition in video sensor and images (Aggarwal & Xia, 2014; Sargano, Angelov, & Habib, 2017; X. Xu, et al., 2013; F. Zhu, Shao, Xie, & Fang, 2016). Recently, authors (Gamboa, 2017; Langkvist, Karlsson, & Loutfi, 2014) reviewed deep learning for time series analysis; another closely related area in human activity recognition. However, the author took a broader view on the applications of deep learning in time series that comprises speech recognition, sleep stage classification and anomaly detection but this review focused on deep learning based human activity recognition using sensor data generated by mobile or wearable devices. From the available literature, there are no studies on review or survey of deep learning based feature representation and extraction for mobile and wearable sensors based on human activity recognition. To fill this gap, this review is a timely exploration of the processes for developing deep learning based human activity recognition and provide in-depth tutorial on the techniques, implementation procedure and feature learning process.

The remainder of this paper is organised as follows: **Section 2** discusses Comparison of deep learning feature representation and conventional handcrafted feature learning approach. **Section 3** discusses the deep learning methods and their subdivisions. **Section 4** review different representative studies in deep learning for human activity recognition using mobile and wearable sensors. The section is subdivided into generative feature extraction techniques such as Deep Belief Network (DBN), Deep Boltzmann Machine (DBM), sparse coding, and discriminative feature extraction with Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) and hybrid methods that combine generative and discriminative deep learning methods. The description, advantages and weakness of these studies are also discussed in details. **Section 5** discusses the training procedure, classification and evaluation of deep learning for human activity recognition. **Section 6** reviews common benchmark datasets for human activity recognition using deep learning. **Section 7** includes the software frameworks for implementation of deep learning algorithms. **Section 8** provides the open research challenges requiring further improvements and attention while Section **9** concludes the review.

2. Comparison of deep learning Feature Representation and Conventional Feature Learning

Feature extraction is a vital part of the human activity recognition process as it helps to identify lower sets of features from input sensor data to minimise classification errors and computational complexity. Effective performance of Human activity recognition system depends on appropriate and efficient feature representation (Abidine, et al., 2016). Therefore, extraction of efficient feature vectors from mobile and wearable sensor data helps to reduce computation time and provide accurate recognition performance. Feature extraction can be performed manually or automatically based on expert knowledge. Manually engineered features follow bottom-up approaches that consist of data collection, signal pre-processing and segmentation, handcrafted features extraction and selection, and classification. Manually engineered feature processes utilise appropriate domain knowledge and expert-driven approach to extract time domain, frequency domain and Hulbert-Huang features using Empirical mode decomposition to represent signal details (Z. L. Wang, Wu, Chen, Ghoneim, & Hossain, 2016; Zdravevski, et al., 2017). Then, appropriate feature selection methods such as Minimal Redundancy Maximal Relevance, correlation based features selection method and RELIEF F are employed to reduce computation time and memory usage due to inability of mobile and wearable devices to support computational intensive applications (Bulling, et al., 2014b). Also, data dimensionality reduction approach such Principal Component analysis(PCA), Linear Discriminative analysis(LDA), Independent Component analysis(ICA) and Empirical Cumulative Distribution Function(ECDF) (Abidine, et al., 2016; Plötz, Hammerla, & Olivier, 2011) are utilised to further reduce features dimensionality and produce compact feature vectors representations.

However, it is very challenging to measure the efficient performances of manually engineered features across different applications and also require time consuming features selection and dimensionality reduction methods specified above to obtain acceptable results(X. Li, et al., 2017; C. A. Ronao & S.-B. Cho, 2016). Moreover, the use of feature selection are often arbitrary and lacks generalisability or ability to model complex activity details. It is highly acknowledged that activity in natural environments are abstracts, hierarchical and translational in nature with temporal and spatial information(X. Li, et al., 2017). In order to consider these mobile and wearable sensor data characteristics for human activity recognition, require intensive feature extraction and selection especially for continuous sensor streams (Ordóñez & Roggen, 2016). Another pertinent issues with handcrafted features are based on the dimensionality reduction commonly used. For instance, principal component analysis (PCA) treat each dimensionality as statistically independent and extract features based on sensor appearance, but activities are performed based on activity windows, and this have been found to affect recognition accuracy (Plötz, et al., 2011).

Clearly, there is need for appropriate techniques to extract discriminative features to achieve optimal performance accuracy. Recent studies in human activity recognition have observed there are no universally best discriminative feature that accurately represent across dataset and applications (Capela, Lemaire, & Baddour, 2015). Therefore, automatic feature representations are required to enable extraction of translational invariant feature vectors without reliance on domain expert knowledge. Deep learning methods for automatic feature representation provide the ability to learn features from raw sensor data with little pre-processing (LeCun, et al., 2015). Using multiple layer of abstraction, deep learning methods learn intricate features representation from raw sensor data and discover the best pattern to improve recognition performance. Recently, studies have indicated the incredible results of deep learning over conventional handcrafted features for human activity recognition (Ordóñez & Roggen, 2016; S. Yao, Hu, Zhao, Zhang, & Abdelzaher, 2017). Also, the use of automatic feature representation helps to capture local dependencies and scale invariants features. Thus, deep learning provide effective means to solve the problem of intra-class variabilities and inter-class similarities that are fundamental challenges for implementing human activity recognition with handcrafted features(Bulling, et al., 2014b). Furthermore, deep learning methods apply unsupervised pre-training to learn structure of high dimensional sensor data to prevent overfitting. With the current influx of unlabelled sensor streams from Internet of Things (IoT), crowdsourcing and cyber-physical systems, implementing efficient human activity recognition would be very challenging without automatic feature representation from raw sensor data (Raffaele Gravina, et al., 2017). In **Table 1**, we summarised the comparison of the two approaches in terms of strengths and weaknesses for mobile and wearable sensor based human activity recognition. The comparisons are summarised using five characteristics. These include feature representation method, generalisation, data preparation, changes in activity details and execution time.

Table 1: Comparison of Deep Learning feature Representation and Conventional feature Learning

Characteristics	Deep learning based feature representation	Conventional Feature Learning Approach
Feature extraction and Representation	Ability to learn features from raw sensors data and discover the most efficient patterns to improve recognition accuracy	Use manually engineered feature vectors that are applications dependent, and unable to model complex activity details
Generalisation and Diversity	Helps to automatically capture spatial, temporal dependencies and scale invariant features from unlabeled raw sensor data	Require labelled sensor data and use arbitrary feature selection, and dimensionality reduction approaches that are hardly generalizable
Data preparations	Data pre-processing and normalisation is not compulsory in deep learning features to obtain improved results	Extract features based on sensor appearance but activities are performed within activity windows. Furthermore, manually engineered features require extensive data pre-processing and normalization to produce improved results
Temporal and Spatial changes in Activities	The use of hierarchical and translational invariant features helps to solve the problem of intra-class variabilities and inter-class similarities inherent in handcrafted features.	Handcrafted features are inefficient at handling inter-class variabilities and inter-class similarities.
Model Training and Execution time	Require large amount of sensor dataset to avoid overfitting and high computation intensive system, therefore require Graphical Processing Unit (GPU) to speed up training	Require small training data with less computation time and memory usage.

3. Automatic Feature Extraction Using Deep Learning Methods

Deep learning as a machine learning method and artificial intelligence techniques for feature extraction has come a long way since its resurgence in 2006 with the work of Hinton et al. (G. E. Hinton, S. Osindero, & Y.-W. Teh, 2006). The upsurge in deep learning research is fuelled by its ability to extract salient features from raw sensor data without relying on laboriously handcrafted features. Furthermore, in the area of human activity recognition, for instance, complex human activities are translational invariant and hierarchical in nature, and the same activities can be performed in different ways by the same participants. In some cases, activities can be a starting point for other complex activities; running and jogging might not be distinguishable depending on the age and health condition for the person performing the activity.

Deep learning (Bengio, 2009; G. E. Hinton, S. Osindero, & Y. W. Teh, 2006; Hollensen & Trappenberg, 2015) is a machine learning technique that uses representational learning to discover feature representation in raw sensor data automatically. Unlike classical machine learning (support vector machine, k-nearest neighbour, k-mean, etc.) that require a human engineered feature to perform optimally (LeCun, et al., 2015). Over the years, deep learning has provided extensive applications in image recognition (Szegedy, et al., 2015), speech recognition (G. Hinton, et al., 2012), medicine and pharmacy (J. Ma, Sheridan, Liaw, Dahl, & Svetnik, 2015), natural language processing (Bordes, Chopra, & Weston, 2014; Sutskever, Vinyals, & Le, 2014) and recently in human activity recognition (Y. Q. Chen, Xue, & Ieee, 2015; L. Lin, et al., 2016; Rahhal, et al., 2016; C. A. Ronao & S. B. Cho, 2016; C. Vollmer, H. M. Gross, & J. P. Eggert, 2013).

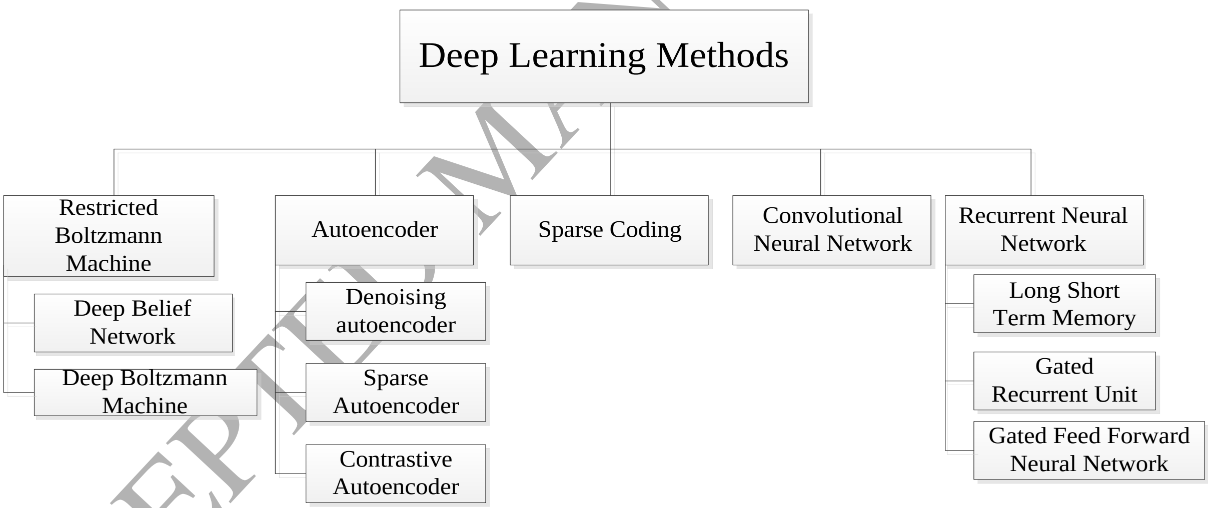


Fig. 1: Different Architecture of Deep Learning Algorithms

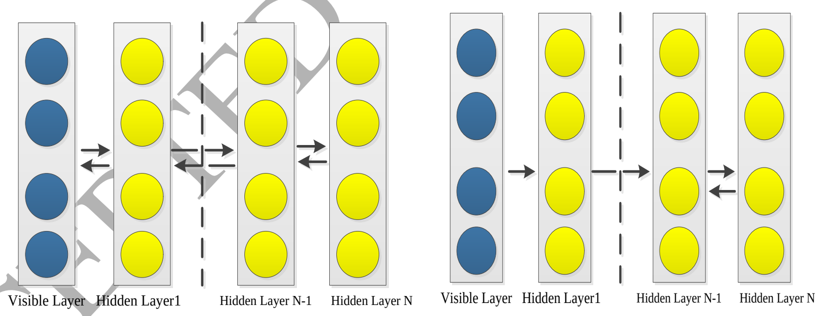
Extensive number of deep learning methods (LeCun, et al., 2015; Schmidhuber, 2015) have been proposed recently, and these methods can be broadly classified into Restricted Boltzmann Machine, Deep Autoencoder, Sparse Coding, Convolutional Neural Network and Recurrent Neural Networks (**Fig. 1**). These methods are reviewed in the subsection below, outlining the characteristics, advantages and drawbacks of each method.

3.1 Restricted Boltzmann Machine

Restricted Boltzmann Machine (Fischer & Igel, 2014; G. E. Hinton & Sejnowski, 1986) is a generative model that serves as a building block in greedy layer by layer feature learning and training of deep neural network. The model is trained with contrastive divergence (CD) to provide unbiased estimates of maximum likelihood learning. However, Restricted Boltzmann Machine is difficult to converge to local minimal and variant of data representation. Furthermore, it is challenging to know how automatic adaptation parameters settings such as learning rate, weight decay, momentum, the size of mini-batch and sparsity can be specified to achieve optimal results (Cho, Raiko, & Ihler, 2011; G. E. Hinton, Srivastava, Krizhevsky, Sutskever, & Salakhutdinov, 2012). Restricted Boltzmann Machine is composed of the visible unit and hidden units that are restricted to form bipartite graph for effective algorithm implementation. Therefore, weights connecting the neurons between visible units and hidden units are conditionally independent without visible-visible or hidden-hidden connections. To provide efficient feature extraction, several RBMs are stacked to form visible to hidden units, and the top layers are fully connected or embedded with classical machine learning to discriminate features vectors (Fischer & Igel, 2014). Although, issues like inactive hidden neuron, class variation, intensity and sensitivity to larger dataset make training RBM difficult. Recently, methods such as regularisation using noisy rectified linear unit (Nair & Hinton, 2010) and temperature based Restricted Boltzmann Machine (G. Li, et al., 2016) have been proposed to resovle the issue. Restricted Boltzmann Machine has been extensively studied in feature extraction and dimensionality reduction (G. E. Hinton & Salakhutdinov, 2006), modelling high dimensional data in video and motion sensors (Taylor, Hinton, & Roweis, 2007), movie rating (Salakhutdinov, Mnih, & Hinton, 2007) and speech recognition (Mohamed & Hinton, 2010). Two well know Restricted Boltzmann Machine methods in literature are Deep Belief Network and Deep Boltzmann Machine (See **Fig. 2**).

Deep Belief Network (G. E. Hinton, et al., 2006) is a deep learning algorithm trained in a greedy-wise layer manner by stacking several Restricted Boltzmann to extract hierarchical features from raw sensor data. Deep Belief Network has directed connection between the lower layer and undirected connection at the top layer that helps to model observed distribution between the vectors space and hidden layers. Likewise, training involves layer by layer at a time with weight fine-tuning using contrastive convergence (CD). Then, the conditional probability distribution of the data is computed to learn robust features that are invariant to transformation, noise and displacement (G. E. Hinton, et al., 2006).

Deep Boltzmann Machine (DBM) (Salakhutdinov & Hinton, 2009; Salakhutdinov & Larochelle, 2010) is a generative model with several hidden layers in undirected connection in the entire network layers. DBM hierarchically learns features from data in which features learned in the first layer are used as latent variables in the next layer. Similar to deep belief network (DBN), Deep Boltzmann machine deploys Markov random field for layer by layer pre-training of massive unlabelled data and provide feedback using bottom-up pass approach. Furthermore, the algorithm is fined through back propagation approach. Fine-tuning allows variation inference and the algorithm to be deployed in specific classification or activity recognition task. Training RBM (Salakhutdinov & Hinton, 2012; Salakhutdinov & Larochelle, 2010) involves maximising the lower bound of likelihood with stochastic maximum likelihood algorithms (Younes, 1999). In this case, training strategies need to adopt a way to determine the training statistics, weight initialization and update after each mini-batch by replacing stochastic binary values with deterministic real probabilities. The major drawback that has been observed in DBM is the time complexity with higher optimisation parameters. In (Montavon & Müller, 2012), a centring optimisation method was proposed for stable learning algorithms and Midsized DBM for faster and good generative and discriminative model.



(a) (b)

Fig. 2: Representation of Restricted Boltzmann Machine: (a) Deep Belief Network (b) Deep Boltzmann Machine

3.2 Deep Autoencoder

The autoencoder method replicates the copies of the input value as output as shown in **Fig. 3.** Using encoder and decoding units, autoencoder methods produces the most discriminative features from unlabeled sensor data by projecting them to lower dimensional space. The encoder transforms the sensor data input into hidden features which are then reconstructed by the decoder to approximate values to minimise error rates (Liou, Cheng, Liou, & Liou, 2014; Lukun Wang, 2016). The method provides data-driven learning feature extraction techniques to avoid problems inherent in handcrafted features. Training autoencoder is done in such a way that the hidden units are smaller than the inputs or outputs to provide a lower dimensional discriminative feature for recognition of activities with reduced computation time (Ravì, Wong, Deligianni, et al., 2017). Moreover, autoencoder algorithm uses multiple layer of encoder units to transform high dimensional data into the low dimensional feature vectors. Autoencoder algorithm is pre-trained using restricted Boltzmann machine due to its complexity (G. E. Hinton & Salakhutdinov, 2006) and then obtains higher feature representations by stacking several level of autoencoder algorithms (Jie Zhang, Shan, Kan, & Chen, 2014). Generally, different variations of autoencoder have been proposed to ensure robust features representation for machine learning applications. These include denoising autoencoder, sparse autoencoder and contractive autoencoder.

Denoising autoencoder was first introduced by Vincent et al. (Vincent, Larochelle, Bengio, & Manzagol, 2008) as method to stochastically learn robust feature representation from corrupted version of data (e.g sensor values) by partial destruction of the raw input sample. Thus, denoising autoencoder is trained to reconstruct sample input data from corrupted version by forcing random sample values of the data to zero through stochastic mapping. Similar to other unsupervised deep learning model, denoising autoencoder is trained through layer to layer initialisation. Each layer of the network is trained to produce input data of the next higher level layer representation. The layer to layer training ensure that autoencoder network is able to capture robust structure and observed statistical dependencies and regularities about input data distributions. Moreover, stacked denoising autoencoder can be stacked to learn useful representation of corrupted version of input sample data which have been found to give less classification error (Vincent, Larochelle, Lajoie, Bengio, & Manzagol, 2010), and this was recently applied to recognise complex activities (Oyedotun & Khashman, 2016).

Sparse autoencoder (Marc'Aurelio Ranzato, et al., 2007) is unsupervised deep learning model developed for sparse and over-complete feature representation from input data by forcing sparsity term to the model loss function and set some of the active units close to zero. Sparse autoencoder is highly applicable in tasks that require analysis of high dimensional and complex input data such as motion sensors, images and videos. Generally, the use of sparsity term allow the model to learn feature representation that are robust, linearly separable and invariant to changes, distortion, displacements and learning applications(Zhou, et al., 2015). Therefore, sparse autoencoder model is very efficient for extraction of low dimensional features from high dimensional input data and compact interpretation of complex input data using supervised learning approach(H. Liu & Taniguchi, 2014).

Recently, Rifai, Vincent, Muller, Glorot, and Bengio (2011) propose *contrative autoencoder* by introducing penalty term of partial derivatives for efficient feature representation. The use of sum of square of all partial derivatives for the feature vectors with respect to size of input data, force the features within neighbourhood of the input data(Dauphin, et al., 2012). Furthermore, penalty term reduces the dimensional feature space with the training datasets and makes it invariant to changes and distortion. Contractive autoencoder is similar to denoising autoencoder as both apply penalty term to the small corrupted data sample. However, unlike the denoising autoencoder, the contractive autoencoder applies an analytic penalty to the whole data instead of the encoding input sample (Mesnil, et al., 2012). Section 4.1.3 discusses the applications of autoencoder in mobile and wearable sensor based human activity recognition and health monitoring.

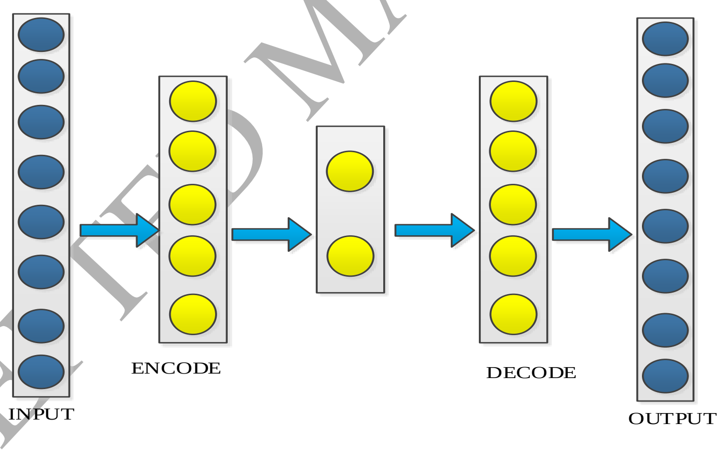


Fig. 3: Deep Autoencoder encoding and decoding process

3.3 Sparse Coding

Sparse coding was first proposed by (Olshausen & Field, 1997) as a machine learning technique for learning over-complete basis in order to produce efficient data representation. Sparse coding provides an effective means of reducing the dimensionality of data and dynamically represent the data as a linear combination of basis vectors. This enable sparse coding model captures the data structure and determines correlations between various input vectors(Y. Guo, et al., 2016). Recently, some studies have proposed sparse coding methods to learn data representation particularly in human activity recognition. These include the shift-invariant method (C. Vollmer, H.-M. Gross, & J. P. Eggert, 2013) and sparse fusion (Ding, Lei, & Rao, 2016). These algorithms provide feature dimensionality reduction strategies to reduce computational complexities for implementation of human activity recognition system using mobile phone and wearable devices.

3.4 Convolutional Neural Network

Convolutional Neural Network (CNN) (LeCun, Huang, & Bottou, 2004) is a Deep Neural Network with interconnected structures. A convolutional neural network performs convolution operations on raw data (e.g. sensor values) and is one of the most researched deep learning techniques which has found extensive applications in image classification, sentence modelling, speech recognition and recently in mobile and wearable sensors based human activity recognition (Y. Guo, et al., 2016; Karpathy, Johnson, & Fei-Fei, 2015; C. A. Ronao & S.-B. Cho, 2016). Generally, convolutional neural network model is composed of *convolutional layer, pooling layer and fully connected layer*. These layers are stacked to form deep architecture for automatic feature extraction in raw sensor data (Ordóñez & Roggen, 2016; Limin Wang, Qiao, & Tang, 2015). The convolutional layer captures the feature maps with different kernel sizes and strides and then pooled the features maps together in order to reduce the number of connections between the convolutional layer and the pooling layer. The pooling layer reduces the feature maps, number of parameters and makes the network translational invariant to changes and distortion. In the past, different pooling strategies have been proposed for Convolutional Neural Network implementation in various area of applications. These include max pooling, average pooling, stochastic pooling and spatial pooling units (Y. Guo, et al., 2016). Recently, theoretical analysis and performance evaluations of these pooling strategies have shown superior performance of max pooling strategies. Thus, max pooling strategy is extensively applied in deep learning training (Boureau, Ponce, & LeCun, 2010; Scherer, Müller, & Behnke, 2010). Moreover, recent studies human activity recognition also applies max pooling strategies due to its robustness to small changes (Kautz, et al., 2017; G. Liu, Liang, Lan, Hao, & Chen, 2016). However, studies in time series analysis with deep learning observed reduction in discriminative ability of max pooling strategies (Abdel-Hamid, Deng, & Yu, 2013). Therefore, further experimental analysis and evaluation is required to ascertain the effectives of these pooling strategies in human activity recognition and time series applications.

The fully connected layer is fused with the inference engine such as SoftMax, Support Vector Machine or Hidden Markov Model that takes the features vectors from sensor data for activity recognition (Erfani, Rajasegarar, Karunasekera, & Leckie, 2016; C. A. Ronao & S.-B. Cho, 2016; Ronaoo & Cho, 2015). In CNN, activation unit values are computed for each region of the network in order to learn patterns across the input data(Ordóñez & Roggen, 2016). The output of convolutional operation is computed as $C_i^{1.j} = \alpha\left(b_j^l + \sum_{m=1}^{M} w_m^{l.j} x_{i+m-1}^{l-1.j}\right)$, where l is the layer index, σ is the activation function, b is the bias term for the feature map, M is the kernel/filter size, W is the weight of the feature map. The weight may be shared to reduce complexity and make the network easy to train. Generally, idea of convolutional neural network (CNN) was inspired by (Hubel & Wiesel, 1962) which noted that the human visual cortex consists of maps of the local receptive field that decrease in granularity as the cortex move along the receptive fields. Since the proposal, a number of other CNN architectures have been developed by researchers. These include the AlexNet (Krizhevsky, Sutskever, & Hinton, 2012), VGG (Krizhevsky, et al., 2012) and GoogleNet (Szegedy, et al., 2015).

Recently, CNN architectures that combine other deep learning techniques or fusion of different CNN architectures (Jing, Wang, Zhao, & Wang, 2017; Ordóñez & Roggen, 2016) were also proposed. For instance, (Ordóñez & Roggen, 2016) proposes DeepConvLSTM, an architecture that replaces the pooling layer of the convolutional neural network with Long Short Term Memory (LSTM) of the recurrent neural network. Also, convolutional deep belief networks (CDBN) was developed by (Lee, Grosse, Ranganath, & Ng, 2009) which exploit the power of discriminative CNN and pre-training technique of Deep Belief Network. Furthermore, Masci et al (Masci, Meier, Cireşan, & Schmidhuber, 2011) proposed deep convolutional autoencoder for feature learning by integrating convolution neural network and autoencoder trained with online stochastic gradient descent optimisation. The architecture of Convolutional neural network is shown in **Fig. 4.**

3.5 Recurrent Neural Network

Recurrent neural network (RNN) was developed to model sequential data such as time series or raw sensor data (**Fig. 5**). RNN incorporates a temporal layer to capture sequential information and then learns complex changes using the hidden unit of the recurrent cell. The hidden unit cells can change based on the information available to the network, and this information is constantly updated to reflect the current status of the network. RNN computes the current hidden state by estimating the next hidden state as activation of the previously hidden state. However, the model is difficult to train and suffer from vanishing or exploding gradients limiting its application for modelling long time activity sequence and temporal dependencies in sensor data (Guan & Ploetz, 2017). Variations of RNN such as Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) integrate varieties of gates and memory cells to capture temporal activity sequence(Graves, 2013). Long Short Term Memory (Hochreiter & Schmidhuber, 1997) incorporated memory cell to store contextual information, thereby control flow of information into the network. With the inclusion of memory cells such input gate, function gate and output gate alongside learn-able weights, allow LSTM model temporal dependencies in sequential data and adequately capture global features to boost recognition accuracy (Zaremba, 2015).

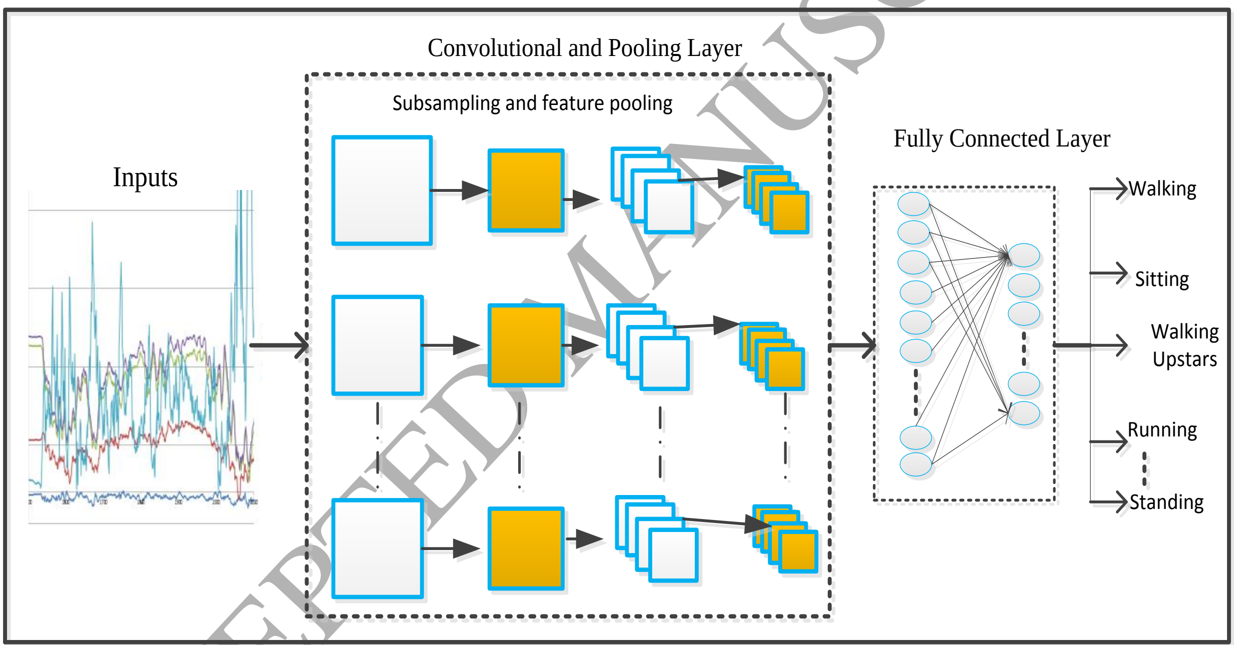


Fig. 4: Deep Convolutional Neural Network

Despite the advantages inherent in LSTM, Cho, et al. (2014) observed that issues of too many parameters that need to be updated during training increases computational complexity of LSTM. To reduce parameter update, they introduced *Gated recurrent units* with fewer parameters that make it faster and less complex to implement. LSTM and Gated Recurrent Unit (GRU) differ in the way the next hidden state are updated and contents exposure mechanism (Valipour, Siam, Jagersand, & Ray, 2016). While LSTM updates by summation operation, GRU updates the next hidden state by taking correlation based on the amount of time needed to keep such information in the memory. Moreover, recent comparative analysis of the performance of LSTM and GRU shown that GRU slightly outperformed LSTM in most of machine learning applications (Chung, Gulcehre, Cho, & Bengio, 2014). An attempt has also been made to improve on GRU by reducing the number of gates in the network and introduce only multiplicative gates to control the flow of information (Y. Gao & Glowacka, 2016). The algorithm was compared with LSTM and GRU, and it outperformed them in terms of memory requirement and computational time. Recently, Chung, Gülçehre, Cho, and Bengio (2015) proposed Gated Feedback Recurrent Neural Network (GF-RNN) to solve the problem of learning at multiplicative scale. This learning process is very challenging in application area such as language modelling and programming language sequence evaluation. Specifically, Gated Feedback Recurrent Neural Networks is developed by stacking multiple recurrent layers and allow control of the signal flowing from upper layer to the lower layer. The mechanism is done by adaptively controlling based on the previously hidden state and assign different layer with different timescale. However, GF-RNN is not popular in human activity recognition. For all the studies review, we find no specific work that apply GF-RNN for human activity. Therefore, the model is omitted in our review of dep learning based human activity recognition in Section 4.2.2.

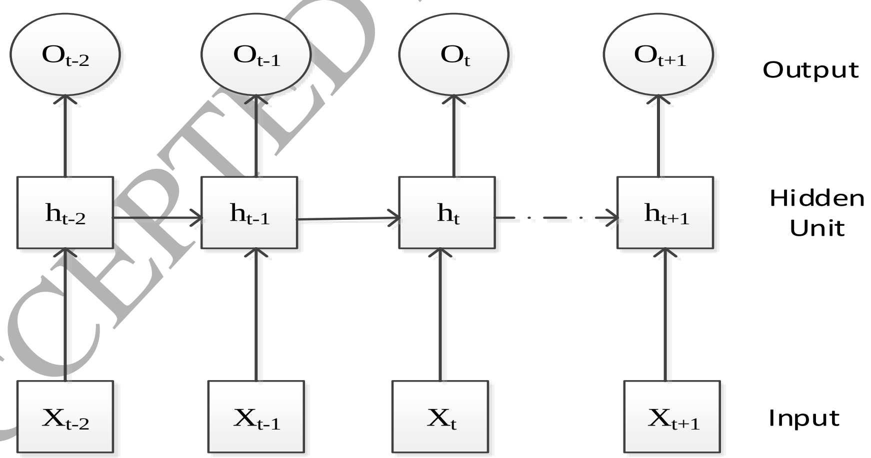


Fig. 5: Simple Recurrent Neural Network

3.6 Strengths and Weaknesses of different deep learning methods

In this section, we compare these methods discussed above noting their strengths and weaknesses for mobile and wearable based human activity recognition. The different deep learning methods discussed in this review have produce state-of-arts performances in mobile and wearable sensor based human activity recognition (**section 4**). The main advantage of deep learning is the ability to automatically learn from unlabelled raw sensor data. However, these methods provide different capabilities for sensor stream processing. For instance, Restricted Boltzmann machine algorithms are efficient for automatic and efficient unsupervised transformation of sensor data into feature vector using layer by layer training leveraging unlabelled data. Also, the methods allow robust feature vectors extraction. Nevertheless, Restricted Boltzmann machine presents major drawback such as high parameter initialisation that make training computationally expensive. Considering the computation capabilities of mobile and wearable sensor devices, it is difficult to support on-board and real-time activity recognition (Yalçın, 2016). On the other hand, Deep autoencoder are efficient for unsupervised feature transformation into lower feature vectors automatically from raw sensor data. Specifically, deep autoencoder methods are trained using greedy layer by layer approach for unsupervised feature learning from continuous sensor streams. Deep autoencoder algorithms are robust to noisy sensor data with ability to learn hierarchical and complex features from sensor data. However, the major drawbacks of deep autoencoder are the inability to search for optimal solutions and high computation time due to high parameter tuning. While sparse coding methods are efficient for reduction of high dimensional sensor data into linear combination feature vectors and ensure compact representation of features. Moreover, sparse coding is invariant to sensor transformation and orientation, and effective for modelling changes in activity progression(M. Zhang & Sawchuk, 2013). Change in sensor orientation is big challenges in human activity recognition system especially for smartphone accelerometers(O. Incel, 2015). In this, accelerometer signal produce by smartphone or wearable devices change with variations in orientation and placement positions. Nevertheless, it is still challenging to effectively perform unsupervised features learning with sparse coding. Convolutional Neural Network are capable of learning deep feature vectors from sensor data for modelling complex and high dimensional sensor data. The main advantage of CNN is the ability to use pooling layer to reduce training data dimensions and make it translational invariant to changes and distortion(C. A. Ronao & S.-B. Cho, 2016). The algorithms is capable of learning long range and repetitive activities through multi-channel approach (Zeng, et al., 2014). Convolutional Neural Networks are more inclined for image processing, therefore, sensor data are converted to image description to support extraction of discriminative features (Sathyanarayana, Joty, Fernandez-Luque, Ofli, Srivastava, Elmagarmid, Taheri, et al., 2016). Convolutional Neural Network are deployed to solve the problem of uncertainty in sensor measurement and conflicting correlation in high dimensional sensor data. However, CNN require high number of hyper-parameter tuning to achieve optimal features. Furthermore, it is challenging to support effective on-board recognition of complex activity details. **Section 4.2.1** provide comprehensive review of Convolutional Neural Networks implementation for human activity recognition. Finally, Recurrent Neural Networks are applied to model temporal dynamics in sensor data, thus enable modelling of complex activity details. RNN such as Long Short Term Memory are efficient at creating global temporal dependencies in sensor data. The major issue in Recurrent Neural Networks especially long short term memory is the high computation time due to large number of parameter update. Techniques such as high throughput parameter update approach may help to reduce computation time(Inoue, Inoue, & Nishida, 2016).

Table *2* summarises the recent applications domain in mobile and wearable sensor based human activity recognition, strength and weakness of each deep learning methods, placing emphasis on sensor data processing. Furthermore, the categorisation of each method for human activity recognition is presented in **section 4**.

Table 2: Deep Learning methods

Methods	**Descriptions**	**Strengths**	**Weaknesses**	**Recent Application in Human activity monitoring and Detection**
Deep Belief Network	Has directed connection at the lower layer and undirected connection at two topmost layer	Unsupervised training with unlabelled sensor streams which is naturally available through cyber-physical systems and Internet of Things and initialisation prevent convergence at local minima	Mobile and wearable sensor on-board training of the network is computationally complex due to extensive parameters initialization process	Activity of daily living (ADL) localisation, detection of posture and hand gestures activities in Alzheimer.
Deep Boltzmann Machine	Has undirected connection at every layer of the network	Allow feedback mechanism for more robust feature extraction through unsupervised training.	Due to resource constraint nature of mobile devices, Joint optimisations are required to reduce operation overhead and execution cost. However, DBM joint optimisation is practically difficult to achieve	Diagnosis of emotional state in elderly and detection of irregular heartbeats during intensive exercise.
Denoising autoencoder	Enable correct reconstruction of corrupted input values	Robust to corrupted sensor data streams	High computational time, lack of scalability to high dimensional data, rely on iterative and numerical optimisation and high parameter tuning (M. Chen, Xu, Weinberger, & Sha, 2012)	Automatic detection of activity of daily living (ADL).
Sparse Autoencoder	Impose sparsity term to the loss function to produce robust features that are invariant to learning applications	Produce more linearly separable features	High computational time due to numerous forward pass for every example of the data sample (Ng, 2011)	Health rate analysis during intensive sports activities and health monitoring
Contractive autoencoder	Add analytic penalty instead of the stochastic penalty to the reconstruction error functions	Reduced dimensional features space and is invariant to changes and local dependencies	Difficult to optimise and greedy pre-training does not find stable nonlinear features especially for one layer autoencoder (Schulz, Cho, Raiko, & Behnke, 2015)	Activity of daily living(ADL), user location and activity context recommendations
Sparse Coding	Over-complete basis for reducing the dimensionality of data as linear combination of basis vector	The use of sparse coding method for dimensionality reduction of input data helps to minimise computational complexity	Efficient handling and computation of feature vectors are non-trivial (Harandi, Sanderson, Hartley, & Lovell, 2012). It is also difficult to develop deep architecture with sparse coding (He, Kavukcuoglu, Wang, Szlam, & Qi, 2014)	Representation of energy related and health monitoring smart homes and Activity of daily living(ADL)
Convolutional Neural Net-	Deep neural network with interconnected	Widely implemented in deep learning with a lot of training	Require large dataset and high number of hyper-parameter	Predict relationship between exercises and sleep patterns,

Methods	Descriptions	Strengths	Weaknesses	Recent Application in Human activity monitoring and Detection
work	structure inspired by biological visual cortex	strategies proposed. Automatically learn features from raw sensor data. Moreover, CNN is invariant to sensor data orientation and change in activity details.	tuning to achieve optimal features. Maybe difficult to support effective on-board recognition of complex activity details.	automatic pain recognition during strenuous sports activities, energy expenditure estimation and tracking of personal activities.
Recurrent Neural Network	Neural network for modelling sequential time series data. Incorporate temporal layer to learn complex changes in data	Used to model time dependencies in data	Difficult to train and suffer from vanishing or exploding gradients. In case of LSTM, require too many parameter updates. Large parameter update is challenging for real-time activity predictions.	Model temporal patterns in activity of daily living (ADL), progressive detection of activity levels, fall and heart failures in elderly.

4. Deep Learning Approaches for Human Activity Recognition Using Mobile and Wearable Sensor Data

Research on the use of deep learning for feature representations and classification is growing rapidly. Generally, deep learning methods can be subdivided into generative model, discriminative model and hybrid model (Deng, 2014). These subdivisions are presented in ***Fig. 6*** . The generative models are graphical models that represent independent or dependent distributions in sensor data where *graphs node* represent the random variable of the given sensor data and *arc* represent the relationship between variables. Generative models capture higher order correlation by identifying joint statistical distributions with associated class. Moreover, generative models use unlabeled datasets that are pre-trained with greedy layer by layer approach and then fine-tuned with labelled data which is then classified with classical machine learning such as Support Vector Machine (SVM) or HMM (Bengio, 2009; Hodo, Bellekens, Hamilton, Tachtatzis, & Atkinson, 2017; Mamoshina, et al., 2016). Among deep learning methods in these categories are Restricted Boltzmann, Autoencoder, Sparse Coding and Deep Gaussian Mixture. In the case of the discriminative models, the posterior distribution provides discriminative power in classification and modelling of label sensor data. A convolutional neural network is an important category of discriminative deep learning model (Mamoshina, et al., 2016). Others are Recurrent Neural Network, Artificial Hydrocarbon and Deep Neural Model. Conversely, hybrid models are used to classify data by deploying the feature output generated by generative models. This involves pre-training of the data to enhance computational time and then classify with classical machine learning algorithms. The generative model reinforces hybrid models through optimisation and regularisation procedures (Deng, 2014). In this review, the studies categorised as a hybrid models are those that combine generative and discriminative or both methods for human activity recognition. Notable examples in this area are Convolutional Restricted Boltzmann Machine(Sarkar, Reddy, Dorgan, Fidopiastis, & Giering, 2016), Convolutional Recurrent Neural Network (Ordóñez & Roggen, 2016) and an ensemble of homogenous convolutional neural network features (Ijjina & Mohan, 2016).

In human activity recognition, deep learning is used in diverse tasks such as estimating changes in the movement pattern for the elderly (Yi, Cheng, & Xu, 2017), labelling of human activity sequence(R. Yao, Lin, Shi, & Ranasinghe, 2017), recognition of emotion in people in need using electroencephalogram (EEG) (Yanagimoto & Sugimoto, 2016) and health anomaly detection using physiological signals. To efficiently achieve these, require automatic feature representation. Therefore, deep learning methods provide effective features representation approach to improve classification errors and reduce computational complexity in human activity recognition. For instance, the variants of Restricted Boltzmann Machine methods play vital role in features dimension reduction and automatically discover discriminative features using a layer by layer pre-training to increase recognition accuracy. Restricted Boltzmann Machine provides an excellent method for learning improved features from unlabeled data and then pre-trained for complex activity recognition. The high-order dependencies and localisation among group activities features are extracted with different deep learning methods (Alsheikh, et al., 2015).

Sensor data processing are classical time series learning and require high input sensor data adaptation to enable efficient processing. Mobile and wearable sensor data generate time series sensor data in one dimension (1D) (Zeng, et al., 2014). It is challenging to processing motion sensor with high dimensional deep learning architectures. Two approaches have been proposed to convert the sensor streams to fit into deep learning algorithms. These include channel or model based approaches. Channel based approach utilise the sensor dimension as the dimension of the network architecture and extract features from each axis for activity recognition and fall detection (Khan & Taati, 2017; Ordóñez & Roggen, 2016). The sensor axes are used to perform 1D convolution for extraction of salient feature and then combined at the fully connected layers (Sathyanarayana, Joty, Fernandez-Luque, Ofli, Srivastava, Elmagarmid, Arora, et al., 2016). Model based methods use temporal correlation of sensor data to convert the sensor data into 2-D image descriptions and apply 2-D convolution operation to extract features. These are common in

Convolutional Neural Network for human activity recognition (Jiang & Yin, 2015; Ravì, Wong, Lo, & Yang, 2017). For instance, Ravì, Wong, Lo, et al. (2017) propose spectrogram representation to transform the motion sensor data (accelerometer and gyroscope) into local temporal convolution to reduce computational complexity. The types of input adaptation employ for motion sensor in human activity recognition depends application domains. Other works modified the convolutional kernel of Convolutional Neural Network to capture temporal dependencies from multiple sensors (Yuqing Chen & Xue, 2015). Therefore, previous studies on deep learning implementation for human activity recognition adopt these input data adaptation approaches to automatically extract relevant features from raw sensor data.

In this section, we discuss recent studies for deep learning implementation of human activity recognition for mobile and wearable sensors. In **Fig. 6**, these methods are depicted while subsequent sections outline their uniqueness for feature extraction in mobile and wearable sensor based human activity recognition.

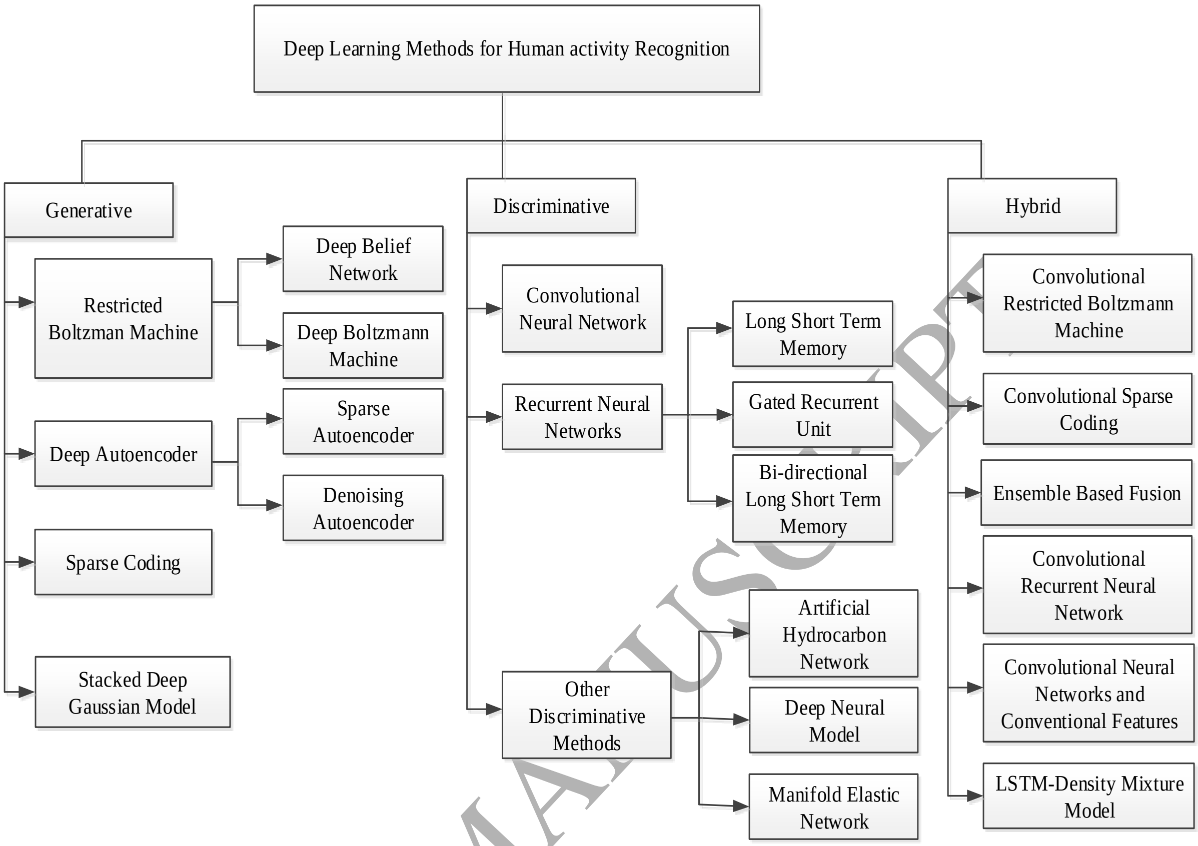


Fig. 6: Taxonomy of Recent Deep Learning Methods for Human Activity Recognition

4.1 Generative Deep Learning Methods

As stated earlier, generative deep learning methods model independent or dependent distributions in data and high order correlation by identifying the joint statistical distribution with associated classes. In the past decade, various studies have been conducted using generative feature extraction models for human activity recognition. Here, we analysed these and their implementation advantages

4.1.1 Deep Restricted Boltzmann Machine methods

Pioneering the use of deep learning based generative feature extraction for human activity recognition was started by (Plötz, et al., 2011) when they proposed the performance evaluation of different generative feature extraction and dimensionality reduction techniques such as autoencoder, principal component analysis, empirical cumulative distribution function and statistical features. An extensive experiment using sensor based on public datasets showed that autoencoder outperforms other feature extraction techniques including handcrafted features. A number of other deep learning methods for human activity recognition have since followed suit. For instance, the deep belief network proposed by (Geoffrey E Hinton, et al., 2006) was used to extract hierarchical features from motion sensor data and then model stochastic temporal activity sequence using Hidden Markov Model(Alsheikh, Niyato, Lin, Tan, & Han, 2016). The work was later extended for on-board mobile phone implementation using mobile Spark platform (Alsheikh, et al., 2016). Also, studies by (Yalçın, 2016; L. Zhang, Wu, & Luo, 2015a) introduce deep belief network for online and real-time feature extraction for human activity recognition. However, due to the computationally intensive nature of deep learning, the algorithm was trained offline with generative backpropagation initialized parameters and activity classification done with SoftMax Regression. Deep learning has also provided feature representation for the online classification task, contextual information provision for sensor and real-time recognition of simple to complex activities details using datasets collected with the aid of mobile devices (L. Zhang, Wu, & Luo, 2015a, 2015b; L. Zhang, et al., 2015a; L. Zhang, Wu, & Luo, 2015b). However, the use of large window size and storing previous data to provide contextual information in some of the studies aid increased computational time and memory usage. Deep Belief Network has also provided excellent means to model temporal dependencies and observable posterior distribution in sensor data with Hidden Markov model for diagnosis and recognition of emotions state in elderly using wearable sensor worn on the patients' scalp (X. Jia, Li, Li, & Zhang, 2014; L. Zhang, et al., 2015b). Also, Z. Y. Wu, Ding, and Zhang (2016) proposed unsupervised feature extraction and recognition of irregular heart beat during intensive exercise by stacking various layers of Restricted Boltzmann machine. The stacked layers enable hierarchical extraction of discriminative features that clearly describe complex activity details. The objective is to provide automatic health monitoring in special cases such as brain activity detection (Electroencephalogram), eye movement (Electrocochleogram), skeletal muscle activity (Electromyogram) and heart rate (Electrocardiogram). This will ensure appropriate independent living and overall health status for the elderly (Längkvist, Karlsson, & Loutfi, 2012; Z. Y. Wu, et al., 2016; H. Xu & Plataniotis, 2016).

Y. Zhao and He (2014) explored implementation of deep Restricted Boltzmann Machine for detection of hand activity in elderly with Alzheimer's disease using Electroencephalogram dataset collected with wearable devices worn by patients. They leverage on incremental learning and support vector machine to classify what features may lead to accurate diagnosis of the disease. In recent study, S. Bhattacharya and Lane (2016) investigated smartwatch-centric activity recognition and the possibility of implementing deep learning in wearable devices. They concluded that GPU-enabled smartwatch could provide deep learning implementation. The framework implemented on Snapdragon 400 SoC wristwatch achieved high accuracy for common daily activity such as hand gesture, indoor/outdoor localisation, and transport model using public datasets. Another key study was presented by H. Fang and Hu (2014), to learn automatic features for recognition of human activities in constrained environment. The dataset was gathered for a fifty (50) day period, leveraging on current and previous activity, and the duration of the activity to ascertain the individual activities. The problem of recognising interleaved and overlapped activities was examined by Radu, et al. (2016), for multimodal and Deep Boltzmann Machine based human activity recognition using pattern mining. With this, the unannotated activity can be discovered by deploying sensors of different modalities.

4.1.2 Deep Autoencoder Methods

Autoencoder, another generative feature learning technique has also dominated human activity recognition landscape. For instance, Plötz, et al. (2011) had earlier argued the superiority of autoencoder over PCA, ECDF and statistical feature extraction methods. Other researchers have also developed autoencoder techniques for human activity recognition. Recent studies by (Mahmudul Hasan & Roy-Chowdhury, 2014; M. Hasan & Roy-Chowdhury, 2015) propose the use of sparse autoencoder for human activity recognition. The algorithm was proposed to learn features from continuous data streams and then activity details were classified using multi-logistic regression classifier (SoftMax). Learning of features in stream sensors are very challenging due to the scarcity of label data, class invariant and concept drift. However, with incremental learning and sparse autoencoder, they automatically learn features without relying on manually annotated data. Performance evaluation of sparse autoencoder, deep autoencoder and principal component analysis was examined by (H. Liu & Taniguchi, 2014). They observed that the use high depth deep sparse autoencoder enable extraction of more discriminative features compared to deep autoencoder and PCA using a dataset from CMU Lab. In Y. Li, Shi, Ding, and Liu (2014), three basic autoencoder methods were evaluated for human activity recognition from data collected using smartphones. They concluded that sparse autoencoder outperformed other feature learning techniques in terms of accuracy. However, due to the small size of the smartphone dataset and computational platform used in the study, the performance cannot be accurately generalised. Similarly, Harasimowicz (2014) evaluated effects of pre-processing on the performance of generative models for feature extraction, examining algorithms comparatively using sparse autoencoder and concluded that pre-processing has a strong influence on the performance of activity classification especially normalisation techniques.

Besides works that parameters evaluation of autoencoder for and preprocessing for human activity recogniton, other studies have further examined mobile based implementation of stacked autoencoder for human motion analysis using motion sensors (accelerometer, gyroscope, gravity sensors etc.) with high performance accuracy(Zhou, et al., 2015). Similarly, Lukun Wang (2016) extracts features from the accelerometer and magnetic sensors using continuous autoencoder for the development of automatic human activity recognition. The proposed continuous autoencoder adds randomness and converts the high dimension inputs into low dimensional vectors by encoding and decoding process at the hidden layers. To increase the learning rate of the algorithm, stochastic gradient descent optimisation was introduced in the hidden layer, and the algorithm was compared with statistical features with enhanced performance obtained. Shared-based autoencoder for separation of multiple input modalities sensors into hierarchical component was proposed by(A. Shahroudy, Liu, Ng, & Wang, 2016). In the study, factorised input modalities were stacked to convert complex and nonlinear input representation into linear vectors for classification. The main advantage of this method is its robustness to noise and ability to extract hierarchical and complex features. Furthermore, Zhou, et al. (2015) proposed stacked autoencoder for feature extraction for Android smartphone based motion recognition using sensor data modalities with high-performance accuracy. In addition to checking human activity to promote a healthy life, mobile sensor data can further help in the diagnosis of lifestyle related illnesses. Related work for such application was recently proposed by(Unger, Bar, Shapira, & Rokach, 2016) using stacked autoencoder. The proposed stacked autoencoder was developed for recognition and recommendation of online based activity leveraging mobile sensor data. The deep learning method helped to reduce the dimensionality of the data and select the feature that best provides the context-aware recommendation, user location and users preference. Stacked autoencoder has also been extended to generate a sequence of time series to characterise human movement pattern based on time elapse window properties (Munoz-Organero & Ruiz-Blazquez, 2017). Related implementation for fall detection using sensor data generated by radar was presented in (Jokanovic, Amin, & Ahmad, 2016). The stacked autoencoder provides mechanism to reduce the dimensionality of the data into lower dimensional features that are feed into SoftMax regression for fall identification. The use of dimensionality reduction strategies helps to reduce computational complexity notably for mobile based implementation.

Stacked denoising autoencoder when combined with active learning provide excellent means for automatic labelling and feature extraction for activity recognition and heart rate analysis during intensive exercise. Moreover, stacked denoising autoencoder implementation are important for morbidity rate prediction(Al Rahhal, et al., 2016; Q. Song, Y.-J. Zheng, Y. Xue, W.-G. Sheng, & M.-R. Zhao, 2017). There is a great need to enable independent living for elderly in different parts of the world due to the high rate of ageing populations. With such assistance, elderly citizens can function optimally by utilising sensor-equipped smart homes. One major challenge is how to increase the performance of the algorithm and automatically extract feature vectors. More so, obtaining labelled data that will be exploited by features engineers is difficult. To solve the problem and improve the performance of human activity recognition in the smart home environment, A. Wang, Chen, Shang, Zhang, and Liu (2016) proposed de-noising autoencoder techniques to learn underlying feature representation in sensor data and then integrate it with a classifier trained into single architecture to obtain powerful recognition model. In general, autoencoder methods have demonstrated excellent approaches for automatic feature representation to learn latent feature representation for human activity monitoring and detection approach. Generally, stacked autoencoder provide compact feature representation from continuous unlabelled sensor streams to enable robust and seamless implementation of human activity recognition system.

4.1.3 Sparse Coding methods

Sparse coding proposed in (Olshausen & Field, 1997) provides a means to reduce sensor data dimension and represent them as an efficient linear combination of basis vectors. Due to the efficient data representation ability of sparse coding, a number of studies have used it to develop feature extraction and representations for human activity recognition. For instance, sparse coding method was presented Y. Zhu, Zhao, Fu, and Liu (2010) to convert feature in activity recognition into linear combination vector trained with dictionary algorithm. Additionally, Sourav Bhattacharya, Nurmi, Hammerla, and Plötz (2014) examined the use of sparse coding algorithm trained on self-taught theorem and codebook basis for combination of feature vectors. The sensor data were converted into a linear combination, and the dimension was reduced to generated movement patterns computed from raw sensor signals. The algorithm outperformed other well-known dimensionality reduction feature learning algorithms such as PCA and semi-supervised En-co Training. Sparse Coding was also used to pre-process and learn basic function that captures high representation in sensor data. Then, activity details were classified using neural network classifier for wireless sensor network based health monitoring(J. Guo, Xie, Bie, & Sun, 2014). A major problem in activity recognition is how to solve the problem of intra-class and inter-class variation and complex nature of human body movement (Bulling, et al., 2014b). To minimize intra-class and inter-class variation, M. Zhang and Sawchuk (2013) proposed sparse representation techniques that employ the use of an over-complete dictionary to represent the human signal as a sparse linear combination of activity classes. In the algorithm, class membership was determined by solving the L_1 minimisation problem. The authors compare the technique with other established classical machine learning method (logistic regression, multinomial regression and decision tree) with impressive results obtain with sparse coding. Sparse coding methods provide the possibility for constrained linear coding representation of energy-related activities in smart home environments using sensor streams. Therefore, sparse coding inherently apply sparse dictionary to reduce manual annotation of data(Q. Zhu, Chen, & Soh, 2015).

4.1.4 Stacked Deep Gaussian methods

Recently, various studies have developed deep learning model by stacking a classical generative model to form a deep architecture. Typical examples are Gaussian process classifier (X. M. Wang, et al., 2016), molecular complex detection method (Lu, et al., 2016), and the Deep Gaussian Model. The Gaussian process model provides unsupervised feature extraction by stacking several layers of Gaussian processes to produce robust features. Lu, et al. (2016) explored the issue of gathering huge amount of sensor data, complex and diverse activities by proposing the molecular complex detection method. The technique was first introduced to study protein interaction by (Bader & Hogue, 2003) and the authors extended the algorithm for effective recognition and detection daily activity, product recommendation and sports activity using accelerometer data. Recent work by Feng, Yuan, and Lu (2017), proposed Deep Gaussian Mixture Model that adaptively uses multilayer nonlinear input transformation to extract salient features from motion sensors for human activity recogniton.

However, majority of the generative models have fully connected layer and cannot capture local and temporal dependencies in sensor data. In general, generative models have difficult optimisation procedures, computationally expensive training processes and suffer from vanishing gradient problem (G. E. Hinton, et al., 2012). **Table 3** summarises the different generative deep learning methods for feature extraction in human activity recognition.

Table 3: Generative deep learning methods for human activity recognition

References	Methods	Description	Advantages
(Alsheikh, et al., 2016; Alsheikh, et al., 2015; Erfani, et al., 2016; H. Fang & Hu, 2014; X. Jia, et al., 2014; Längkvist, et al., 2012; Z. Y. Wu, et al., 2016; Yalçın, 2016; L. Zhang, et al., 2015a, 2015b; L. Zhang, et al., 2015a, 2015b)	Deep Belief Network	Generative model that learn greedy layer-wise compact representation of sensor data and learn high-dimensional manifold from unlabelled data	Generate feature from unlabelled sensor data that are invariant to irrelevant variation. Used for nonlinear dimensionality reduction of high dimensional sensor data
(S. Bhattacharya & Lane, 2016; Radu, et al., 2016; Y. Zhao & He, 2014)	Deep Boltzmann Machine	Generative undirected bipartite graphs composed of stochastic visible and hidden units. The layers are stacked into deep layers for extracting salient features from sensor observations	Use sparse representation techniques to reduce data sensitivity. Allow cross-correlation feature extraction and sensor fusion for innate feature representation
(Al Rahhal, et al., 2016; Jokanovic, et al., 2016; Munoz-Organero & Ruiz-Blazquez, 2017; Plötz, et al., 2011; Amir Shahroudy, Ng, Gong, & Wang, 2016; Shimizu, et al., 2016; Unger, et al., 2016; Zhou, et al., 2015)	Deep Autoencoder	Unsupervised feature algorithm that discovers correlation between features and extracts low dimensional representation using backpropagation to reconstruct sensor sample	Reduce feature dimensionality, minimise undesirable activities and extract hierarchical features. Learn identity approximation and compressed version to select the most suitable feature vectors
(Q. Song, Y. J. Zheng, Y. Xue, W. G. Sheng, & M. R. Zhao, 2017; A. Wang, et al., 2016)	Denoising Autoencoder	Generative model for partial reconstruction of raw sensor input corrupted by adding stochastic mapping term	Learn robust and compressed representation of features from raw sensor data
(Harasimowicz, 2014; M. Hasan & Roy-Chowdhury, 2015; Y. Li, et al., 2014; H. Liu &	Sparse Autoencoder	Introduce sparsity penalty to Autoencoder hidden units to extract robust and compressed features from the visible units	Extract high-level features from high-dimensional sensor data and select the most suitable feature by sparsity penalty to

References	Methods	Description	Advantages
Taniguchi, 2014; Lukun Wang, 2016)			the reconstructed inputs sensor
(Sourav Bhattacharya, et al., 2014; J. Guo, et al., 2014; M. Zhang & Sawchuk, 2013; Q. Zhu, et al., 2015; Y. Zhu, et al., 2010)	Sparse Coding	The techniques help to extract salient features and convert feature vectors for human activity recognition from raw sensor data into linear vectors	Enable location of optimal feature, reduce computational complexity and time, and speed up data annotation from unlabelled data
(Feng, et al., 2017; Jänicke, Tomforde, & Sick, 2016; L. Liu, Cheng, Liu, Jia, & Rosenblum, 2016; X. M. Wang, et al., 2016)	Stacked Deep Gaussian models	Deep fusion of generative and probabilistic models for nonlinear transformation and adaptive extraction of salient and robust features from sensor data.	Reduce number of parameters and model complexity during feature extraction. Furthermore, helps to convert high dimensional vectors to enhance complex activity detection

4.2 Discriminative Deep Learning Methods

Discriminative feature learning algorithms are modelled with posterior distribution classes to provide discriminative powers for activity classification and recognition. In recent years, there has been a tremendous growth in the amount of activity recognition that deploys the use of discriminative deep learning methods. The methods traverse from Convolutional Neural Network to Recurrent Neural Networks. Researchers in ubiquitous sensing have proposed different algorithms in this regard. In this section of the review, we discuss these implementations for human activity recognition using mobile and wearable sensor data.

4.2.1 Convolutional Neural Networks

A comprehensive implementation of Convolutional Neural Network (CNN) for human activity recognition using mobile phone sensor data was reported by (C. A. Ronao & S.-B. Cho, 2016; Ronaoo & Cho, 2015). In their study, Convolutional Neural Network was deployed to extract hierarchical and translational invariant features from accelerometer and gyroscope sensor data and activity details classified using Multinomial Logistic regression (SoftMax). However, the method failed to capture temporal variance and change in complex activity detail and generalisation to different activity models. Furthermore, intra-class and inter-class variations can be solved by incorporating time-frequency convolution which was not implemented in the study. In study by Yuqing Chen and Xue

(2015), instead of developing new CNN architecture modified the convolutional kernel using transfer learning to suit the tri-axial characteristics of acceleration signal for human activity recognition. While Charalampous and Gasteratos (2016) examined the use of the convolutional neural network for online deep learning feature extraction using the whole data sequence. Moreover, they introduce Viterbi algorithm using optimisation criterion and a network of computational nodes in hierarchical form to increase performance of the network. However, the proposed approach applied entire sample of the sensor dataset to implement the CNN and this may increase the computation time for mobile and wearable devices implementation. On the other hand, Ha, Yun, and Choi (2015) proposed a 2-D kernel convolutional neural network to capture local dependencies over time and spatial dependencies over sensors and this is important where multiple sensors are attached to different part of the body. When using 1-D kernel convolution, it will be difficult to capture features from different sensor modalities. The use of a convolutional neural network can also predict the relationship between physical exercises and sleep pattern using accelerometer and gyroscope sensors. In recent study, Sathyanarayana, Joty, Fernandez-Luque, Ofli, Srivastava, Elmagarmid, Taheri, et al. (2016) observed that convolutional neural network outperformed handcrafted features in terms of robust feature generation, high dimensional data and classification accuracy when applied to predict the link between exercises and sleep. Furthermore, similar studies have comparatively explore the performances of convolutional neural network and handcrafted features(Egede, Valstar, & Martinez, 2017; H. Gjoreski, Bizjak, Gjoreski, & Gams, 2015). The experimental analysis showed convolutional neural network conveniently outperform handcrafted features using sensor data generated by wearable devices attached to the wrist for human activity recognition and automatic pain detection during intensive sports activities. However, wrist sensor placement produce irregular movement pattern and it is challenging to ascertain best feature combinations to achieve higher performance accuracy(M. Gjoreski, Gjoreski, Luštrek, & Gams, 2016) for such location placement. Therefore, the results obtain by the comparative analysis cannot be active generalised.

Implementation of deep learning algorithm on low-power wearable devices was recently reported in (D. Ravi, C. Wong, B. Lo, & G. Z. Yang, 2016). They proposed a temporal convolutional neural network that limits the number of hidden layer connections with few input nodes to avoid computational complexity and enable real-time activity recognition. Furthermore, the authors applied spectral representation of the inertial sensor to achieve invariance to sensor placement, orientation and data collection rate. The authors later reported successive implementation combined handcrafted features to reduce computation time and enhance on-board wearable devices implementation

(Ravì, Wong, Lo, et al., 2017). In other way, scale invariant features and local dependencies can also be achieved through weight sharing in convolutional layer (Zeng, et al., 2014). Weight sharing helps to reduce the number of training parameters and computational complexity as closely related filters share similar weights. The issue of computational complexity of convolutional neural network algorithm implemented on low power devices was also analysed by (Jiang & Yin, 2015). The sensor data were transferred and transformed into activity image that has descriptive information about the data. The activity image is then transferred to the deep convolutional neural network to extract discriminative features. They noted that to reduce computational complexity, there is a need to adopt carefully chosen techniques such as feature selection and extraction, sensor selection and use of frequency reduction.

For full implement of automate activity recognition techniques for wearable, Vepakomma, De, Das, and Bhansali (2015) proposed "*A-Aristocracy*", a wristband platform to recognise simple and complex activity using a Deep Neural Network (DNN) classifier for the elderly health monitoring. The propose platform was tested for its performance on detection of daily living and instrumental activity of daily living (cooking, washing plates, doing laundry) (ADL/IADL). The use of wearable sensors ensures the privacy of the elderly are maintained, which is a big issue when camera-based sensors are deployed for activity recognition. Moreover, the work employed affordable wearable devices and multimodal information such as locomotion sensing, environmental condition and contextual location signal sensing to achieve high recognition accuracy. However, the study only used a Deep Neural Network with two layers for classification and extracted statistical and manual features defeating the purpose of automatic feature extraction. Sheng, et al. (2016) proposed quick and short time activity recognition using convolutional neural network for wearable devices. Long time activities comprise series of short-term activity which is segmented using short window length. Therefore, by constructing an over-complete pattern library of long time activities into short time activities using sliding window techniques, feature extraction was implemented offline and learning for recognition was performed online to ensure real-time and continuous activity description. However, the use of short time window length may result in loss of vital information for complex activity recognition (O. Banos, et al., 2015).

Autism Spectrum Disorder can affect the functional ability and activity performance by individuals, social interaction and communication ability. Recognition of such activities can help seamless management of the condition. However, detection of stereotypical motor movement (SMM) is challenging due to intra-subject and inter-subject variability, and may portray different degree of mental and physical health behaviour. For this, the convolutional neural network has been utilised to learn movement such as hand tapping, body rocking or simultaneous combination of body movement to detect stereotypical motor movement(Rad, et al., 2015; Rad & Furlanello, 2016). In the same way, studies conducted by (Castro, et al., 2015; Singh, Arora, & Jawahar, 2016) developed the first person and egocentric activity recognition using the wearable sensor. They combined contextual information and egocentric cues to capture human motion and extract robust and discriminative features using the convolutional neural network. The incorporation of cues and contextual information enable the techniques to capture time-dependent activities and variation in viewpoints.

Conversely, J. Zhu, Pande, Mohapatra, and Han (2015) examined how features extracted by a convolutional neural network can lead to the high estimation of energy expenditure during intensive physical exercises. Energy expenditure estimations enable tracking of personal activity to prevent chronic diseases common in individuals living a sedentary lifestyle. Combining accelerometer sensor and heart rate data, they developed online mechanisms to track daily living activity. Energy expenditure prediction was done on the feature extracted using a backpropagation neural network. However, the dataset used for prediction were collected from sensors placed at the waist which does not indicate movement location. Therefore, there is need to test data collected from sensors placed on the wrist, chest or ankle that accurately detect and monitor total body movements. G. Liu, et al. (2016) modelled binary sensor based human activity recognition by converting the sensor value into a binary number and extracting discriminative features with convolutional neural network. The far-reaching effect of the study is the ability to reduce computational time using fewer binary values during feature extraction from sensor data. Gait assessment based Convolutional Neural Network in a patient with Sclerosis was presented by (J. Q. Gong, Goldman, & Lach, 2016) with body-worn sensors. Convolutional Neural Networks were implemented to learn the temporal and spectral association among the multichannel time series motion data and learn holistic gait patterns for robust and efficient feature representation. In related study, Eskofier, et al. (2016) propose deep learning algorithm for assessment of movement disorders for patients with idiopathic Parkinson diseases. Patients were attached with inertial measurement unit sensor nodes to collect accelerometer data and extract salient features with two convolutional neural network layers and achieved 90.9% accuracy. However, due to the limited number of sensor data used for training the Convolutional Neural network, it may be challenging to generalise the performances accuracy achieved.

In some cases, convolutional neural network are optimised with classical machine learning techniques such as meta-heuristic algorithms to model hyper-parameter tuning to obtain higher accuracy. This techniques were recently implemented for detection of Parkinson disease and measurement of calories consumption to combat obesity and recommend physical activities (Pereira, Pereira, Papa, Rosa, & Yang, 2016; Pouladzadeh, Kuhad, Peddi, Yassine, & Shirmohammadi, 2016). In a related research for the elderly, Yin, Yang, Zhang, and Oki (2016) proposed the cascade convolutional neural network for monitoring of heart-related diseases using impulse radio ultra-wideband radar data. Different convolutional neural network modules were implemented to extract robust ECG features and impulse radio ultra-wideband radar feature, which are then combined to form a cascade to distinguish normal heart bits from abnormal ones. The essence of the cascade is to take care of the different sampling rate and dimensionality of the various data source. Also, Junming Zhang and Wu (2017) proposed the use of the convolutional neural network for automatic stage sleep classification using electrocardiography data.

Other similar Convolutional Neural networks approach were lately implemented for automatic data labelling, variable sliding window segmentation and multi-sensor and multi-channel time series fusion. For instance, Zebin, Scully, and Ozanyan (2016) introduce multichannel sensor time series to acquire sensor data from body-worn inertial sensors. The authors modelled feature extraction using a convolutional neural network and monitored different hyperparameter setting at the pooling layer, rectified linear units and max pooling to achieve high accuracy. R. Yao, et al. (2017) proposed the use of CNN for dense labelling in human activity recognition. The use of dense labelling provides an approach to avoid missing information, and the algorithm was implemented using publicly available datasets with an overall accuracy of 91.2%. Another important applications of convolutional neural network is in multi-sensor fusion for human activity detection. Fusion of multiple sensor are essential for enhanced activity recognition rate (R. Gravina, et al., 2017). However, many issues are yet unresolved, such as imprecision and uncertainty in measurement, noise and conflicting correlation, high data dimensions and the best techniques to select the fusion level. To that effect, Jing, et al. (2017) propose adaptive multi-sensor fusion using the deep convolutional neural network. The proposed techniques learn features and optimise the combination of sensor fusion level such as extraction, selection, data, features, and decision fusion levels to build complex recognition patterns for higher activity detections. These processes go through from the lower layer of the network to the higher layer and implement the robust feature extraction process.

Automatic feature extraction in wearable sensors with the convolutional neural network provide means to monitor beach volley ball players' skills from a tri-axial accelerometer (Kautz, et al., 2017). To achieve that, the authors deploy data collected from 30 subjects wearing sensors attached to the right hand with a thin wristband. However, the proposed architecture of the CNN suffered from overfitting as it performed better on training data than on testing data. Therefore, the use improve regularisation techniques, increase the training datasets and use batch normalisation (Ioffe & Szegedy, 2015) may enhance the performance of the proposed model. Moreover, adding artificial noise to the data may also improve the prediction accuracy (G. E. Hinton, et al., 2012).

4.2.2 Recurrent Neural Networks

Human activity recognition is a classical time series classification problem made up of complex motor movements and vary with time. Capturing the temporal dynamic in movement pattern will help to model complex activity details and enhance the performance of recognition algorithms. Convolutional neural network architecture can only extract translational invariant local features but become ineffective when modelling global temporal dependencies in sensor data. However, Recurrent Neural Network (RNN) is naturally designed for time series data in which sensor data is a prominent part.

Recently various studies have explored different recurrent neural network models for modelling human activity recognition. For instance, studies such as (Yuwen Chen, Zhong, Zhang, Sun, & Zhao, 2016; X. Ma, Tao, Wang, Yu, & Wang, 2015) proposed long short term memory (LSTM) for feature extraction to recognise activity of daily living using WISDM data, a publicly available dataset by Wireless Sensor Data Mining Lab (Kwapisz, Weiss, & Moore, 2011) and achieved a classification accuracy of 95.1%. Despite the high performance obtained, the result cannot be generalised due to the simplicity of the specified activities and small sample sizes of the dataset. Therefore, larger datasets are required to improve the robustness of the algorithm. Large-scale study on the prediction of activity of daily living was examined by (Moon & Hamm, 2016) with Long Short Term Memory to capture the randomness in activity patterns and model the temporal dependencies using multi-step look ahead approach. Long short memory provides the possibility to automatically detect and characterise eating pattern using the wearable necklace, and early or progressive detection of activities (S. Ma, Sigal, & Sclaroff, 2016; Nguyen, Cohen, Pourhomayoun, & Alshurafa, 2016). However, issues on the modelling of motion movement of head and neck are difficult as piezoelectric sensors do not detect such motions. Furthermore, Long short term memory methods provide technique to rank activity progression and penalise incorrect activity prediction that may lead to serious consequence especially for detection of fall in elderly (S. Ma, et al., 2016).

Inoue, et al. (2016) investigated the use of the deep recurrent neural network for human activity recognition in real time scenario. They looked at the best combination of architecture and optimal parameter values for increased performance. The authors noted that, increasing the layer of deep RNN will greatly increase computational time and memory usage and recommend a three-layer architecture for optimal performance. To reduce memory usage, (Edel & Köppe, 2016) developed optimised binary version of Bidirectional LSTM for human activity recognition in a resource constrained environment such as mobile or wearable devices. The extended version of Bidirectional LSTM (Graves & Schmidhuber, 2005) achieved real-time and online activity recognition by applying binary values to the network weight and activation parameters.

Subsequent studies introduced other aspects of the recurrent neural network. Notably, Palumbo, Gallicchio, Pucci, and Micheli (2016) proposed the Recurrent Neural Network for real-time human activity recognition trained with *echo state network* leveraging smartphones and Reciprocal Received Signal Strength (RSS). Echo State Network is a Recurrent Neural Network with a non-trainable reservoir and linear readout in which the weights are randomly generated during training (Rodan & Tino, 2011). However, a number of issues have deterred the practical application of the Echo State Network. These include the unclear properties of the reservoir and lack of training strategies to achieve optimal performance but rely on a game of chance. Furthermore, Choi, Schuetz, Stewart, and Sun (2016) develop the Gated Recurrent Unit Model (Cho, et al., 2014) to detect heart failure from clinical time series data. Gated recurrent unit is an RNN model that is similar in structure to LSTM but with simple parameter update and recently achieved superior results in similar classification tasks (Zaremba, 2015).

4.2.3 Other Discriminative Deep Learning Models

Various studies have also proposed other discriminative feature extraction methods for human activity recognition. For instance, studies in (Ponce, de Lourdes Martínez-Villaseñor, & Miralles-Pechúan, 2015; Ponce, Martínez-Villaseñor, & Miralles-Pechuán, 2016) proposed and analysed the use of Artificial Hydrocarbon Network (AHN) for human activity recognition. Artificial Hydrocarbon Network is an algorithm inspired by an organic chemistry that use heuristic mechanism to generate organise structure to ensure modularity and stability in activity recognition. The algorithm is tolerant to noisy sensor data. However, it needs to be combined with heuristic feature extraction and selection techniques to increase recognition time. Similarly, Rogers, Kelleher, and Ross (2016) exploited deep neural language model for the discovery of interleaved and overlapping activities. The model builds hierarchical activities and captures the inherent complexities in activity details. Similarly, Hongqing Fang, He, Si, Liu, and Xie (2014) initiated backpropagation techniques to train feedforward neural for complex human activity recognition in smart home environment. Although the algorithm outperformed the Hidden Markov Model and Naïve Bayes, it requires combined handcrafted feature extraction for high-performance accuracy. Y.-L. Chen, et al. (2016) proposed manifold elastic network for feature extraction and dimensionality reduction by mapping motion sensor data from high dimensional to low dimensional subspace through minimization algorithm. **Table 4** summarises recently discriminative model for human activity recognition and their advantages.

Table 4: Discriminative Deep Learning Methods for human Activity Recognition

References	Methods	Description	Advantages
(Castro, et al., 2015; Charalampous & Gasteratos, 2016; Yuqing Chen & Xue, 2015; Eskofier, et al., 2016; M. Gjoreski, et al., 2016; J. Q. Gong, et al., 2016; Ha, et al., 2015; Jiang & Yin, 2015; Jing, et al., 2017; Kautz, et al., 2017; G. Liu, et al., 2016; Page, et al., 2015; Pereira, et al., 2016; Pouladzadeh, et al., 2016; Rad, et al., 2015; D. Ravi, et al., 2016; C. A. Ronao & S.-B. Cho, 2016; Ronaoo & Cho, 2015; Sathyanarayana, Joty, Fernandez-Luque, Ofli, Srivastava, Elmagarmid, Taheri, et al., 2016; Sheng, et al., 2016; Singh, et al., 2016; Vepakomma, et al., 2015; Yang, et al., 2015; R. Yao, et al., 2017; Yin, et al., 2016; Junming Zhang & Wu, 2017; Zheng, Ling, & Xue, 2014; J. Zhu, et al., 2015)	Convolutional Neural Network	Multilayer neural network that combines convolution and pooling operations to extract translation invariant, temporally correlated and hierarchical feature vectors from sensor data. The architecture use convolutional operation to handle and extract local features and cancel the effect of translation and displacement in sensor data	Extract hierarchical and translational invariant features from sensor data with or without preprocessing to enhance performance and recognition accuracy
(Y. Chen, et al., 2016; Inoue, et al., 2016; S. Ma, et al., 2016; X. Ma, et al., 2015; Moon & Hamm, 2016; Nguyen, et al., 2016)	Long Short Term Memory	Recurrent neural network (RNN) that incorporate memory block to overcome backpropagation problem and detect activities with long-term temporal dependencies	Capture temporal dependencies and complex activities dynamic in raw sensor data
(Edel & Köppe, 2016)	Binarise-Bidirectional Long Short Term Memory	Recurrent Neural Network in which the network parameters are binary values trained and evaluated with bits logics	Has low computational complexity and applicable in resource constrained environment such as mobile and

References	Methods	Description	Advantages
			wearable devices with low energy resources. The extracted features are invariant to distortion and transformation
(Choi, et al., 2016)	Gated Recurrent Unit	Recurrent Neural Network with reduced parameter for detection and recognition of time sensitive events	Gated Recurrent unit has fewer parameters and easy to train
(Ponce, Miralles-Pechuán, & Martínez-Villaseñor, 2016)	Artificial Hydrocarbon Network	Nature inspired meta-heuristic and chemical organic algorithm that organise activity details in modules	Ability to model noisy and unlabelled data and also robust to sensor data characteristics and data point
(Rogers, et al., 2016)	Deep Neural Model	A form of deep learning for modelling natural language problem. The algorithm is trained to approximate model distribution by taking encoding of sensor distribution and produce posterior distribution of all possible values	Can handle problem of multiple activities occurring in parallel (interleaved activities)
(Y.-L. Chen, et al., 2016)	Manifold Elastic Network	Dimensionality reduction methods that encode local geometry to find best feature representation in raw sensor data	Minimise error mechanisms to select appropriate feature subspace

4.3 Hybrid Deep Learning Methods

Various research efforts have been geared toward obtaining robust and effective features for human activity recognition by combining generative, discriminative or both methods. From the available literature on hybrid implementation, the convolutional neural network seems to be the best choice method for many studies to be hybridised with other generative or discriminative models for human activity recognition. For instance, Convolutional Neural Network and Denoising Autoencoder (G. Ma, Yang, Zhang, & Shi, 2016), Convolutional Neural Network and Sparse Coding (S. Bhattacharya & Lane, 2016), Convolutional Neural Network and Recurrent Neural Network

(Ordóñez & Roggen, 2016; Sathyanarayana, Joty, Fernandez-Luque, Ofli, Srivastava, Elmagarmid, Taheri, et al., 2016), Convolutional Neural Network and Restricted Boltzmann Machine (J. Gao, Yang, Wang, & Li, 2016).

In most of these studies, the convolutional neural network is incorporated to produce hierarchical and translational invariant features. To reduce the source of instability and extract translational invariant features, J. Gao, et al. (2016) introduce the centred factor Convolutional Restricted Boltzmann Machine (CRBM) while in Sarkar, et al. (2016), a combination of Deep Belief Network and convolutional neural network were examined for activity recognition in prognostic and health monitoring related services. The authors compare the performance using electroencephalogram sensor data with deep learning outperforming handcrafted features. However, the result deteriorated when it was tested on four recognition tasks due to the limited amount of training and testing data. Recently, other studies incorporated the convolutional neural network and sparse coding to produce sparse representation and reduce computational time. This can be seen in recent work by S. Bhattacharya and Lane (2016), which proposed sparse coding-based convolutional neural network for mobile based activity recognition. To reduce computation time, memory and processor usage, they introduced sparsification of the fully connected layer and separation of the convolutional kernel. The techniques ensure full optimisation of CNN to be implemented for mobile devices.

Another work for hybridization of deep learning methods for robust features extraction was reported in (G. Ma, et al., 2016). In the work, the authors proposed the fusion of features extracted with deep autoencoder to obtain more abstract features. While Khan and Taati (2017)proposed a channel-wise ensemble of autoencoder to detect unseen falls using wearable devices. In the study, stacked autoencoder was used to learn accelerometer and gyroscope data separately, using interquartile range and then training a new autoencoder on data with no outliers to accurately identify unseen fall. Ijjina and Mohan (2016) developed ensemble deep learning approach based on Convolutional Neural network by altering the inputs and weights of network of each convolutional neural network to create network structures variabilities and then combined the results with different ensemble fusion techniques. Recently, an ensemble of diverse long short term memory (Guan & Ploetz, 2017) was evaluated on publicly available datasets for human activity recognition. The proposed method outperformed other methods in real life activity prediction.

To recognise and detect complex activity details, there is a need to capture spatial and temporal dependencies involve in human activity recognition. The convolutional neural network and recurrent neural network are important deep learning methods in this regard. The techniques are common in multimodal and multi-sensor activity recognition frameworks. X. Li, et al. (2017) investigated the use of CNN and LSTM for recognition of concurrent activities. The authors introduced encoder to output binary code prediction that denotes whether the activity is in progress or not in progress. Furthermore, the architecture can accept input from the sensor of different modalities. Similarly, Ordóñez and Roggen (2016) proposed a convolutional neural network and long short term memory to automatically learn translational invariant features and model temporal dependencies in multimodal sensor comprise of accelerometer and gyroscope sensor. The pooling layer in the network was replaced with a recurrent layer (LSTM) that models the temporal sequence, whereas the final layer is the SoftMax regression that produces the class prediction. The technique was compared with baseline CNN using OPPORTUNITY and Skoda datasets with $0.61F_1$ score performance. The ensemble of Convolutional neural network and bidirectional long short term memory (BLSTM) were proposed for health monitoring using the accelerometer and acoustic emission data. CNN extract local features, and while BLSTM encodes temporal dependencies and model sequential structure, past and present contextual information (R. Zhao, Yan, Wang, & Mao, 2017).

Furthermore, other authors have also proposed fusion along multimodal and multi-sensor lines. For instance, Song, et al. (2016) proposed the fusion of the video and accelerometer sensor model using the convolutional neural network and long short term memory. CNN extract spatial-temporal features from video data while the LSTM models temporal dependencies features from the accelerometer and gyroscope. These feature vectors were integrated using a two-level fusion approach for egocentric activity recognition. However, the result obtained in multimodal fusion performed below expectation due to the small number of training examples. In Neverova, et al. (2016), the authors proposed the recurrent neural network and convolutional neural network to extract feature vectors optimised with shift-invariant dense mechanism to reduce computation complexity. In order to develop effective deep learning fusion approach, Nils Y Hammerla, Halloran, and Ploetz (2016) explored the effect of hyper-parameter setting such as regularisation, learning process, the number of architecture on the performance of deep learning for human activity recognition. The authors concluded that hyper-parameters have great impact on the performance of deep architectures and recommend extensive hyper-parameter tuning strategies to obtain enhance activity recognition rate. To develop a multi-fusion architecture of CNN and LSTM, F. J. O. Morales and Roggen (2016) examined the effect of transfer learning at the network kernel between users, applications domains, sensor modalities and sensor placements in human activity recognition. They noted that transfer learning greatly reduced training time and are sensitive to sensor characteristics, placement and motion dynamic. They utilised the above automatic feature representation method to develop a hybrid of CNN and LSTM for extraction of robust features for human activity recognition in a wearable device. In Sathyanarayana, Joty, Fernandez-Luque, Ofli, Srivastava, Elmagarmid, Taheri, et al. (2016), CNN-LSTM was used to model the impact of sleep on physical activity detection with *actigraphy* dataset. CNN models robust feature extraction while LSTM was used to build sleep prediction. Alternatively, a convolutional neural network with Gated Recurrent Unit (GRU) was proposed by (S. Yao, Hu, Zhao, Zhang, & Abdelzaher, 2016) for activity recognition and car tracking using accelerometer, gyroscope and magnetometer data. CNN and GRU were integrated to extract local interaction among identical mobile sensor, merged into global interaction and then extract temporal interaction to model signal dynamics.

Various studies have proposed fusion of deep learning model and handcrafted features for human activity recognition. Fusion of handcrafted features and deep learning are effective for increased recognition accuracy, real time and on-board human activity recognition in wearable devices. Furthermore, the techniques allow extraction of interpretable feature vectors using spectrogram and to capture intensity among data points(Ravì, Wong, Lo, et al., 2017). Interestingly, some studies have also found that such fusion are important means to model lateral and temporal variation in activity details by adaptively decomposing complex activity into simpler activity details and then train the algorithm using radius margin bound for network regularisation and improve performance generalisation(Liang Lin, et al., 2015). In recent work, Alzantot, Chakraborty, and Srivastava (2017) explored generation of artificial activity data by fusion of mixture density network and long short term memory. The approach was proposed to resolve the issue of lack of training data using mobile phones and discriminate robust feature vectors. Developing protocol to collect large training data for human activity recognition project is very tedious and may result to privacy violations. Therefore, the study generated synthetic data to augment the training sensor data generated using mobile phone. Moreover, the developed fusion of mixture density networks and long short term memory will help to reduce reliance on real training data for evaluation of deep learning. **Table 5** summarises the different hybrid deep learning based feature extraction techniques for human activity recognition.

Table 5: Hybrid deep learning methods for human activity recognition

References	Methods	Descriptions	Advantages
(J. Gao, et al., 2016; Sarkar, et al., 2016)	CNN, RBM	Propose integration of Deep Belief Network and Convolutional Neural Network for real-time multimodal feature extraction in unconstrained environment	Provide automatic feature extraction and selection without extensive pre-processing procedure
(S. Bhattacharya & Lane, 2016)	Sparse coding and Convolutional Neural Networks	Automatically produce compact representation of features vectors from raw sensor data for mobile based activity recognition.	The use of sparse coding helps to reduce computation time and memory usage by utilising sparsification approach to separate fully connected layer and convolutional kernel.
(Khan & Taati, 2017)	Ensemble of Channel-wise Autoencoder	Channel-wise autoencoder algorithms fusion of autoencoder trained separately with accelerometer and gyroscope sensor data and combine with reconstruction error values	Automatically learn generic features from raw sensor data.
(Ijjina & Mohan, 2016)	Ensemble of Deep Convolutional Neural Networks	Develop fusion of extracted features of homogenous CNN architecture built by alternating the initialisation of the network parameters.	Achieve high model diversity and enhance performance generalisation
(Guan & Ploetz, 2017; X. Li, et al., 2017; F. J. O. Morales & Roggen, 2016; Neverova, et al., 2016; Ordóñez & Roggen, 2016; Sathyanarayana, Joty, Fernandez-Luque, Ofli, Srivastava, Elmagarmid,	Convolutional Neural Network (CNN) and Recurrent Neural Networks (RNN)	Propose multimodal and spatial-temporal feature extraction with CNN and LSTM for concurrent activity recognition	Suitable for multimodal, Multi-feature and multi-sensory for recognition of complex and concurrent activity details

References	Methods	Descriptions	Advantages
Taheri, et al., 2016; Song, et al., 2016; R. Zhao, et al., 2017)			
(S. Yao, et al., 2016)	CNN, Gated Recurrent Unit (GRU)	Integrate convolutional neural network and Gated recurrent unit that exploits local interaction within activities and merges them into global interaction to extract temporal relationship	Provide low energy consumption and low latency services for implementation in mobile and wearable devices. Gated recurrent unit has expressible terms with reduce network complexity for mobile based implementation
(Liang Lin, et al., 2015; D. Ravi, C. Wong, B. Lo, & G.-Z. Yang, 2016)	CNN, Conventional feature	Combine deep feature learned with CNN and statistical feature for real-time mobile based implementation of activity recognition. Also, the fusion provides effective means of decomposing complex activity into sub activities by modelling temporal variation and extract transition invariant features.	Enable real-time on-board implementation with reduced feature vectors. The method can handle optimal decomposition of complex activity details and enhance generalisation ability deep learning algorithms for human activity recognition.
(Alzantot, et al., 2017)	LSTM, Mixture Density Network	Deep stacked long short term memory for generation and discriminating artificial sensory data in human activity recognition	Distinguish between real and synthetic data set to improve privacy in data collection

5. Classification Algorithms and Performance Evaluation of Human Activities

Classification is a vital part of human activity recognition processes. Classification involves training, testing and use of evaluation metrics to measure the performance of the proposed algorithms. Over the years, different classifiers have been implemented in human activity recognition to categorise activity details during training and testing. The commonly used classifiers are the Support Vector Machine (SVM), Hidden Markov Model (HMM), K-

Nearest Neighbour (KNN), and Decision Tress, Neural Network (NN). In deep learning based human activity recognition, most studies favour multinomial logistic regression (SoftMax) (Ordóñez & Roggen, 2016; D. Ravi, et al., 2016; Song, et al., 2016) or Hidden Markov Model (Alsheikh, et al., 2015) trained with the deep neural network for activity recognition. The training process extracts the feature vectors that are fed to the classifiers through fully connected layers to yield probability distribution classes for every single time step of the sensor data (D. Ravi, et al., 2016). The performance of the extracted feature vectors is evaluated with pre-set evaluation metrics and access the recognition accuracy and computational complexity. Performance metrics such as accuracy, precision, recall and F-measure provide essential information to access recognition ability of the features vectors. In this section, training, classifiers and performance evaluation metrics of human activity recognition system with deep learning methods are explained. We begin by presenting the training of both deep learning methods and classification inference algorithm and then the performance evaluation metrics for human activity recognition.

5.1 Training

Early works using deep neural networks were trained with gradient descent optimisation where the weights and biases are adjusted to obtain low-cost function. However, training neural network with such strategies will cause its output to get stuck in local minima due to the high number of parameters involve. To solve the problem, Geoffrey E Hinton, et al. (2006) introduced the greedy layer-wise unsupervised pre-training techniques in which the neural network algorithm is trained one layer at a time then the deep architecture is fine-tuned in a supervised way with gradient optimisation. In his work, G. Hinton (2010) showed how to train deep learning algorithm and set the different hyperparameter settings. Deep learning researchers adopt these strategies when validating their methods.

In training deep learning algorithms, the main aim is to find network parameters that minimise reconstruction errors between inputs and outputs (Erfani, et al., 2016). Using the pre-training and fine-tuning, the networks will learn to extract salient features from sensor data which is then passed to multi-linear logistic regression (SoftMax Regression) or any other classifiers to discriminate the activity details. Therefore, numerous regularisation methods have been proposed to modify the learning algorithm to reduce generalisation errors by applying hyper-parameter settings to control the network behaviour. According to G. Hinton (2010), these hyper-parameters include the values of learning rate, momentum, weight decay, initial values of the weight and weight update mechanism. Others are pre-training and fine-tuning parameter values, optimisation procedures, activation functions, sizes of mini-batch, training epochs, network depth and pooling procedure to use when training convolutional neural networks. In deep learning based human activity recognition, different studies specify varying values of these hyper-parameters relying on the network and size of the training sensor data. Different hyper-parameter settings that were recently implemented for mobile and wearable sensor based human activity recognition is shown in **Table 6.** Here we present brief explanations of these hyper-parameters with examples of value settings in recent works.

Learning rate provides the value that shows how much the network has learned during neural network training iterations. The learning rates need to be initialised in such a way that it is not too large or small. A large value will cause the network weight to explode; a value between 0.0001 multiplied by the weight is recommended. Past studies in human activity recognition using mobile and wearable sensor implement varying values that range from 0.0001 (Castro, et al., 2015), 0.001 (Alsheikh, et al., 2015; Kautz, et al., 2017), 0.01 (Eskofier, et al., 2016; C. A. Ronao & S.-B. Cho, 2016), 0.0 5 (Jing, et al., 2017) to as high as 0.1 (S. Ma, et al., 2016).

Momentum (Qian, 1999) increases the velocity of learning and the rate of convergence of deep neural networks. Previous studies in deep learning based human activity recognition adopted the recommended values between 0.5 to 0.99 (Kautz, et al., 2017; C. A. Ronao & S.-B. Cho, 2016). The *size of mini-batch* is another important parameter used to avoid overfitting. The mini-batch size divides the training data into small size of 10 to 100 training set, and then total gradients are computed using these sizes. When the network is trained with stochastic gradient descent, there is need to maintain relative sizes to reduce sampling bias. In activity recognition, too large mini-batches will be the equivalent of using large window size, and therefore may increase computation time and miss important activity details. Therefore, factors such as the size of data and implementation platform play vital roles in choosing the size of mini-batch (C. A. Ronao & S.-B. Cho, 2016).

Another key insight for improving deep learning model is the use of weight regularisation. Regularising large weight in deep learning to avoid overfitting is imperative during training due to large parameter updates. Overfitting is monitored by measuring the free energy of training data (G. Hinton, 2010). Previous studies have proposed various regularisation techniques for training deep neural networks. For instance, *Dropout* (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014) randomly deletes half of the feature values to prevent complex co-adaptation and increase generalisation ability of the model. Dropout regularisation technique were recently improved by (Wan, Zeiler, Zhang, Cun, & Fergus, 2013) into DropConnect by randomly dropping weight vectors instead of the activation function. However, Dropout is still the most popular and is utilised by the majority of the studies reviewed (Alsheikh, et al., 2015; Jing, et al., 2017; Ordóñez & Roggen, 2016) with a probability of dropout ranging from 0.5 to 0.8.

In addition to dropout, *weight decay* techniques such as L1/L2 regularisations prevent overfitting by introducing penalty term for large weights and this help to improve generalisation and shrink useless weights. Studies apply different weight decaying terms with varying values. Also, optimisation techniques such as batch normalisation that compute gradients on whole datasets, stochastic gradient descent (SGD) using each training examples or mini-batch gradient descent that compute update on every mini-batch will further help to reduce invariance of the parameter update (Ruder, 2016). However, batch normalisation is slow and does not allow online weight update. Stochastic gradient provides faster convergence and helps to choose proper learning rate. It is widely applied in deep learning based human activity recognition (Ravì, Wong, Lo, et al., 2017; Vepakomma, et al., 2015; Lukun Wang, 2016).

Other optimisation algorithms have also been implemented for deep learning training. For instance, Adagrad (Duchi, Hazan, & Singer, 2011) apply adaptive learning rate to the network parameter to improve robustness to Stochastic gradient descent, while (Zeiler, 2012) proposed ADADelta that applied adaptive methods to decrease the learning rate. Furthermore, to solve the problem of diminishing weights, algorithms such as RMSProp (Tieleman & Hinton, 2012) and Adaptive Moment Estimation (ADAM) (Kingma & Ba, 2014) were proposed. RMSProp adopts adaptive learning rate to solve the diminishing weights issues by adapting different step size for each neural network weights. ADAM applies an exponentially decaying average of past square gradient with default values ranging from 0.9 to 0.999 and momentum of 8E-10. Adaptive optimisation is important and widely used because of its ability to adapt to learning rate and momentum without manual intervention. Furthermore, Q. Song, et al. (2017) proposed an evolutionary based optimisation algorithm called Ecogeography Based Optimisation (EBO) that adaptively optimises the autoencoder algorithm layer by layer to achieve optimal performance. Another important optimisation technique is the use of early stopping criteria that monitor errors on each validation set and stop when the validation error stops increasing. **Table 6** shows some of the training techniques in some of reviewed studies with their value settings.

Table 6: Sample hyper-Parameter Setting and Optimisation for Deep Learning Training for human activity recognition

Settings	(Ordóñez & Roggen, 2016)	(C. A. Ronao & S.-B. Cho, 2016)	(Castro, et al., 2015)	(Jing, et al., 2017)	(Eskofier, et al., 2016)	(Kautz, et al., 2017)	(S. Ma, et al., 2016)
Learning Rate	0.001	0.01	0.0001	0.05	0.01	0.01	0.1
Momentum	0.9	0.5-0.99	0.9	0.5	0.9-0.999	0.9-0.999	0.9
Size of Mini-batch	100	128	100	20	500	200	100
Dropout	✓	✓	✓	✓	✓	✓	
Activation Function	ReLU, Tanh	ReLU	ReLU	ReLU	ReLU	ReLU	Tanh
Decay Rate	0.9	0.00005	0.0005	0.04	1E-8	1E-8	0.05
Optimisation	RMSProp	SGD	SGD	SGD	ADAM	SGD	
Training Epoch		5000	100000	200			30
Method	CNN-LSTM	CNN	CNN	CNN	CNN	CNN	LSTM

5. 2 Classification

Deep learning algorithms are applied on sensor data to extract discriminative and salient features and then flattened and pass to an inference engine to recognise activities classes. The outputs of the deep neural network model feature at the fully connected layer of the model are connected with classifiers. The most commonly used classifiers are Multinomial Regression (SoftMax) (Alvear-Sandoval & Figueiras-Vidal, 2018; Alzantot, et al., 2017; Guan & Ploetz, 2017; Ordóñez & Roggen, 2016; C. A. Ronao & S.-B. Cho, 2016), Support Vector Machine (Erfani, et al., 2016) or Hidden Markov Model(Alsheikh, et al., 2015) and provide probability distribution classes over activity details. Most of the studies reviewed use SoftMax to model the probability of the activity classes.

SoftMax is a variant of logistic regression that model Multiclass classification(J. Gao, et al., 2016; O'Donoghue & Roantree, 2015) using cost minimization approach. Therefore, given training sets $\{(x^{(i)}, y^{(i)}), (x^{(i)}, y^{(i)}), \ldots\ldots\ldots\ldots (x^{(m)}, y^{(m)})\}$ with corresponding m label examples, where $y^{(i)} \in \{1, 2, 3\ldots\ldots, k\}$ and x is the input feature space. The SoftMax parameters are trained by minimising the cost function and then fine-tuned to minimise the likelihood function and improve adaptability. The cost function with the decay terms is as stated below.

$$J(\theta) = \frac{1}{m}\left[\sum_{i=1}^{m} \sum_{j=1}^{k} 1\{y(y-j)\} \log \frac{\ell^{\theta_j^T x^{(i)}}}{\sum_{i=1}^{k} \ell_i^{T\, x^{(i)}}}\right] + \frac{\lambda}{2} \sum_{i=1}^{k} \sum_{j=0}^{n} \theta_{jk}^2 (\lambda > 0) \tag{1}$$

The fine-tuned algorithm through backpropagation to improve performance is given as:

$$p\left(y^{(i)} = k / x^{(i)}; \theta\right) = \frac{\exp\left(\theta^{(k)T} x^{(i)}\right)}{\sum_{j=1}^{k} \exp\left(\theta^{(j)T} x^{(i)}\right)} \tag{2}$$

The above equation provides the probability of the activity classes with possible values of labels (Yan, et al., 2015). Also, (C. A. Ronao & S.-B. Cho, 2016) noted that the last layer of the convolutional neural network that infers activity classes is given as:

$$p(c / p) = \arg\max_{c=C} \exp \frac{\exp\left(p^{L-1} w^L + b^L\right)}{\sum_{k=1}^{NC} \exp\left(p^{L-1} w^k\right)} \tag{3}$$

Where c is the activity class, L is the last layer index of the convolutional neural network (CNN), and NC is the total number of activity classes.

5.3 Evaluation Metrics

The performance of features representation for human activity recognition using mobile and wearable sensors is evaluated with pre-set evaluation techniques. Criteria such as accuracy, computation time and complexity, robustness, diversity, data size, scalability, types of sensor, users and storage requirements are used to evaluate how the features extracted, and classifiers perform in relation to other studies. Alternatively, deep learning methods can also be evaluated on how varying the hyper-parameters affect their performances during training, filter size, pre-training and fine-tuning, pooling layers and number of temporal sequences (Alsheikh, et al., 2015; Ordóñez & Roggen, 2016; C. A. Ronao & S.-B. Cho, 2016). These parameters evaluation is still an open research challenge to establish their effects on deep learning network performance (Erfani, et al., 2016; Munoz-Organero & Ruiz-Blazquez, 2017).

Like the handcrafted features based human activity recognition methods, deep learning features are evaluated with different performance metrics. Hold-out cross-validation techniques are utilised to test the performance of features representation on different datasets. Hold-out cross-validation techniques include leave-one-out, leave one person out when testing the performance of single-user, 10-fold cross validation, or leave one day out when using data collected for a specific number of days for activity details (Nils Yannick Hammerla, et al., 2015). These different hold-outs cross-validation techniques allow the deep learning training to be repeated a number of times to ensure generalisation across datasets. Different performance evaluation metrics used in the studies review is presented in **Table 7** below.

The most common performance metrics are accuracy, precision, recall, confusion matrices and Receiver Operating Characteristics (ROC) curve. Therefore, the activity can be classified as True Positive (TP), True Negative (TN) when correctly recognised or False Positive (FP) or False Negative (FN) when incorrectly classified. Other performance metrics are derived with True positive or True Negative. These metrics are discussed below:

Accuracy provides the overall correctly classified instances. It is the sum of correct classification divide by the total number of classification.

$$\frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

Precision (Specificity) measures the accuracy and provides the value based on the fraction of the negative instance that are classified as negative.

$$\frac{TP}{TP + FP} \tag{5}$$

Recall measures the performance of correctly predicted instances as positive instances.

$$\frac{TP}{TP + FN} \qquad (6)$$

F-Measure (Score), F-Measure is mainly applied in unbalanced datasets and provides a geometric mean of sensitivity and specificity. F-measure

$$2.\frac{\Pr ecison.\mathrm{Re}\,call}{\Pr ecision + \mathrm{Re}\,call} \qquad (7)$$

Confusion Matrices: Confusion matrices are important performance measure, and the matrix provide the overall misclassifications rate in human activity recognition (Nils Yannick Hammerla, 2015). The known classes are represented with rows while the columns correspond to the predicted classes made by the classifiers. The use of confusion matrices allows the analysis of Null class which is common in Human Activity Recognition and further enables visualisation of the recognition performance of the system.

Receiver Operating Characteristics (ROC) Curve: The ROC curve is also known as precision-recall rate and provides mechanism to analyse the true positive rate against the true negative rate give as (FPR). However, the ROC curve is only suitable for detection model as it depends on the number of True Negative classes and may not be used in imbalance dataset which is common in deep learning based human activity recognition. Metrics such as Equal Error Rate that show the values at which precision is equal to recall, average precision and Area Under the Curve (AUC) the show the overall performance of classifiers and probability that chosen positive instances will be ranked higher than negative instances (Bulling, et al., 2014b; Nils Yannick Hammerla, 2015).

Accuracy, precision and recall are suitable for two classes and balance datasets. For imbalance data, average accuracy, precision and recall are computed for the overall activities. These values are averages of the summation of their individual values.

$$\text{Average accuracy} = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{\left(TP + FP\right)_i} \qquad (8)$$

$$\text{Precision} = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TI_i}, \qquad (9)$$

$$\text{Average Recall} = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TT_i} \quad (10)$$

where *N* is the number of classes, *TI*, the total number of inferred label and *TT* is the ground truth label. However, it has become an issue of contention in deep learning as most of the data are unlabelled data and ground truth labels are missing in most cases. The use of average precision and recall require manual annotation of data which is tedious and laborious especially for mobile based and real time human activity recognition (D. Ravi, et al., 2016; Ravì, Wong, Lo, et al., 2017). Studies adopting deep learning methods test for precision and recall instead.

Table 7: Evaluation Metrics of Deep Learning Methods for Human Activity Recognition

References	Accuracy	Precision	Recall	Confusion Matrix	F_1-Score	ROC/AUC
(Plötz, et al., 2011)	✓	-	-	-	-	-
(Sourav Bhattacharya, et al., 2014)	-	-	-	✓	✓	-
(Al Rahhal, et al., 2016)	✓	✓	✓	-	-	-
(Jokanovic, et al., 2016)	-	-	-	✓	-	-
(Munoz-Organero & Ruiz-Blazquez, 2017)	-	✓	✓	-	✓	-
(Jing, et al., 2017)	✓	-	-	-	-	-
(Alsheikh, et al., 2015)	✓	-	-	-	-	-
(Erfani, et al., 2016)	-	-	-	-	-	✓
(D. Ravi, et al., 2016)	✓	-	-	-	-	-
(L. Zhang, et al., 2015a)	✓	✓	✓	-	-	-
(Ravì, Wong, Lo, et al., 2017)	✓	✓	✓	-	-	-
(Q. Song, et al., 2017)	✓	-	-	-	-	-
(Lukun Wang, 2016)	-	-	-	✓	-	-
(Kautz, et al., 2017)	✓	✓	✓	✓	-	-
(Guan & Ploetz, 2017)	-	-	-	-	✓	
(C. A. Ronao & S.-B. Cho, 2016)	✓	-	-	-	-	-
(Sathyanarayana, Joty, Fernandez-Luque, Ofli, Srivastava, Elmagarmid, Taheri, et al., 2016)	✓	✓	✓	-	✓	✓
(X. Li, et al., 2017)	✓	✓	-	-	✓	-
(Ordóñez & Roggen, 2016)	-	-	-	✓	✓	-
(Song, et al., 2016)	✓	-	-	✓	-	-
(Yang, et al., 2015)	✓	-	-	✓	✓	-

6. Common Datasets for Deep Learning Based Human Activity Recognition

Benchmark datasets are important for human activity recognition with deep learning methods. With benchmark datasets, researchers can test the performance of their proposed methods and how the results compare with previous studies. Some studies used datasets collected purposely for their research while others rely on public datasets to evaluate and validate their methods which are the most popular procedure among researchers in human activity recognition.

The main advantages of benchmark dataset are the ability to provide varieties of activity details both ambulatory, ambient living, daily, gesture and skill assessment activities (Nils Yannick Hammerla, et al., 2015). The most widely used benchmark datasets and the number of sensors, activities and subjects are shown in **Table 8**.

OPPORTUNITY Dataset (Roggen, et al., 2010) is a set of complex, hierarchical and interleaved dataset for activity of daily living (ADL) collected with multiple sensors of different modalities in naturalistic environments. During the data collection, the sensors were integrated into objects, environments and on-body that ensure multimodal data fusion and activity modelling. The OPPORTUNITY dataset is composed of sessions, daily living activities and drills. In the daily living activity section, the subjects were asked to perform different kitchen-related activities such as preparing and drinking coffee, eating sandwich, cleaning up, etc. while in the drill session, the subjects were asked to perform 20 set of repeated activities like “Opening and close the fridge”, “Open and close the dishwasher”, “Open and close the door”, “Clean the table” etc. for a period of 6 hours. All the datasets were gathered with Inertia Measurement Unit (IMU) sensors with different modalities inform of accelerometers, gyroscope and magnetometer. In a total of seventeen (17) activities were performed with twelve (12) subjects.

The Skoda Mini Checkpoint Dataset (Zappi, et al., 2008) was collected to check quality assurance checkpoint among assembly lines workers in car production environment. In the study, one subject wore twenty (20) 3D sensors on both arms and performed different manipulative gestures recorded for 3hours for seventy (70) repetitions in each gesture. The activities considered are “Write on notepad”, “Open hood”, “Close hood”, Check steering wheel” etc. using on-body sensors placed on the right and left arms.

Daily and Sports Activity (Barshan & Yüksek, 2014) was collected at Bilkent University in Turkey for human activity classification using on-body sensors placed on different parts of the body. The dataset involved five inertial measurement unit sensors by eight (8) subjects and performed nineteen (19) different ambulatory activities. The IMU collected multimodal data: accelerometers, gyroscope and magnetometer for activities involving walking, climbing stairs, standing, walking on the treadmill etc. It was made public after their research with intra-subject variability. It is a challenging dataset for human activity recognition.

WISDM dataset (Kwapisz, et al., 2011) by Wireless Sensor Data Mining Lab Fordham University describes a dataset collected for human activity recognition using Android based mobile phone accelerometer sensors. The data was collected from twenty-nine (29) users with single mobile phones doing simple ambulatory activities such as working, jogging, sitting, standing, etc.

PAMAP2 (Reiss & Stricker, 2012), *Physical Activity monitoring for Aging People* comprises daily activity dataset collected with three inertial measurement (IMU) and heart rate monitor sensors for a 10 hour period using nine (9) subjects. The sensors were placed at different body positions (dominant arm, ankle and chest region) and measured activities ranging sitting, jogging, watching TV to using the computers.

mHealth (Oresti Banos, et al., 2014) comprises 12 daily activity dataset collected using accelerometer, gyroscope, magnetometer and electrocardiogram sensor for health monitoring applications. It uses diverse mobile and wearable biomedical devices to collect sensor data. The architecture of the mobile app includes components such as data collection, storage, data processing and classification, data visualisation and service enablers that provide complete health monitoring systems.

Table 8: Benchmark Dataset for Human Activity Recognition Methods Evaluation

Authors	**Dataset**	**Sensor Modalities**	**Number of Sensors**	**#Participant**	**Activities**
(Roggen, et al., 2010)	OPPORTUNITY	Accelerometer, gyroscope, magnetometer	19	4	Open and close door, open and close fridge, open and close dishwasher, open and close drawer, clean table, drink from cup, Toggle switch, Groom, prepare coffee, Drink coffee, prepare Sandwich, eat sandwich, Clean up
(Zappi, et al., 2008)	Skoda	Accelerometer, gyroscope, magnetometer	20	1	Write on Notepad, open hood, close hood, check Gap door, open door, check steering wheel, open and close trunk, close both doors, close doors, check trunks
(Barshan & Yüksek, 2014)	Daily and Sports Activities	Accelerometer, gyroscope, magnetometer	5	8	Sitting, standing, lying on back, lying on right side, ascending stair descending stairs, standing in an elevator still, moving around in an elevator, walking in a parking lot, walking on a treadmill with a speed of 4 km/h in flat, walking on a treadmill with a speed of 4 km/h and 15 degree inclined positions, running on a treadmill with a speed of 8 km/h, exercising on a stepper, exercising on a cross trainer, cycling on an exercise bike in horizontal positions, cycling

					on an exercise bike in vertical position, rowing, jumping and playing basketball
(Kwapisz, et al., 2011)	WISDM v2	Accelerometer	1	29	Walking, Jogging, Upstairs, Downstairs, Sitting, Standing
	PAMAP2	Accelerometer, gyroscope and magnetometer	4	18	Lying, sitting, standing, walking, running, cycling, Nordic walking, Watching TV, Computer work, Car driving, Ascending stairs, Vacuum cleaning, descending stairs, ironing, folding laundry, house cleaning, playing soccer, rope jumping
(Oresti Banos, et al., 2014)	mHealth	Accelerometer, gyroscope, magnetometer, electrocardiogram	4	10	Standing still, sitting and relaxing, lying down, walking, climbing stairs, waist bends forward, frontal elevation of arms, knees bending, cycling, jogging, running, jumping front and back

7. Deep Learning Implementation Frameworks

Deep learning has come a long way and has become an important area of research. A number of software and hardware implementation platforms have been developed that exploit high-performance computing platforms to extract discriminative features for activity recognitions and other application areas. Some of these deep learning frameworks are open source, and others are proprietary developed by different organisations for use in cutting-edge technological development. NVidia[1] has become a driving force in the development of hardware technologies such as Graphical Processing Unit (GPU) and other processors that accelerate learning and improve the performance of deep learning methods. Recently, the organisation developed deep learning purpose-built microprocessors such as

[1] www.nvidia.co.uk

NVidia Tesla 40 GPU acceleration, Tesla M4 Hyperscale Accelerator and DGX-1 deep learning system (NVidia-Corps, 2017). Other companies like Mathematica, Wolfram, Nervana Systems, IBM and Intel Curie have followed suit in the development of deep learning implementation hardware (Ravì, Wong, Deligianni, et al., 2017).

One important aspect of the NVidia GPU is their support for the majority of the Machine learning and deep learning implementation tools and packages. Below, we discussed some of these tools and frameworks for implementation of deep learning and their various characteristics as shown in **Table 9**. Although the parameters used in the discussion were presented in (Ravì, Wong, Deligianni, et al., 2017), the frameworks were updated to reflect the current development in the area.

- *TensorFlow* (Abadi, et al., 2016) is an open source framework developed by Google Research Team for Numerical computation using data flow graph. TensorFlow has the highest number of community support for implementation of deep learning models. TensorFlow is very popular in deep learning research due to its flexibility for a variety of algorithms, portability and can run inference on mobile phones devices. Furthermore, it provides support for low level and high-level network training with multiple GPU, robust and provides consistency of parameter updates.

- *Theano* (Bergstra, et al., 2010) is a Python library used to define, optimise and evaluate the mathematical expression for multi-dimensional array. Theano provides high network modelling capability, dynamic code generation and speed with multiple GPU support. However, Theano provides low-level API and involves a lot of complex compilations that are often slow. Meanwhile, Theano has a wide range of learning resources and is still used by many researchers and developers.

- *Caffe* (Y. Jia, et al., 2014) is a framework for expressing algorithms in modular form. It provides C++ core language and binding support in Python and MATLAB. Caffe provides a complete architecture for training, testing and deployment of the deep learning model. Moreover, NVidia GPU provides Caffe support for accelerated learning of deep learning.

- *Pylearn2* (Goodfellow, et al., 2013) Pylearn2 was proposed in 2013 as machine learning library composed of several components that can be combined to form complete machine learning algorithms with deep learning models such as Autoencoder, Deep Belief Network, Deep Boltzmann machine implementation module. It is built on top of Theano and provides CPU and GPU support for intensive machine learning implementation. The major drawback of Pylearn is its low-level API that requires expert knowledge to implement any deep learning method.

- *Torch* (Collobert, Kavukcuoglu, & Farabet, 2011), scientific computing framework that provides model for machine learning implementation. The framework was developed to extend Lua programming Language and provide the flexibility needed to design and train machine learning algorihtms. It is equipped with tensor; standard MATLAB and Neural Network model functionalities that describe neural network architectures.
- *Cognitive Network Toolkit* (Microsoft, 2017) was developed by Microsoft Research to provide a unified framework for well-known deep learning algorithms. It provides multi-GPU parallelisation of learning techniques and implements stochastic gradient descent and automatic differentiation. The toolkit was released in 2015 and still has high community contribution in GitHub.
- *Lasagne* (Lasagne, 2015) provides a light library for implementation of deep learning algorithms such as convolutional neural network and recurrent neural network in Theano. It allows multiple input architectures with many popular optimisation techniques such as RMSprop and ADAM. The algorithm also provides CPU and Multiple GPU support for the implementation of deep learning methods.
- *Keras* (Chollet, 2015) was developed for deep learning implementation in Theano and TensorFlow written in Python programming language. It enables high-level neural network API for speedy implementation of deep learning algorithms. The main key point of Keras is its support for Theano and TensorFlow, popular deep learning implementation framework and allows modular, extensible and user platform using Python.
- *MXNet* (T. Chen, et al., 2015) combines symbolic and imperative programming to enable deep neural network implementation on heterogeneous devices (Mobile or GPU clusters). It automatically derives neural network gradients and graph optimisation layer to provide fast and memory efficient execution.

- *Deeplearning4j* (Skymind, 2017) developed by Skymind is an open source, distributed and commercial machine learning toolkits for deep learning implementation. The framework integrates Hadoop and Spark, with CPU and GPU-enabled for easy and quick prototyping of deep neural network implementation.

- *Neon* (Nervana-Systems, 2017) is developed for cross-platform implementation in all hardware with support for popular deep learning methods, convolutional neural network and recurrent neural network. Once codes are written in Neon, it can be deployed on different hardware platforms, and it provides the best performance among deep learning libraries.

- *Pytorch* (Erickson, Korfiatis, Akkus, Kline, & Philbrick, 2017) was recently developed at Facebook and is a front-end integration of Torch for high performance deep learning development with excellent GPU support. It provides Python front-end that enables dynamic neural network construction. However, the toolkit was recently released and does not have a lot of community support, learning resources and evaluation for its performance.

- *CuDNN* (Chetlur, et al., 2014) was developed as GPU-accelerated library for implementation of common deep learning methods. The framework with developing with the same intent as BLAS for optimised high-performance computing, to ease development, training and implementation of deep learning such as convolutional layer, recurrent neural network and backpropagation techniques. CuDNN supports both GPU and other platforms and provides straightforward integration with other frameworks such as TensorFlow, Caffe, Theano and Keras. Also, the context based API of CuDNN allows for multithreading and evaluation of complete deep learning algorithms.

Various other frameworks are still being developed that will simplify deep learning implementation across platforms and heterogeneous devices. For instance, frameworks such as DIGIT, Convnet and MATLAB based CNN toolbox for feature extraction, Cudanet, CUDA and C++ implementation of CNN and others are being fine-tuned to enable deep learning development. There are a number of evaluations of these frameworks that were reported recently (Bahrampour, Ramakrishnan, Schott, & Shah, 2015a, 2015b; Erickson, et al., 2017) using parameters such as language support, documentation, development environment, extension speed, training speed, GPU support, maturity level, model library, etc. From these, TensorFlow has the highest GitHub interest and contribution, surpassing Caffe and CNTK. Also, some of the frameworks support GPU or have limited support in which the GPU has to be resident on the workstation (e.g., MXNet).

With the development of deep learning based human activity recognition, these frameworks have become dominant choices for developers and researcher for mobile and wearable sensor based applications. With different implementation frameworks and varying programming support, the choice of the framework depends on the programming and technical ability of the users. The software frameworks recently used for mobile-based human activity recognition are TensorFlow (Eskofier, et al., 2016; Kautz, et al., 2017), Theano (Ordóñez & Roggen, 2016; C. A. Ronao & S.-B. Cho, 2016), Caffe (Yin, et al., 2016), Keras(X. Li, et al., 2017), Torch (Daniele Ravi, et al., 2016) and Lasagne (Guan & Ploetz, 2017). Other studies develop the algorithm using programming platforms such as MATLAB (S. Bhattacharya & Lane, 2016; Erfani, et al., 2016; Sheng, et al., 2016; Zebin, et al., 2016) and C++ (Ding, et al., 2016).

Table 9: Software Frameworks for Deep Learning Implementation

Name	Organisation	Licence	Platform	Language Support	OpenMP Support	Support Techniques			Cloud Computing Support
						RNN	CNN	DBN	
Theano	Universite de Montreal	BSD	Cross Platform	Python	-	✓	✓	✓	-
TensorFlow	Google Research	Apache 2.0	Linux, OSX	Python	✓	-	✓	✓	-
Caffe	Berkeley Vision and Learning Centre	FreeBSD	Linux, Win, OSX, Android,	C++,Python, MATLAB	-	-	✓	-	-
Torch	Ronan Collobert et al.	BSD	Linux, Win, OSX, Android, iOS	Lua, LuaJIT, C	✓	-	✓	✓	-
CNTK	Microsoft	MIT	Linux, Window	C++, Python, C#, Command Line	✓	✓	✓	-	-
Deeplearning4jK	Skymind	Apache 2.0	Linux, Win, OSX, Android	Java, Scala, Clojure, Spark	✓	✓	✓	✓	-
Keras	Francois Chollet	MIT Licence	Linux, Win, OSX	Python	-	✓	✓		
Neon	Nervana Systems	Apache 2.0	OSX, Linux	Python	✓	✓	✓	✓	✓
Lasagne	Universite de Montreal	BSD	Linux, Win, OSX, Android	Python	✓	✓	✓	✓	-
MXNet	Chen et al	Apache 2.0	Linux, Win, Andriod	Python, R, C++, Julia	-	✓	✓	-	-
Pylearn	LISA Lab Universite de Montreal	BSD	Cross Platform	Python	✓	✓	✓	✓	-
PyTorch	Facebook	BSD	Linux	Python	✓	✓	✓	✓	
CuDNN	NVIDIA	Free BSD	Linux, Win, Android, OSX	C	✓	✓	✓	-	✓

8. Open Research Challenges

In this section, we present some research challenges that require further discussion. Many open research issues in the area of sensor fusion, real-time and on-board implementation on mobile and wearable devices, data pre-processing and evaluation, collection of large dataset and class imbalance problems are some of the areas that required further research. Here, we discuss these research directions in seven important themes:

- *Real-time and on-board implementation of deep learning algorithm on mobile and wearable devices:* On-board implementation of deep learning algorithms on mobile and wearable devices will help to reduce computation complexity on data storage and transfer. However, this technique is hampered by data acquisition and memory constrained in the current mobile and wearable devices. Furthermore, a high number of parameters tuning and initialisation in deep learning increases computational time and is not suitable for low energy mobile devices. Therefore, utilising methods such as optimal compression and use of mobile phone enabled GPU to minimise computation time and resources consumptions is highly needed. Other methods that may provide enabling techniques for real-time implementation is leveraging mobile cloud computing platforms for training to reduce training time and memory usage. With this type of implementation, the system can become self-adaptive and require minimal user inputs for a new source of information.

- *Comprehensive evaluation of pre-processing and hyper-parameter settings on learning algorithms:* Pre-processing and dimensionality reduction is an important aspect of the human activity recognition process. Dimensionality reduction provide mechanism to minimize computational complexity especially in mobile and wearable devices with limited computation powers and memory by projecting high dimensional sensor data into lower dimensional vectors. However, the method and extent of pre-processing on the performance of deep learning is an open research challenge. A number of pre-processing techniques such as normalisation, standardisation and different dimensionality reduction methods need to be experimented with, to know the effects on performances, computational time and accuracy of deep-learning methods. Issues such as learning rate optimisation to accelerate computation and reduce model and data size, kernel reuse, filter size, computation time, memory analysis and learning process still require further research as current studies depend on heuristics method to apply these hyper-parameters. Moreover, the use of grid search and evolutionary optimisation methods on mobile based deep learning methods that support lower energy consumption, dynamic and adaptive applications, and new techniques that enable mobile GPUs to reduce computational time are very significant research directions(Ordonez & Roggen, 2016).

- *A collection of large sensor datasets for evaluation of deep learning methods:* Training and evaluation of deep learning techniques require large datasets that abound through different sensor based Internet of Thing (IoT) devices and technologies. The current review indicates that most studies on deep learning implementation of mobile and wearable based human activity recognition depend on benchmark dataset from conventional machine learning algorithms such as OPPORTUNITY, Skoda and WSDM for evaluation. Data collection methods through cyber-physical systems and mobile crowdsourcing to leverage data collected through the smart home and mobile location data for transportation mode, smart home environment for elderly care and monitoring, GPS data for context aware location recognition and other important applications. Therefore, collection of large dataset through the synergy of these technologies are important for performance improvements.

- *Transfer learning for mobile and wearable devices implementation of deep learning algorithms:* Transfer learning based activity recognition is a challenging task to accomplish. Transfer learning leverage experience acquired in different domains to improve the performance of new areas yet to be experienced by the system. The main vital reasons for application of transfer learning are to reduce training time, provide robust and versatile activity details and reuse of existing knowledge into new domains and a critical issue in activity recognition. Further research in area related to kernel, convolutional layer, inter-location and inter-modalities transferability will improve implementation of deep learning based human activity recognition(Ordonez & Roggen, 2016). Moreover, transfer learning in mobile wearable sensor based human activity recognition will minimize source, target and environment specific applications implementation which have not received the needed attention.

- *Implementation of deep learning based decision fusion for human activity recognition in mobile and wearable devices:* Decision fusion is an essential step to improve the performance and diversity of human activity recognition systems by combining several architectures, sensors and classifiers into a single decision. Typical areas that require further researches are heterogeneous sensor fusion, combining expert knowledge with deep learning algorithm and combination of different unsupervised feature learning methods to improve performance of activity recognition systems.

- *Solving the class imbalance problem for deep learning in mobile and wearable based human activity recognition:* Class imbalance issues can be found in datasets for human activity recognition and detection of abnormal activities. Class imbalance problem is vital in healthcare monitoring especially fall detection in which what constitute actual fall is difficult. For mobile and wearable sensor based human activity recognition, class imbalance maybe as a result of a distortion in the dataset and sensor data calibration which reduce performance generalisation (Edel & Köppe, 2016). Existing studies have proposed a range of solutions such as mixed kernel based weighted extreme learning machine and cost sensitive learning strategies (D. Wu, Wang, Chen, & Zhao, 2016). However, there are no studies on how class imbalance affect deep learning implementation especially for mobile wearable sensors. Therefore, strategies to reduce class imbalance will significantly improve human activity recognition using deep learning methods.

- *Augmentation of mobile and wearable sensor data to enhance deep learning performance:* Another aspect of open research challenge is the use of data augmentation techniques to improve the performance of deep learning methods for motion sensors (accelerometer, gyroscopes, etc.) based human activity recognition with the convolutional neural network. Data augmentation methods exploit limited amount of mobile and wearable sensor data by transforming the existing training sensor data to generate new data. These processes are important as it help to generate enough training data to avoid overfitting, improve translation invariance to sensor orientation, distortion and changes especially in convolutional neural network (CNN) model. In image classification, data augmentation is a common training strategy (Y. Guo, et al., 2016). However, there is need to evaluate the impacts and performances of data augmentation in mobile and wearable sensor-based human activity recognition to generate more training examples and prevent overfitting resulting from small datasets. Different data augmentation approaches such as change of sensor placements, arbitrary rotations, permutation of locations with sensor events, time warping and scaling will provide effective means to enhance performance of deep learning based human activity recognition(Um, et al., 2017).

9. Conclusion

Automatic feature learning in human activity recognition is increasing in momentum. This is as results of the steady rise in computation facilities and large datasets available through mobile and wearable sensing, Internet of Things (IoT) and crowd sourcing. In this paper, we reviewed various deep learning methods that enable automatic feature extraction in human activity recognition. Deep learning methods such as Restricted Boltzmann Machine, Autoencoder, and Convolutional Neural Networks and Recurrent neural network were presented and their characteristics, advantages and drawback were equally exposed. Deep learning methods can be classified as generative, discriminative and hybrid methods. We utilise the categorisations to review and outline deep learning implementation of human activity recognition. Those in the generative categories are the Restricted Boltzmann Machine, autoencoder, sparse coding and deep mixture model while the discriminative approaches include the convolutional neural network, recurrent neural network, deep neural model and hydrocarbon. Similarly, hybrid methods combine generative and discriminative model to enhance feature learning and such combination dominant research landscape of deep learning for human activity recognition lately. Hybrid methods incorporate diverse generative model such as autoencoder, Restricted Boltzmann Machine with the convolutional neural network or combine discriminative models such as convolutional neural network and long short term memory. These approaches are an important step to achieving automatic feature learning and enhancing performance generalisation across datasets and activities.

On the other hand, the implementation of deep learning methods is driven by the availability of high-performance computing GPU and software frameworks. A number of these software frameworks were recently released to the research community as open sources projects. These software frameworks were discussed, taking into cognizance their characteristics and what inform developers' choice in using particular frameworks. Also, training, classification and evaluation of deep learning algorithm for human activity recognition is not always a trivial case. To provide the best comparison and categorisations of recent events in the research community, we reviewed the training and optimisation strategies adopted by different studies recently proposed for mobile and wearable based human activity recognition. Furthermore, classification and performance metrics with different validation techniques are important to ensure generalisation across datasets. These approaches are adopted to avoid overfitting the model on the training set. Also, we provide some of the publicly available benchmark datasets for modelling and testing deep learning algorithms for human activity recognition. Some of these datasets that are widely used for evaluation are OPPORTUNITY, Skoda, and PAMAP2 which are also popular with classical machine learning algorithms.

To provide further insight on the directions of the research progress, we presented the open research challenges that require the attention of researchers. For instance, areas such as deep learning based decision fusion, implementation of deep learning on-board mobile devices, transfer learning and class imbalance problems that enable implementation of human activity recognition for enhanced performance accuracy. With further development of high computational resources that increase the online and real-time deep learning implementation on mobile and wearable devices, such machine learning techniques are projected to improve human activity recognition researches.

REFERENCES

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., & Devin, M. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*.

Abdel-Hamid, O., Deng, L., & Yu, D. (2013). Exploring convolutional neural network structures and optimization techniques for speech recognition. In *Interspeech* (pp. 3366-3370).

Abidine, B. M. h., Fergani, L., Fergani, B., & Oussalah, M. (2016). The joint use of sequence features combination and modified weighted SVM for improving daily activity recognition. *Pattern Analysis and Applications*, 1-20.

Aggarwal, J. K., & Xia, L. (2014). Human activity recognition from 3D data: A review. *Pattern Recognition Letters, 48*, 70-80.

Ahmad, M., Saeed, M., Saleem, S., & Kamboh, A. M. (2016). Seizure detection using EEG: A survey of different techniques. In *Emerging Technologies (ICET), 2016 International Conference on* (pp. 1-6): IEEE.

Al Rahhal, M., Bazi, Y., AlHichri, H., Alajlan, N., Melgani, F., & Yager, R. (2016). Deep learning approach for active classification of electrocardiogram signals. *Information Sciences, 345*, 340-354.

Alsheikh, M. A., Niyato, D., Lin, S., Tan, H.-P., & Han, Z. (2016). Mobile big data analytics using deep learning and apache spark. *IEEE Network, 30*, 22-29.

Alsheikh, M. A., Selim, A., Niyato, D., Doyle, L., Lin, S., & Tan, H.-P. (2015). Deep Activity Recognition Models with Triaxial Accelerometers. *arXiv preprint arXiv:1511.04664*.

Alvear-Sandoval, R. F., & Figueiras-Vidal, A. R. (2018). On building ensembles of stacked denoising auto-encoding classifiers and their further improvement. *Information Fusion, 39*, 41-52.

Alzantot, M., Chakraborty, S., & Srivastava, M. B. (2017). SenseGen: A Deep Learning Architecture for Synthetic Sensor Data Generation. *arXiv preprint arXiv:1701.08886*.

Angermueller, C., Parnamaa, T., Parts, L., & Stegle, O. (2016). Deep learning for computational biology. *Molecular Systems Biology, 12*.

Anguita, D., Ghio, A., Oneto, L., Parra, X., & Reyes-Ortiz, J. L. (2012). Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine. In *International Workshop on Ambient Assisted Living* (pp. 216-223): Springer.

Attal, F., Mohammed, S., Dedabrishvili, M., Chamroukhi, F., Oukhellou, L., & Amirat, Y. (2015). Physical Human Activity Recognition Using Wearable Sensors. *Sensors, 15*, 31314-31338.
Bader, G. D., & Hogue, C. W. (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC bioinformatics, 4*, 2.
Bahrampour, S., Ramakrishnan, N., Schott, L., & Shah, M. (2015a). Comparative study of caffe, neon, theano, and torch for deep learning. *arXiv preprint arXiv:1511.06435*.
Bahrampour, S., Ramakrishnan, N., Schott, L., & Shah, M. (2015b). Comparative study of deep learning software frameworks. *arXiv preprint arXiv:1511.06435*.
Banos, O., Galvez, J. M., Damas, M., Guillen, A., Herrera, L. J., Pomares, H., Rojas, I., Villalonga, C., Hong, C. S., & Lee, S. (2015). Multiwindow Fusion for Wearable Activity Recognition. In I. Rojas, G. Joya & A. Catala (Eds.), *Advances in Computational Intelligence, Pt Ii* (Vol. 9095, pp. 290-297).
Banos, O., Garcia, R., Holgado-Terriza, J. A., Damas, M., Pomares, H., Rojas, I., Saez, A., & Villalonga, C. (2014). mHealthDroid: a novel framework for agile development of mobile health applications. In *International Workshop on Ambient Assisted Living* (pp. 91-98): Springer.
Barshan, B., & Yüksek, M. C. (2014). Recognizing daily and sports activities in two open source machine learning environments using body-worn sensor units. *The Computer Journal, 57*, 1649-1667.
Bengio, Y. (2009). Learning deep architectures for AI. *Foundations and trends® in Machine Learning, 2*, 1-127.
Benuwa, B., Zhan, Y. Z., Ghansah, B., Wornyo, D. K., & Kataka, F. B. (2016). A Review of Deep Machine Learning. *International Journal of Engineering Research in Africa, 24*, 124-136.
Bergstra, J., Breuleux, O., Bastien, F., Lamblin, P., Pascanu, R., Desjardins, G., Turian, J., Warde-Farley, D., & Bengio, Y. (2010). Theano: A CPU and GPU math compiler in Python. In *Proc. 9th Python in Science Conf* (pp. 1-7).
Bhattacharya, S., & Lane, N. D. (2016). From smart to deep: Robust activity recognition on smartwatches using deep learning. In *2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)* (pp. 1-6).
Bhattacharya, S., Nurmi, P., Hammerla, N., & Plötz, T. (2014). Using unlabeled data in a sparse-coding framework for human activity recognition. *Pervasive and Mobile Computing, 15*, 242-262.
Bordes, A., Chopra, S., & Weston, J. (2014). Question answering with subgraph embeddings. *arXiv preprint arXiv:1406.3676*.
Boureau, Y.-L., Ponce, J., & LeCun, Y. (2010). A theoretical analysis of feature pooling in visual recognition. In *Proceedings of the 27th international conference on machine learning (ICML-10)* (pp. 111-118).
Bulling, A., Blanke, U., & Schiele, B. (2014a). A tutorial on human activity recognition using body-worn inertial sensors. *Acm Computing Surveys, 46*, 1-33.
Bulling, A., Blanke, U., & Schiele, B. (2014b). A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys (CSUR), 46*, 33.
Cao, L., Wang, Y., Zhang, B., Jin, Q., & Vasilakos, A. V. (2017). GCHAR: An efficient Group-based Context–aware human activity recognition on smartphone. *Journal of Parallel and Distributed Computing*.
Capela, N. A., Lemaire, E. D., & Baddour, N. (2015). Feature selection for wearable smartphone-based human activity recognition with able bodied, elderly, and stroke patients. *Plos One, 10*, e0124414.
Castro, D., Hickson, S., Bettadapura, V., Thomaz, E., Abowd, G., Christensen, H., & Essa, I. (2015). Predicting daily activities from egocentric images using deep learning. In *proceedings of the 2015 ACM International symposium on Wearable Computers* (pp. 75-82): ACM.
Charalampous, K., & Gasteratos, A. (2016). On-line deep learning method for action recognition. *Pattern Analysis and Applications, 19*, 337-354.

Chen, M., Xu, Z., Weinberger, K., & Sha, F. (2012). Marginalized denoising autoencoders for domain adaptation. *arXiv preprint arXiv:1206.4683*.

Chen, T., Li, M., Li, Y., Lin, M., Wang, N., Wang, M., Xiao, T., Xu, B., Zhang, C., & Zhang, Z. (2015). Mxnet: A flexible and efficient machine learning library for heterogeneous distributed systems. *arXiv preprint arXiv:1512.01274*.

Chen, Y.-L., Wu, X., Li, T., Cheng, J., Ou, Y., & Xu, M. (2016). Dimensionality reduction of data sequences for human activity recognition. *Neurocomputing, 210*, 294-302.

Chen, Y., & Xue, Y. (2015). A deep learning approach to human activity recognition based on single accelerometer. In *Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on* (pp. 1488-1492): IEEE.

Chen, Y., Zhong, K., Zhang, J., Sun, Q., & Zhao, X. (2016). LSTM Networks for Mobile Human Activity Recognition.

Chen, Y. Q., Xue, Y., & Ieee. (2015). A Deep Learning Approach to Human Activity Recognition Based on Single Accelerometer. In *2015 Ieee International Conference on Systems, Man and Cybernetics* (pp. 1488-1492). Los Alamitos: Ieee Computer Soc.

Chetlur, S., Woolley, C., Vandermersch, P., Cohen, J., Tran, J., Catanzaro, B., & Shelhamer, E. (2014). cudnn: Efficient primitives for deep learning. *arXiv preprint arXiv:1410.0759*.

Cho, K., Raiko, T., & Ihler, A. T. (2011). Enhanced gradient and adaptive learning rate for training restricted Boltzmann machines. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)* (pp. 105-112).

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

Choi, E., Schuetz, A., Stewart, W. F., & Sun, J. (2016). Using recurrent neural network models for early detection of heart failure onset. *Journal of the American Medical Informatics Association*, ocw112.

Chollet, F. (2015). Keras: Deep learning library for theano and tensorflow. *URL: https://keras. io/k*.

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.

Chung, J., Gülçehre, C., Cho, K., & Bengio, Y. (2015). Gated Feedback Recurrent Neural Networks. In *ICML* (pp. 2067-2075).

Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A., & Oliva, A. (2016). Deep neural networks predict hierarchical spatio-temporal cortical dynamics of human visual object recognition. *arXiv preprint arXiv:1601.02970*.

Collobert, R., Kavukcuoglu, K., & Farabet, C. (2011). Torch7: A matlab-like environment for machine learning. In *BigLearn, NIPS Workshop*.

Cornacchia, M., Ozcan, K., Zheng, Y., & Velipasalar, S. (2017). A Survey on Activity Detection and Classification Using Wearable Sensors. *IEEE Sensors Journal, 17*, 386-403.

Dauphin, G. M. Y., Glorot, X., Rifai, S., Bengio, Y., Goodfellow, I., Lavoie, E., Muller, X., Desjardins, G., Warde-Farley, D., Vincent, P., Courville, A., & Bergstra, J. (2012). Unsupervised and Transfer Learning Challenge: a Deep Learning Approach. In G. Isabelle, D. Gideon, L. Vincent, T. Graham & S. Daniel (Eds.), *Proceedings of ICML Workshop on Unsupervised and Transfer Learning* (Vol. 27, pp. 97--110). Proceedings of Machine Learning Research: PMLR.

Deng, L. (2014). A tutorial survey of architectures, algorithms, and applications for deep learning. *APSIPA Transactions on Signal and Information Processing, 3*, e2.

Ding, X., Lei, H., & Rao, Y. (2016). Sparse codes fusion for context enhancement of night video surveillance. *Multimedia Tools and Applications, 75*, 11221-11239.

Dolmans, D., Loyens, S. M. M., Marcq, H., & Gijbels, D. (2016). Deep and surface learning in problem-based learning: a review of the literature. *Advances in Health Sciences Education, 21*, 1087-1112.

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM, 55*, 78-87.

Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research, 12*, 2121-2159.

Edel, M., & Köppe, E. (2016). Binarized-BLSTM-RNN based Human Activity Recognition. In *2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN)* (pp. 1-7).

Egede, J., Valstar, M., & Martinez, B. (2017). Fusing Deep Learned and Hand-Crafted Features of Appearance, Shape, and Dynamics for Automatic Pain Estimation. *arXiv preprint arXiv:1701.04540*.

Erfani, S. M., Rajasegarar, S., Karunasekera, S., & Leckie, C. (2016). High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning. *Pattern Recognition, 58*, 121-134.

Erickson, B. J., Korfiatis, P., Akkus, Z., Kline, T., & Philbrick, K. (2017). Toolkits and Libraries for Deep Learning. *Journal of Digital Imaging*, 1-6.

Eskofier, B. M., Lee, S. I., Daneault, J.-F., Golabchi, F. N., Ferreira-Carvalho, G., Vergara-Diaz, G., Sapienza, S., Costante, G., Klucken, J., & Kautz, T. (2016). Recent machine learning advancements in sensor-based mobility analysis: Deep learning for Parkinson's disease assessment. In *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the* (pp. 655-658): IEEE.

Fang, H., He, L., Si, H., Liu, P., & Xie, X. (2014). Human activity recognition based on feature selection in smart home using back-propagation algorithm. *ISA transactions, 53*, 1629-1638.

Fang, H., & Hu, C. (2014). Recognizing human activity in smart home using deep learning algorithm. In *Proceedings of the 33rd Chinese Control Conference* (pp. 4716-4720).

Feng, Y., Yuan, Y., & Lu, X. (2017). Learning deep event models for crowd anomaly detection. *Neurocomputing, 219*, 548-556.

Figo, D., Diniz, P. C., Ferreira, D. R., & Cardoso, J. M. (2010). Preprocessing techniques for context recognition from accelerometer data. *Personal and Ubiquitous Computing, 14*, 645-662.

Figo, D., Diniz, P. C., Ferreira, D. R., Jo, #227, & Cardoso, o. M. (2010). Preprocessing techniques for context recognition from accelerometer data. *Personal Ubiquitous Comput., 14*, 645-662.

Fischer, A., & Igel, C. (2014). Training restricted Boltzmann machines: An introduction. *Pattern Recognition, 47*, 25-39.

Gamboa, J. C. B. (2017). Deep Learning for Time-Series Analysis. *arXiv preprint arXiv:1701.01887*.

Gao, J., Yang, J., Wang, G., & Li, M. (2016). A novel feature extraction method for scene recognition based on Centered Convolutional Restricted Boltzmann Machines. *Neurocomputing, 214*, 708-717.

Gao, Y., & Glowacka, D. (2016). Deep Gate Recurrent Neural Network. *arXiv preprint arXiv:1604.02910*.

Gawehn, E., Hiss, J. A., & Schneider, G. (2016). 6 Deep Learning in Drug Discovery. *Molecular Informatics, 35*, 3-14.

Gjoreski, H., Bizjak, J., Gjoreski, M., & Gams, M. (2015). Comparing Deep and Classical Machine Learning Methods for Human Activity Recognition using Wrist Accelerometer.

Gjoreski, M., Gjoreski, H., Luštrek, M., & Gams, M. (2016). How accurately can your wrist device recognize daily activities and detect falls? *Sensors, 16*, 800.

Gong, J., Cui, L., Xiao, K., & Wang, R. (2012). MPD-Model: A distributed multipreference-driven data fusion model and its application in a WSNs-based healthcare monitoring system. *International Journal of Distributed Sensor Networks, 8*, 602358.

Gong, J. Q., Goldman, M. D., & Lach, J. (2016). DeepMotion: A Deep Convolutional Neural Network on Inertial Body Sensors for Gait Assessment in Multiple Sclerosis. *2016 Ieee Wireless Health (Wh)*, 164-171.

Goodfellow, I. J., Warde-Farley, D., Lamblin, P., Dumoulin, V., Mirza, M., Pascanu, R., Bergstra, J., Bastien, F., & Bengio, Y. (2013). Pylearn2: a machine learning research library. *arXiv preprint arXiv:1308.4214*.

Graves, A. (2013). Generating sequences with recurrent neural networks. *arXiv preprint arXiv:1308.0850*.

Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks, 18*, 602-610.

Gravina, R., Alinia, P., Ghasemzadeh, H., & Fortino, G. (2017). Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges. *Information Fusion, 35*, 68-80.

Gravina, R., Ma, C., Pace, P., Aloi, G., Russo, W., Li, W., & Fortino, G. (2017). Cloud-based Activity-aaService cyber–physical framework for human activity monitoring in mobility. *Future Generation Computer Systems, 75*, 158-171.

Guan, Y., & Ploetz, T. (2017). Ensembles of Deep LSTM Learners for Activity Recognition using Wearables. *arXiv preprint arXiv:1703.09370*.

Guo, J., Xie, X., Bie, R., & Sun, L. (2014). Structural health monitoring by using a sparse coding-based deep learning algorithm with wireless sensor networks. *Personal and Ubiquitous Computing, 18*, 1977-1987.

Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. *Neurocomputing, 187*, 27-48.

Ha, S., Yun, J. M., & Choi, S. (2015). Multi-modal Convolutional Neural Networks for Activity Recognition. In *2015 IEEE International Conference on Systems, Man, and Cybernetics* (pp. 3017-3022).

Habib, C., Makhoul, A., Darazi, R., & Couturier, R. (2016). Multisensor Data Fusion and Decision Support in Wireless Body Sensor Networks. In S. O. Badonnel, M. Ulema, C. Cavdar, L. Z. Granville & C. R. P. DosSantos (Eds.), *Noms 2016 - 2016 Ieee/Ifip Network Operations and Management Symposium* (pp. 708-712).

Hammerla, N. Y. (2015). Activity recognition in naturalistic environments using body-worn sensors.

Hammerla, N. Y., Fisher, J., Andras, P., Rochester, L., Walker, R., & Plötz, T. (2015). PD Disease State Assessment in Naturalistic Environments Using Deep Learning. In *AAAI* (pp. 1742-1748).

Hammerla, N. Y., Halloran, S., & Ploetz, T. (2016). Deep, Convolutional, and Recurrent Models for Human Activity Recognition using Wearables. *arXiv preprint arXiv:1604.08880*.

Harandi, M. T., Sanderson, C., Hartley, R., & Lovell, B. C. (2012). Sparse coding and dictionary learning for symmetric positive definite matrices: A kernel approach. In *Computer Vision–ECCV 2012* (pp. 216-229): Springer.

Harasimowicz, A. (2014). Comparison of Data Preprocessing Methods and the Impact on Auto-encoder's Performance in Activity Recognition Domain.

Hasan, M., & Roy-Chowdhury, A. K. (2014). Continuous learning of human activity models using deep nets. In *European Conference on Computer Vision* (pp. 705-720): Springer.

Hasan, M., & Roy-Chowdhury, A. K. (2015). A Continuous Learning Framework for Activity Recognition Using Deep Hybrid Feature Models. *IEEE Transactions on Multimedia, 17*, 1909-1922.

He, Y., Kavukcuoglu, K., Wang, Y., Szlam, A., & Qi, Y. (2014). Unsupervised feature learning by deep sparse coding. In *Proceedings of the 2014 SIAM International Conference on Data Mining* (pp. 902-910): SIAM.

Hinton, G. (2010). A practical guide to training restricted Boltzmann machines. *Momentum, 9*, 926.

Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A.-r., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., & Sainath, T. N. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *Ieee Signal Processing Magazine, 29*, 82-97.

Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural computation, 18*, 1527-1554.

Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation, 18*, 1527-1554.

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science, 313*, 504-507.

Hinton, G. E., & Sejnowski, T. J. (1986). Learning and releaming in Boltzmann machines. *Parallel distributed processing: Explorations in the microstructure of cognition, 1*, 282-317.

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation, 9*, 1735-1780.

Hodo, E., Bellekens, X., Hamilton, A., Tachtatzis, C., & Atkinson, R. (2017). Shallow and Deep Networks Intrusion Detection System: A Taxonomy and Survey. *arXiv preprint arXiv:1701.02145*.

Hollensen, P., & Trappenberg, T. P. (2015). An Introduction to Deep Learning. In D. Barbosa & E. Milios (Eds.), *Advances in Artificial Intelligence* (Vol. 9091). Berlin: Springer-Verlag Berlin.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology, 160*, 106-154.

Ijjina, E. P., & Mohan, C. K. (2016). Hybrid deep neural network model for human action recognition. *Applied Soft Computing, 46*, 936-952.

Incel, O. (2015). Analysis of Movement, Orientation and Rotation-Based Sensing for Phone Placement Recognition. *Sensors, 15*, 25474.

Incel, O. D., Kose, M., & Ersoy, C. (2013). A review and taxonomy of activity recognition on mobile phones. *BioNanoScience, 3*, 145-171.

Inoue, M., Inoue, S., & Nishida, T. (2016). Deep Recurrent Neural Network for Mobile Human Activity Recognition with High Throughput. *arXiv preprint arXiv:1611.03607*.

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*.

Jänicke, M., Tomforde, S., & Sick, B. (2016). Towards Self-Improving Activity Recognition Systems Based on Probabilistic, Generative Models. In *Autonomic Computing (ICAC), 2016 IEEE International Conference on* (pp. 285-291): IEEE.

Jia, X., Li, K., Li, X., & Zhang, A. (2014). A novel semi-supervised deep learning framework for affective state recognition on eeg signals. In *Bioinformatics and Bioengineering (BIBE), 2014 IEEE International Conference on* (pp. 30-37): IEEE.

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., & Darrell, T. (2014). Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia* (pp. 675-678): ACM.

Jia, Y., Song, X., Zhou, J., Liu, L., Nie, L., & Rosenblum, D. S. (2016). Fusing Social Networks with Deep Learning for Volunteerism Tendency Prediction. In *Thirtieth AAAI Conference on Artificial Intelligence*.

Jiang, W., & Yin, Z. (2015). Human activity recognition using wearable sensors by deep convolutional neural networks. In *Proceedings of the 23rd ACM international conference on Multimedia* (pp. 1307-1310): ACM.

Jing, L., Wang, T., Zhao, M., & Wang, P. (2017). An Adaptive Multi-Sensor Data Fusion Method Based on Deep Convolutional Neural Networks for Fault Diagnosis of Planetary Gearbox. *Sensors, 17*, 414.

Jokanovic, B., Amin, M., & Ahmad, F. (2016). Radar fall motion detection using deep learning. In *Radar Conference (RadarConf), 2016 IEEE* (pp. 1-6): IEEE.

Kanaris, L., Kokkinis, A., Liotta, A., & Stavrou, S. (2017). Fusing Bluetooth Beacon Data with Wi-Fi Radiomaps for Improved Indoor Localization. *Sensors, 17*, 812.

Karpathy, A., Johnson, J., & Fei-Fei, L. (2015). Visualizing and understanding recurrent networks. *arXiv preprint arXiv:1506.02078*.

Kautz, T., Groh, B. H., Hannink, J., Jensen, U., Strubberg, H., & Eskofier, B. M. (2017). Activity recognition in beach volleyball using a Deep Convolutional Neural Network. *Data Mining and Knowledge Discovery*, 1-28.

Khan, S. S., & Taati, B. (2017). Detecting unseen falls from wearable devices using channel-wise ensemble of autoencoders. *Expert Systems with Applications, 87*, 280-290.

Kim, Y., & Ling, H. (2009). Human activity classification based on micro-Doppler signatures using a support vector machine. *IEEE Transactions on Geoscience and Remote Sensing, 47*, 1328-1337.

Kingma, D., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).

Kumari, P., Mathew, L., & Syal, P. (2017). Increasing trend of wearables and multimodal interface for human activity monitoring: A review. *Biosensors and Bioelectronics, 90*, 298-307.

Kwapisz, J. R., Weiss, G. M., & Moore, S. A. (2011). Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter, 12*, 74-82.

Langkvist, M., Karlsson, L., & Loutfi, A. (2014). A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognition Letters, 42*, 11-24.

Längkvist, M., Karlsson, L., & Loutfi, A. (2012). Sleep stage classification using unsupervised feature learning. *Advances in Artificial Neural Systems, 2012*, 5.

Lasagne. (2015) (Vol. 2017).

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*, 436-444.

LeCun, Y., Huang, F. J., & Bottou, L. (2004). Learning methods for generic object recognition with invariance to pose and lighting. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on* (Vol. 2, pp. II-104): IEEE.

Lee, H., Grosse, R., Ranganath, R., & Ng, A. Y. (2009). Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In *Proceedings of the 26th annual international conference on machine learning* (pp. 609-616): ACM.

Li, G., Deng, L., Xu, Y., Wen, C., Wang, W., Pei, J., & Shi, L. (2016). Temperature based Restricted Boltzmann Machines. *Scientific reports, 6*.

Li, X., Zhang, Y., Zhang, J., Chen, S., Marsic, I., Farneth, R. A., & Burd, R. S. (2017). Concurrent Activity Recognition with Multimodal CNN-LSTM Structure. *arXiv preprint arXiv:1702.01638*.

Li, Y., Shi, D., Ding, B., & Liu, D. (2014). Unsupervised feature learning for human activity recognition using smartphone sensors. In *Mining Intelligence and Knowledge Exploration* (pp. 99-107): Springer.

Lin, L., Wang, K., Zuo, W., Wang, M., Luo, J., & Zhang, L. (2015). A deep structured model with radius–margin bound for 3D human activity recognition. *International Journal of Computer Vision*, 1-18.

Lin, L., Wang, K. Z., Zuo, W. M., Wang, M., Luo, J. B., & Zhang, L. (2016). A Deep Structured Model with Radius-Margin Bound for 3D Human Activity Recognition. *International Journal of Computer Vision, 118*, 256-273.

Liou, C.-Y., Cheng, W.-C., Liou, J.-W., & Liou, D.-R. (2014). Autoencoder for words. *Neurocomputing, 139*, 84-96.

Liu, G., Liang, J., Lan, G., Hao, Q., & Chen, M. (2016). Convolution neutral network enhanced binary sensor network for human activity recognition. In *SENSORS, 2016 IEEE* (pp. 1-3): IEEE.

Liu, H., & Taniguchi, T. (2014). Feature extraction and pattern recognition for human motion by a deep sparse autoencoder. In *Computer and Information Technology (CIT), 2014 IEEE International Conference on* (pp. 173-181): IEEE.

Liu, L., Cheng, L., Liu, Y., Jia, Y., & Rosenblum, D. S. (2016). Recognizing Complex Activities by a Probabilistic Interval-Based Model. In *AAAI* (pp. 1266-1272).

Liu, W., Ma, H. D., Qi, H., Zhao, D., & Chen, Z. N. (2017). Deep learning hashing for mobile visual search. *EURASIP Journal on Image and Video Processing*.

Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., & Alsaadi, F. E. (2016). A survey of deep neural network architectures and their applications. *Neurocomputing*.

Lu, Y., Wei, Y., Liu, L., Zhong, J., Sun, L., & Liu, Y. (2016). Towards unsupervised physical activity recognition using smartphone accelerometers. *Multimedia Tools and Applications*, 1-19.

Ma, G., Yang, X., Zhang, B., & Shi, Z. (2016). Multi-feature fusion deep networks. *Neurocomputing, 218*, 164-171.

Ma, J., Sheridan, R. P., Liaw, A., Dahl, G. E., & Svetnik, V. (2015). Deep neural nets as a method for quantitative structure–activity relationships. *Journal of Chemical Information and Modeling, 55*, 263-274.

Ma, S., Sigal, L., & Sclaroff, S. (2016). Learning activity progression in LSTMs for activity detection and early detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1942-1950).

Ma, X., Tao, Z., Wang, Y., Yu, H., & Wang, Y. (2015). Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transportation Research Part C: Emerging Technologies, 54*, 187-197.

Mamoshina, P., Vieira, A., Putin, E., & Zhavoronkov, A. (2016). Applications of deep learning in biomedicine. *Molecular pharmaceutics, 13*, 1445-1454.

Marc'Aurelio Ranzato, C. P., Chopra, S., & LeCun, Y. (2007). Efficient learning of sparse representations with an energy-based model. In *Proceedings of NIPS*.

Masci, J., Meier, U., Cireşan, D., & Schmidhuber, J. (2011). Stacked convolutional auto-encoders for hierarchical feature extraction. In *International Conference on Artificial Neural Networks* (pp. 52-59): Springer.

Mesnil, G., Dauphin, Y., Glorot, X., Rifai, S., Bengio, Y., Goodfellow, I. J., Lavoie, E., Muller, X., Desjardins, G., & Warde-Farley, D. (2012). Unsupervised and Transfer Learning Challenge: a Deep Learning Approach. *ICML Unsupervised and Transfer Learning, 27*, 97-110.

Microsoft. (2017). Microsoft Cognitive Toolkit. In.

Mohamed, A.-r., & Hinton, G. (2010). Phone recognition using restricted boltzmann machines. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on* (pp. 4354-4357): IEEE.

Montavon, G., & Müller, K.-R. (2012). Deep Boltzmann machines and the centering trick. In *Neural Networks: Tricks of the Trade* (pp. 621-637): Springer.

Moon, G. E., & Hamm, J. (2016). A Large-Scale Study in Predictability of Daily Activities and Places.

Morales, F. J. O., & Roggen, D. (2016). Deep convolutional feature transfer across mobile activity recognition domains, sensor modalities and locations. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers* (pp. 92-99): ACM.

Morales, J., & Akopian, D. (2017). Physical activity recognition by smartphones, a survey. *Biocybernetics and Biomedical Engineering*.

Munoz-Organero, M., & Ruiz-Blazquez, R. (2017). Time-Elastic Generative Model for Acceleration Time Series in Human Activity Recognition. *Sensors, 17*, 319.

Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)* (pp. 807-814).

Natarajasivan, D., & Govindarajan, M. (2016). Filter Based Sensor Fusion for Activity Recognition using Smartphone. *International Journal of Computer Science and Telecommunications, 7*, 26-31.
Nervana-Systems. (2017). Neon. In.
Neverova, N., Wolf, C., Lacey, G., Fridman, L., Chandra, D., Barbello, B., & Taylor, G. (2016). Learning human identity from motion patterns. *IEEE Access, 4*, 1810-1820.
Ng, A. (2011). Sparse autoencoder. *CS294A Lecture notes, 72*, 1-19.
Nguyen, D. T., Cohen, E., Pourhomayoun, M., & Alshurafa, N. (2016). SwallowNet: Recurrent Neural Network Detects and Characterizes Eating Patterns.
NVidia-Corps. (2017). NVidia DGX-1. In.
O'Donoghue, J., & Roantree, M. (2015). A framework for selecting deep learning hyper-parameters. In *British International Conference on Databases* (pp. 120-132): Springer.
Olshausen, B. A., & Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision research, 37*, 3311-3325.
Onofri, L., Soda, P., Pechenizkiy, M., & Iannello, G. (2016). A survey on using domain and contextual knowledge for human activity recognition in video streams. *Expert Systems with Applications, 63*, 97-111.
Ordonez, F. J., & Roggen, D. (2016). Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. *Sensors (Basel), 16*.
Ordóñez, F. J., & Roggen, D. (2016). Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. *Sensors, 16*, 115.
Oyedotun, O. K., & Khashman, A. (2016). Deep learning in vision-based static hand gesture recognition. *Neural Computing and Applications*, 1-11.
Page, A., Sagedy, C., Smith, E., Attaran, N., Oates, T., & Mohsenin, T. (2015). A flexible multichannel EEG feature extractor and classifier for seizure detection. *IEEE Transactions on Circuits and Systems II: Express Briefs, 62*, 109-113.
Palumbo, F., Gallicchio, C., Pucci, R., & Micheli, A. (2016). Human activity recognition using multisensor data fusion based on reservoir computing. *Journal of Ambient Intelligence and Smart Environments, 8*, 87-107.
Pereira, C. R., Pereira, D. R., Papa, J. P., Rosa, G. H., & Yang, X.-S. (2016). Convolutional Neural Networks Applied for Parkinson's Disease Identification. In *Machine Learning for Health Informatics* (pp. 377-390): Springer.
Pires, I. M., Garcia, N. M., Pombo, N., & Flórez-Revuelta, F. (2016). From data acquisition to data fusion: a comprehensive review and a roadmap for the identification of activities of daily living using mobile devices. *Sensors, 16*, 184.
Plötz, T., Hammerla, N. Y., & Olivier, P. (2011). Feature learning for activity recognition in ubiquitous computing. In *IJCAI Proceedings-International Joint Conference on Artificial Intelligence* (Vol. 22, pp. 1729).
Ponce, H., de Lourdes Martínez-Villaseñor, M., & Miralles-Pechúan, L. (2015). Comparative analysis of artificial hydrocarbon networks and data-driven approaches for human activity recognition. In *International Conference on Ubiquitous Computing and Ambient Intelligence* (pp. 150-161): Springer.
Ponce, H., Martínez-Villaseñor, M. d. L., & Miralles-Pechuán, L. (2016). A novel wearable sensor-based human activity recognition approach using artificial hydrocarbon networks. *Sensors, 16*, 1033.
Ponce, H., Miralles-Pechuán, L., & Martínez-Villaseñor, M. d. L. (2016). A Flexible Approach for Human Activity Recognition Using Artificial Hydrocarbon Networks. *Sensors, 16*, 1715.
Pouladzadeh, P., Kuhad, P., Peddi, S. V. B., Yassine, A., & Shirmohammadi, S. (2016). Food calorie measurement using deep learning neural network. In *Instrumentation and Measurement Technology Conference Proceedings (I2MTC), 2016 IEEE International* (pp. 1-6): IEEE.

Qian, N. (1999). On the momentum term in gradient descent learning algorithms. *Neural Networks, 12*, 145-151.

Rad, N. M., Bizzego, A., Kia, S. M., Jurman, G., Venuti, P., & Furlanello, C. (2015). Convolutional neural network for stereotypical motor movement detection in autism. *arXiv preprint arXiv:1511.01865*.

Rad, N. M., & Furlanello, C. (2016). Applying Deep Learning to Stereotypical Motor Movement Detection in Autism Spectrum Disorders. In *International Conference on Data Mining (ICDM 2016)*: IEEE.

Radu, V., Lane, N. D., Bhattacharya, S., Mascolo, C., Marina, M. K., & Kawsar, F. (2016). Towards multimodal deep learning for activity recognition on mobile devices. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct* (pp. 185-188): ACM.

Rahhal, M. M. A., Bazi, Y., AlHichri, H., Alajlan, N., Melgani, F., & Yager, R. R. (2016). Deep learning approach for active classification of electrocardiogram signals. *Information Sciences, 345*, 340-354.

Ravì, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., & Yang, G. Z. (2017). Deep Learning for Health Informatics. *IEEE journal of Biomedical and Health Informatics, 21*, 4-21.

Ravi, D., Wong, C., Lo, B., & Yang, G.-Z. (2016). A deep learning approach to on-node sensor data analytics for mobile or wearable devices. *IEEE journal of Biomedical and Health Informatics*.

Ravi, D., Wong, C., Lo, B., & Yang, G. Z. (2016). Deep learning for human activity recognition: A resource efficient implementation on low-power devices. In *2016 IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN)* (pp. 71-76).

Ravì, D., Wong, C., Lo, B., & Yang, G. Z. (2017). A Deep Learning Approach to on-Node Sensor Data Analytics for Mobile or Wearable Devices. *IEEE journal of Biomedical and Health Informatics, 21*, 56-64.

Reiss, A., & Stricker, D. (2012). Introducing a New Benchmarked Dataset for Activity Monitoring. In *2012 16th International Symposium on Wearable Computers* (pp. 108-109).

Rifai, S., Vincent, P., Muller, X., Glorot, X., & Bengio, Y. (2011). Contractive auto-encoders: Explicit invariance during feature extraction. In *Proceedings of the 28th international conference on machine learning (ICML-11)* (pp. 833-840).

Rodan, A., & Tino, P. (2011). Minimum Complexity Echo State Network. *IEEE Transactions on Neural Networks, 22*, 131-144.

Rodriguez, M., Orrite, C., Medrano, C., & Makris, D. (2016). One-Shot Learning of Human Activity With an MAP Adapted GMM and Simplex-HMM. *IEEE Trans Cybern*, 1-12.

Rogers, E., Kelleher, J. D., & Ross, R. J. (2016). Towards a Deep Learning-based Activity Discovery System.

Roggen, D., Calatroni, A., Rossi, M., Holleczek, T., Förster, K., Tröster, G., Lukowicz, P., Bannach, D., Pirkl, G., & Ferscha, A. (2010). Collecting complex activity datasets in highly rich networked sensor environments. In *Networked Sensing Systems (INSS), 2010 Seventh International Conference on* (pp. 233-240): IEEE.

Ronao, C. A., & Cho, S.-B. (2016). Human activity recognition with smartphone sensors using deep learning neural networks. *Expert Systems with Applications, 59*, 235-244.

Ronao, C. A., & Cho, S. B. (2016). Human activity recognition with smartphone sensors using deep learning neural networks. *Expert Systems with Applications, 59*, 235-244.

Ronaoo, C. A., & Cho, S.-B. (2015). Evaluation of deep convolutional neural network architectures for human activity recognition with smartphone sensors. In *Proc. of the KIISE Korea Computer Congress* (pp. 858-860).

Ruder, S. (2016). An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*.

Safi, K., Mohammed, S., Attal, F., Khalil, M., & Amirat, Y. (2016). Recognition of different daily living activities using hidden Markov model regression. In *Biomedical Engineering (MECBME), 2016 3rd Middle East Conference on* (pp. 16-19): IEEE.

Salakhutdinov, R., & Hinton, G. (2012). An efficient learning procedure for deep Boltzmann machines. *Neural computation, 24*, 1967-2006.

Salakhutdinov, R., & Hinton, G. E. (2009). Deep Boltzmann Machines. In *AISTATS* (Vol. 1, pp. 3).

Salakhutdinov, R., & Larochelle, H. (2010). Efficient Learning of Deep Boltzmann Machines. In *AISTATs* (pp. 693-700).

Salakhutdinov, R., Mnih, A., & Hinton, G. (2007). Restricted Boltzmann machines for collaborative filtering. In *Proceedings of the 24th international conference on Machine learning* (pp. 791-798): ACM.

Sargano, A. B., Angelov, P., & Habib, Z. (2017). A comprehensive review on handcrafted and learning-based action representation approaches for human activity recognition. *Applied Sciences, 7*, 110.

Sarkar, S., Reddy, K., Dorgan, A., Fidopiastis, C., & Giering, M. (2016). Wearable EEG-based Activity Recognition in PHM-related Service Environment via Deep Learning. *international Journal of Prognostics and Health Management, 7*, 10.

Sathyanarayana, A., Joty, S., Fernandez-Luque, L., Ofli, F., Srivastava, J., Elmagarmid, A., Arora, T., & Taheri, S. (2016). Sleep Quality Prediction From Wearable Data Using Deep Learning. *JMIR mHealth and uHealth, 4*.

Sathyanarayana, A., Joty, S., Fernandez-Luque, L., Ofli, F., Srivastava, J., Elmagarmid, A., Taheri, S., & Arora, T. (2016). Impact of Physical Activity on Sleep: A Deep Learning Based Exploration. *arXiv preprint arXiv:1607.07034*.

Savazzi, S., Rampa, V., Vicentini, F., & Giussani, M. (2016). Device-Free Human Sensing and Localization in Collaborative Human–Robot Workspaces: A Case Study. *IEEE Sensors Journal, 16*, 1253-1264.

Scherer, D., Müller, A., & Behnke, S. (2010). Evaluation of pooling operations in convolutional architectures for object recognition. In *International Conference on Artificial Neural Networks* (pp. 92-101): Springer.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks, 61*, 85-117.

Schulz, H., Cho, K., Raiko, T., & Behnke, S. (2015). Two-layer contractive encodings for learning stable nonlinear features. *Neural Networks, 64*, 4-11.

Shahroudy, A., Liu, J., Ng, T. T., & Wang, G. (2016). NTU RGB+D: A Large Scale Dataset for 3D Human Activity Analysis. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1010-1019).

Shahroudy, A., Ng, T.-T., Gong, Y., & Wang, G. (2016). Deep multimodal feature analysis for action recognition in RGB+ D videos. *arXiv preprint arXiv:1603.07120*.

Sheng, M., Jiang, J., Su, B., Tang, Q., Yahya, A. A., & Wang, G. (2016). Short-time activity recognition with wearable sensors using convolutional neural network. In *Proceedings of the 15th ACM SIGGRAPH Conference on Virtual-Reality Continuum and Its Applications in Industry-Volume 1* (pp. 413-416): ACM.

Shimizu, R., Yanagawa, S., Monde, Y., Yamagishi, H., Hamada, M., Shimizu, T., & Kuroda, T. (2016). Deep learning application trial to lung cancer diagnosis for medical sensor systems. In *SoC Design Conference (ISOCC), 2016 International* (pp. 191-192): IEEE.

Shoaib, M., Bosch, S., Incel, O. D., Scholten, H., & Havinga, P. J. (2014). Fusion of smartphone motion sensors for physical activity recognition. *Sensors, 14*, 10146-10176.

Shoaib, M., Bosch, S., Incel, O. D., Scholten, H., & Havinga, P. J. (2016). Complex human activity recognition using smartphone and wrist-worn motion sensors. *Sensors, 16*, 426.

Singh, S., Arora, C., & Jawahar, C. (2016). First person action recognition using deep learned descriptors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2620-2628).

Skymind. (2017). Deeplearning4j. In (Vol. 2017).

Song-Mi, L., Sang Min, Y., & Heeryon, C. (2017). Human activity recognition from accelerometer data using Convolutional Neural Network. In *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)* (pp. 131-134).

Song, Q., Zheng, Y.-J., Xue, Y., Sheng, W.-G., & Zhao, M.-R. (2017). An evolutionary deep neural network for predicting morbidity of gastrointestinal infections by food contamination. *Neurocomputing, 226*, 16-22.

Song, Q., Zheng, Y. J., Xue, Y., Sheng, W. G., & Zhao, M. R. (2017). An evolutionary deep neural network for predicting morbidity of gastrointestinal infections by food contamination. *Neurocomputing, 226*, 16-22.

Song, S., Chandrasekhar, V., Mandal, B., Li, L., Lim, J.-H., Sateesh Babu, G., Phyo San, P., & Cheung, N.-M. (2016). Multimodal multi-stream deep learning for egocentric activity recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 24-31).

Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research, 15*, 1929-1958.

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. In *Advances in neural information processing systems* (pp. 3104-3112).

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1-9).

Taylor, G. W., Hinton, G. E., & Roweis, S. T. (2007). Modeling human motion using binary latent variables. *Advances in neural information processing systems, 19*, 1345.

Tieleman, T., & Hinton, G. (2012). Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning, 4*.

Turaga, P., Chellappa, R., Subrahmanian, V. S., & Udrea, O. (2008). Machine Recognition of Human Activities: A Survey. *Ieee Transactions on Circuits and Systems for Video Technology, 18*, 1473-1488.

Um, T. T., Pfister, F. M. J., Pichler, D., Endo, S., Lang, M., Hirche, S., Fietzek, U., & Kulić, D. (2017). Data Augmentation of Wearable Sensor Data for Parkinson's Disease Monitoring using Convolutional Neural Networks. *arXiv preprint arXiv:1706.00527*.

Unger, M., Bar, A., Shapira, B., & Rokach, L. (2016). Towards latent context-aware recommendation systems. *Knowledge-Based Systems, 104*, 165-178.

Valipour, S., Siam, M., Jagersand, M., & Ray, N. (2016). Recurrent Fully Convolutional Networks for Video Segmentation. *arXiv preprint arXiv:1606.00487*.

Vepakomma, P., De, D., Das, S. K., & Bhansali, S. (2015). A-Wristocracy: Deep learning on wrist-worn sensing for recognition of user complex activities. In *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)* (pp. 1-6).

Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P.-A. (2008). Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th international conference on Machine learning* (pp. 1096-1103): ACM.

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., & Manzagol, P.-A. (2010). Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research, 11*, 3371-3408.

Vollmer, C., Gross, H.-M., & Eggert, J. P. (2013). Learning features for activity recognition with shift-invariant sparse coding. In *International Conference on Artificial Neural Networks* (pp. 367-374): Springer.

Vollmer, C., Gross, H. M., & Eggert, J. P. (2013). Learning Features for Activity Recognition with Shift-Invariant Sparse Coding. In V. Mladenov, P. KoprinkovaHristova, G. Palm, A. E. P. Villa, B. Appollini & N. Kasabov (Eds.), *Artificial Neural Networks and Machine Learning - Icann 2013* (Vol. 8131, pp. 367-374).

Wan, L., Zeiler, M., Zhang, S., Cun, Y. L., & Fergus, R. (2013). Regularization of neural networks using dropconnect. In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)* (pp. 1058-1066).

Wang, A., Chen, G., Shang, C., Zhang, M., & Liu, L. (2016). Human Activity Recognition in a Smart Home Environment with Stacked Denoising Autoencoders. In *International Conference on Web-Age Information Management* (pp. 29-40): Springer.

Wang, L. (2016). Recognition of human activities using continuous autoencoders with wearable sensors. *Sensors, 16*, 189.

Wang, L., Qiao, Y., & Tang, X. (2015). Action recognition with trajectory-pooled deep-convolutional descriptors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4305-4314).

Wang, X. M., Zhang, B., Zhang, F. P., Teng, G. W., Sun, Z. L., & Wei, J. M. (2016). Toward robust activity recognition: Hierarchical classifier based on Gaussian Process. *Intelligent Data Analysis, 20*, 701-717.

Wang, Z. L., Wu, D. H., Chen, J. M., Ghoneim, A., & Hossain, M. A. (2016). A Triaxial Accelerometer-Based Human Activity Recognition via EEMD-Based Features and Game-Theory-Based Feature Selection. *Ieee Sensors Journal, 16*, 3198-3207.

Wu, D., Wang, Z., Chen, Y., & Zhao, H. (2016). Mixed-kernel based weighted extreme learning machine for inertial sensor based human activity recognition with imbalanced dataset. *Neurocomputing, 190*, 35-49.

Wu, Z. Y., Ding, X. Q., & Zhang, G. R. (2016). A Novel Method for Classification of ECG Arrhythmias Using Deep Belief Networks. *International Journal of Computational Intelligence and Applications, 15*.

Xu, H., & Plataniotis, K. N. (2016). EEG-based affect states classification using Deep Belief Networks. In *Digital Media Industry & Academic Forum (DMIAF)* (pp. 148-153): IEEE.

Xu, X., Tang, J. S., Zhang, X. L., Liu, X. M., Zhang, H., & Qiu, Y. M. (2013). Exploring Techniques for Vision Based Human Activity Recognition: Methods, Systems, and Evaluation. *Sensors, 13*, 1635-1650.

Yalçın, H. (2016). Human activity recognition using deep belief networks. In *2016 24th Signal Processing and Communication Application Conference (SIU)* (pp. 1649-1652).

Yan, Y., Qin, X., Wu, Y., Zhang, N., Fan, J., & Wang, L. (2015). A restricted Boltzmann machine based two-lead electrocardiography classification. In *Wearable and Implantable Body Sensor Networks (BSN), 2015 IEEE 12th International Conference on* (pp. 1-9): IEEE.

Yanagimoto, M., & Sugimoto, C. (2016). Recognition of persisting emotional valence from EEG using convolutional neural networks. In *Computational Intelligence and Applications (IWCIA), 2016 IEEE 9th International Workshop on* (pp. 27-32): IEEE.

Yang, J. B., Nguyen, M. N., San, P. P., Li, X. L., & Krishnaswamy, S. (2015). Deep convolutional neural networks on multichannel time series for human activity recognition. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI), Buenos Aires, Argentina* (pp. 25-31).

Yao, R., Lin, G., Shi, Q., & Ranasinghe, D. (2017). Efficient Dense Labeling of Human Activity Sequences from Wearables using Fully Convolutional Networks. *arXiv preprint arXiv:1702.06212*.

Yao, S., Hu, S., Zhao, Y., Zhang, A., & Abdelzaher, T. (2016). DeepSense: A Unified Deep Learning Framework for Time-Series Mobile Sensing Data Processing. *arXiv preprint arXiv:1611.01942*.

Yao, S., Hu, S., Zhao, Y., Zhang, A., & Abdelzaher, T. (2017). Deepsense: A unified deep learning framework for time-series mobile sensing data processing. In (pp. 351-360): International World Wide Web Conferences Steering Committee.

Yi, Y., Cheng, Y., & Xu, C. (2017). Mining human movement evolution for complex action recognition. *Expert Systems with Applications, 78*, 259-272.

Yin, W., Yang, X., Zhang, L., & Oki, E. (2016). ECG Monitoring System Integrated With IR-UWB Radar Based on CNN. *IEEE Access, 4*, 6344-6351.

Younes, L. (1999). On the convergence of Markovian stochastic algorithms with rapidly decreasing ergodicity rates. *Stochastics: An International Journal of Probability and Stochastic Processes, 65*, 177-228.

Zappi, P., Lombriser, C., Stiefmeier, T., Farella, E., Roggen, D., Benini, L., & Tröster, G. (2008). Activity recognition from on-body sensors: accuracy-power trade-off by dynamic sensor selection. In *Wireless sensor networks* (pp. 17-33): Springer.

Zaremba, W. (2015). An empirical exploration of recurrent network architectures.

Zdravevski, E., Lameski, P., Trajkovik, V., Kulakov, A., Chorbev, I., Goleva, R., Pombo, N., & Garcia, N. (2017). Improving Activity Recognition Accuracy in Ambient-Assisted Living Systems by Automated Feature Engineering. *Ieee Access, 5*, 5262-5280.

Zebin, T., Scully, P. J., & Ozanyan, K. B. (2016). Human activity recognition with inertial sensors using a deep learning approach. In *SENSORS, 2016 IEEE* (pp. 1-3): IEEE.

Zeiler, M. D. (2012). ADADELTA: an adaptive learning rate method. *arXiv preprint arXiv:1212.5701*.

Zeng, M., Nguyen, L. T., Yu, B., Mengshoel, O. J., Zhu, J., Wu, P., & Zhang, J. (2014). Convolutional Neural Networks for human activity recognition using mobile sensors. In *6th International Conference on Mobile Computing, Applications and Services* (pp. 197-205).

Zhang, J., Shan, S., Kan, M., & Chen, X. (2014). Coarse-to-fine auto-encoder networks (cfan) for real-time face alignment. In *European Conference on Computer Vision* (pp. 1-16): Springer.

Zhang, J., & Wu, Y. (2017). Automatic sleep stage classification of single-channel EEG by using complex-valued convolutional neural network. *Biomedical Engineering/Biomedizinische Technik*.

Zhang, L., Wu, X., & Luo, D. (2015a). Human activity recognition with HMM-DNN model. In *2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC)* (pp. 192-197).

Zhang, L., Wu, X., & Luo, D. (2015b). Improving activity recognition with context information. In *2015 IEEE International Conference on Mechatronics and Automation (ICMA)* (pp. 1241-1246).

Zhang, L., Wu, X., & Luo, D. (2015a). Real-Time Activity Recognition on Smartphones Using Deep Neural Networks. In *Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom), 2015 IEEE 12th Intl Conf on* (pp. 1236-1242): IEEE.

Zhang, L., Wu, X., & Luo, D. (2015b). Recognizing Human Activities from Raw Accelerometer Data Using Deep Neural Networks. In *Machine Learning and Applications (ICMLA), 2015 IEEE 14th International Conference on* (pp. 865-870): IEEE.

Zhang, M., & Sawchuk, A. A. (2013). Human daily activity recognition with sparse representation using wearable sensors. *IEEE journal of Biomedical and Health Informatics, 17*, 553-560.

Zhao, R., Yan, R., Wang, J., & Mao, K. (2017). Learning to monitor machine health with convolutional bi-directional lstm networks. *Sensors, 17*, 273.

Zhao, Y., & He, L. (2014). Deep learning in the EEG diagnosis of Alzheimer's disease. In *Asian Conference on Computer Vision* (pp. 340-353): Springer.

Zheng, Y.-J., Ling, H.-F., & Xue, J.-Y. (2014). Ecogeography-based optimization: enhancing biogeography-based optimization with ecogeographic barriers and differentiations. *Computers & Operations Research, 50*, 115-127.
Zhou, X., Guo, J., & Wang, S. (2015). Motion recognition by using a stacked autoencoder-based deep learning algorithm with smart phones. In *International Conference on Wireless Algorithms, Systems, and Applications* (pp. 778-787): Springer.
Zhu, C., & Sheng, W. (2009). Multi-sensor fusion for human daily activity recognition in robot-assisted living. In *2009 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 303-304).
Zhu, F., Shao, L., Xie, J., & Fang, Y. (2016). From handcrafted to learned representations for human action recognition: a survey. *Image and vision computing, 55*, 42-52.
Zhu, J., Pande, A., Mohapatra, P., & Han, J. J. (2015). Using Deep Learning for Energy Expenditure Estimation with wearable sensors. In *E-health Networking, Application & Services (HealthCom), 2015 17th International Conference on* (pp. 501-506): IEEE.
Zhu, Q., Chen, Z., & Soh, Y. C. (2015). Using unlabeled acoustic data with locality-constrained linear coding for energy-related activity recognition in buildings. In *Automation Science and Engineering (CASE), 2015 IEEE International Conference on* (pp. 174-179): IEEE.
Zhu, Y., Zhao, X., Fu, Y., & Liu, Y. (2010). Sparse coding on local spatial-temporal volumes for human action recognition. In *Asian Conference on Computer Vision* (pp. 660-671): Springer.
Zouba, N., Bremond, F., & Thonnat, M. (2009). Multisensor fusion for monitoring elderly activities at home. In *Advanced Video and Signal Based Surveillance, 2009. AVSS'09. Sixth IEEE International Conference on* (pp. 98-103): IEEE.