CNN Based Approach for Activity Recognition using a Wrist-Worn Accelerometer

Madhuri Panwar[1], S. Ram Dyuthi[1], K. Chandra Prakash[1], Dwaipayan Biswas[2], Amit Acharyya[1], Koushik Maharatna[3], Arvind Gautam[1], Ganesh R. Naik[4]

Abstract— **In recent years, significant advancements have taken place in human activity recognition using various machine learning approaches. However, feature engineering have dominated conventional methods involving the difficult process of optimal feature selection. This problem has been mitigated by using a novel methodology based on deep learning framework which automatically extracts the useful features and reduces the computational cost. As a proof of concept, we have attempted to design a generalized model for recognition of three fundamental movements of the human forearm performed in daily life where data is collected from four different subjects using a single wrist worn accelerometer sensor. The validation of the proposed model is done with different pre-processing and noisy data condition which is evaluated using three possible methods. The results show that our proposed methodology achieves an average recognition rate of 99.8% as opposed to conventional methods based on *K*-means clustering, linear discriminant analysis and support vector machine.**

Keywords—**Activity recognition, Accelerometer, Deep learning, Convolutional Neural Network.**

I. Introduction

In the past decade, HAR has received a growing attention towards a variety of areas such as context-aware computing, surveillance-based security, mobile and pervasive computing, and ambient assistive living [1].

Activity recognition can be broadly classified into two parts - vision and sensor based recognition [1]. Vision based recognition is intuitive and information-rich but it usually suffers from privacy and ethical issues [2]. In particular, sensor-based activity recognition has the advantage of continuous monitoring of activities in a pervasive manner and has less dependence on their surroundings compared to vision based approaches [3]. There has been substantial progress in sensor based activity recognition and there exists a vast amount of work on low-cost, low-power, high-capacity, and miniaturized wearable sensors [1-6], [8-9].

Accelerometers are the most frequently used wearable sensors for measuring different activities due to their small size, low-power, low-cost, comfort, and convenience [4]. A variety of studies have explored using single as well as multiple sensors. Although multiple sensors can predict more accurate results [5] but they are too cumbersome and difficult to handle in practical cases [6]. Therefore, our proposed approach is framed using a single tri-axial accelerometer.

Over the past years, researchers have analyzed simple and complex activities using time series data and many of them focus on getting the most informative features. Feature selection is a critical step in the development of any classifier because feeding raw data directly on classifier doesn't produce the appropriate results. Therefore, classification is generally performed after the difficult process of appropriate feature selection that distinguishes different activities and subjects [4], [7]. Researches have focused on deep learning framework, primarily convolutional neural network (CNN) as it considerably reduces the effort on feature engineering (feature extraction and selection) [8] and also the strong generalization capability of this multi layered network has pushed the classification beyond human accuracy [10].

Inspired by the aforementioned points, in this paper, we have explored a deep learning algorithm i.e. CNN to recognize three elementary arm movements which are generally associated with activities of daily living. Here, we aimed to develop a model which is capable of recognizing these arm movements under real-world conditions using a single tri-axial accelerometer. The recognition of these arm movements in daily living activities over a period of time, can help in monitoring arm rehabilitation in pathologies associated with neurodegenerative diseases (stroke or cerebral palsy). Hence, in this *proof-of-concept* paper we explore with healthy subjects which can be further extended towards health monitoring applications. To the best of our knowledge, this is the first work where deep leaning framework using CNN is used to recognize elementary arm movements by considering possible real time conditions such as noise and subject variability.

The rest of the paper is structured as follows. Section II describes the experimental setup and data collection process. Section III presents the architectural details of the proposed CNN framework and overall model for recognizing upper limb movements. Section IV presents the results with a comparative analysis over existing approaches applied on same dataset. The conclusion are drawn in Section V.

II. Experimental Setup and Data Collection

A wearable Shimmer 9DoF sensing platform is used for this study, having tri-axis accelerometer, gyroscope, and a magnetometer. For our experiments we excluded gyroscope

This research is supported by Visvesvaraya PhD Scheme for Electronics & IT by the Department of Electronics and Information Technology (DeitY), Visvesvaraya Young Faculty Research Fellowship India and Indian Institute of Technology Hyderabad, Hyderabad, India.

[1]M.P, S.R.D, K.C.P, A.G and A.A are with Department of Electrical Engineering, Indian Institute of Technology Hyderabad, 502205, Telangana, India. (ee15resch01004, amit_acharyya)@iith.ac.in.

[2]D.B is with the Biomedical Circuits and Systems group in IMEC, Heverlee, Belgium. Dwaipayan.Biswas@imec.be.

[3]K.M is with Sch. of Electronics & Computer Science, University of Southampton, United Kingdom. km3@ecs.soton.ac.uk.

[4]G.N is with University of Technology Sydney, Australia. Ganesh.Naik@uts.edu.au.

978-1-5090-2809-2/17/$31.00 ©2017 IEEE

and magnetometer and only used the accelerometer (range ± 1.5 g). The data is collected by placing the device on wrist with a sampling rate of 50 Hz which is deemed enough to capture all the information of hand movements. This study involved four male subjects between 20 to 40 years of age, all right arm dominant. The experiments were conducted within an open laboratory with an attached kitchen at University of Southampton (UoS). Each subject performed a bespoke activity-list of "Making a cup of tea", a common activity performed in daily life, having repeated occurrences of the three arm movements (actions) used in daily life:

- *Action A* – Reach and retrieve an object (extension and flexion of the forearm).
- *Action B* – Lift cup to mouth (rotation of the forearm about the elbow).
- *Action C* – Perform pouring or (un)locking action (rotation of the wrist about long axis of forearm).

The activity-list "Making a cup of tea" (cf. Table 1) comprises 20 individual activities including 10 occurrences of Action A, and 5 each of Action B and Action C as shown in Table I. Four repetitions of the activity-list (Table 1) was performed at a comfortable speed in a kitchen and 10 min rest period between repetitions. The start and stop time of the activities are noted down by researcher; and segmented using the annotations from the researcher.

TABLE I- Activity list for 'Making-a-cup-of-tea'

Activity	Action
1. Fetch cup from desk	A
2. Place cup on kitchen surface	A
3. Fetch kettle	A
4. Pour out extra water from kettle	C
5. Put kettle onto charging point	A
6. Reach out for the power switch on the wall	A
7. Drink a glass of water while waiting for kettle to boil	B
8. Reach out to switch off the kettle	A
9. Pour hot water from the kettle in to cup	C
10. Fetch milk from the shelf	A
11. Pour milk into cup	C
12. Put the bottle of milk back on shelf	A
13. Fetch cup from kitchen surface	A
14. Have a sip and taste the drink	B
15. Have another sip while walking back to desk	B
16. Unlock drawer	C
17. Retrieve biscuits from drawer	A
18. Eat a biscuit	B
19. Lock drawer	C
20. Have a drink	B

III. Proposed Methodology

The classification performance of any model highly depends on data used for classification. The data collected from sensors contain noise and artifacts. Therefore, after collecting the raw data from the sensor, it needs to be pre-processed prior to classification. In our proposed model, we have first pre-processed the data using the different types of preprocessing techniques and is fed to CNN to identify the performed arm movements. Fig. 1 shows the overall process of recognizing the arm movements.

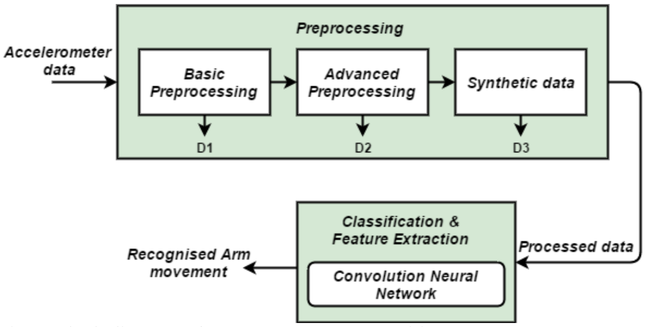

Fig. 1 Block diagram of arm movements recognition process.

A. Preprocessing

1) Basic Preprocessing: The main motivation of this pre-processing module is to meet the requirement of CNN model. The data collected for each activity has different sample lengths but CNN always takes fixed length input for training and testing for a particular design. Therefore, raw (Fig. 2(a)) data are normalized and then resampled to fixed length i.e. 64 points (Fig. 2(b)). The dataset created using basic pre-processing is named as D1.

2) Advanced Preprocessing: Observations have revealed that segmented data has the region where no activity is being done. Hence, to remove the undesired parts, mode of the data is computed and the corresponding data points are removed. In this process, we have also included smoothing on top of the basic pre-processing. The Dataset D2 is created by applying this processing on data (Fig. 2(c)). The different steps followed in advanced pre-processing are

- *Smoothing*- Makes the model insignificant to small perturbations.
- *Normalization*- Transformation of all data into specific range (0 -1).
- *Mode capturing*- Removal of undesired part where no activity has occurred and extract the required signal (Fig. 2(c).
- *Resampling*- Conversion of all data from variable size to fixed size.

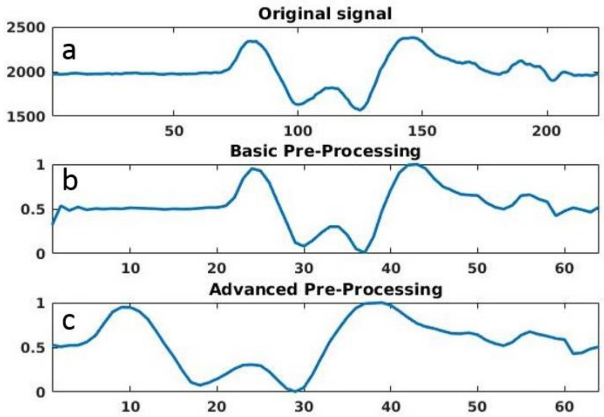

Fig. 2 Sample of raw and processed data where a represent the raw data, and b and c represent processed data with basic and advanced pre-processing respectively.

B. Classification

Inspired by the success of deep learning, we have explored CNN to recognize arm movements and proposed a new architecture for the same. CNN is a type of feed-forward neural network or a sequence of multiple layers which is inspired by biological processes. It eliminates the dependency on hand-crafted features and directly learns useful features from data itself. It is a combination of both feature extractor and classifier and mainly consists convolutional, pooling and fully connected layers which are explained later. Two different architectures *(architecture 1* and *architecture 2*) are proposed in this paper. The *architecture 1* is designed for smaller dataset (D1 and D2) and *architecture 2* is for D3, a relatively larger dataset. Fig. 3 presents the *architecture 1* and *2* in the same Fig. as both the architecture are having the same structure except the inclusion of extra *convolutional layer 3* in *architecture 2*, depicted by dotted line and also can be seen in Table II. This is due to the fact that D3 is having more training data due to inclusion of 20 different noise signals therefore *architecture 1* is not able to learn it properly which results in less accuracy. Therefore, one more layer is included in the *architecture 1* to improve the performance and named as *architecture 2*. The brief details of both the architectures are given in Table II where layer information, (*d*, *f*, *s*) in convolutional layer indicates *d* filters of spatial size *f*s*. Similarly, *Pooling layer*, *Fully connected* and *Dense layer* represent the spatial size used for max pooling (2*1), total number of nodes (480) and classes used for classification (3) respectively. The output shape gives the spatial size detail of output feature maps of each layer and parameter represents the total number of weights including biases. The description of training phase and all the layers used in proposed convolutional neural network are given below.

1) Training: It learns the parameters (weights) in each layer of the network using the back-propagation algorithm. For back propagation, stochastic gradient descent (SGD) is used to update the parameters. The forward and back propagation process are repeated until a stopping criterion is met i.e. maximum number of epochs is reached. There are four subjects and each has done the 4 trials so three possible case are considered for training, these are-

- *Case1*: training and testing with cross-validation
- *Case2*: training with 3 subjects data and testing with remaining 4th subject
- *Case3*: training with dataset collected from 3 trials and testing with remaining 1 trail for each subject.

2) Convolutional layer: It extracts the useful features by convolving the input feature map with the different filters. The weights of the filters are initialized using Gaussian distribution and later it automatically gets updated during back propagation. The output of each convolutional layer may have many number of feature map based on the filters used in that particular layer. It can be expressed as-

$$C_i^{I,j} = \sigma(b_j + \sum_{m=1}^{M} w_m^j x_{i+m-1}^{0j}) \qquad (1)$$

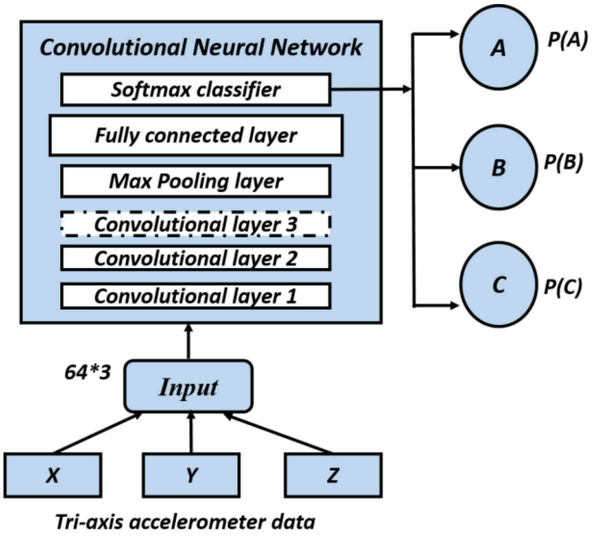

Fig. 3 Architectures of proposed CNN model for recognition of elementary arm movements where dotted line around convolutional layer 3 shows that architecture 1 for D1,D2 doesn't has this layer whereas it is included in architecture 2 (D3).

TABLE II. ARCHITECTURE DETAILS OF CNN

Architecture 1 : For dataset D1 and D2			
Layer (type)	**layer information**	**Output shape**	**Parameter**
Convolution 1 (conv2d_1)	(5,9,3)	(5,64,3)	140
Convolution 2 (conv2d_2)	(5,5,3)	(5,64,3)	300
Max-pooling (Max2d_1)	(2,1)	(5,32,3)	0
Fully-connected (Flatten_1)	(480)		0
Dense (dense_1)	(3)		1443
Architecture 2 : For dataset D3			
Convolution 1 (conv2d_1)	(5,9,3)	(5,64,3)	140
Convolution 2 (conv2d_2)	(5,7,3)	(5,64,3)	530
Convolution 3 (conv2d_3)	(5,5,3)	(5,64,3)	380
Max-pooling (Max2d_1)	(2,1)	(5,32,3)	0
Fully-connected (Flatten_1)	(480)		0
Dense (dense_1)	(3)		1443

Where *I is* the layer index, σ is the activation function, b_j is the bias term for the jth feature map, M is the filter size, and w_m^j is the weight of the jth feature map for mth filter index.

3) Rectified linear unit (ReLU): This is a non-linear activation function which has almost similar response as neuron. Therefore, widely used in deep learning framework.

4) Pooling layer: It is a form of nonlinear subsampling which is used to reduce the size of the feature map. This results in less number of parameters and computations. It also provides slight translation invariance.

5) Fully connected layer: High level reasoning in neural network is done through fully connected layer. It is similar to traditional MLP where all input neurons are connected to previous layer. Al the output features of previous layer are taken to determine the correlation between the features and a particular class.

6) Softmax: It is used as a classifier where each class is presented with some probability and a particular class is predicted based on highest probability.

IV. Results and analysis

The proposed methodology is implemented in Keras with Tensorflow backend using python environment. A number of experiments are carried out using three different dataset created by different pre-processing steps that highlight the role of pre-processing. The validation of proposed approach is done by a comparative study with existing methods. These experiments also show the impact of pre-processing on model performance. According to [10-11], the accuracy of activity model is affected by interclass variability. Therefore, evaluation of each dataset is done by taking possible real time cases for training and testing as explained in section III. These three cases are used to evaluate the performance of model and named as-

- *Cross-validation evaluation*-Training and testing with cross-validation.
- *Inter-subject evaluation* - Training with 3 subjects and testing with other subject
- *Hybrid evaluation*- Out of 4 repetitions, Training with 3 sets of data taken from each subject and testing with remaining one set from each subject.

A. Case study

It can be seen from table III that D1, D2 and D3, have less accuracy during inter subject evaluation than cross-validation and hybrid evaluation, due to the fact that testing is done with subject's data not used for training. It is also noted that basic per-processing is having less accuracy than advanced during all evaluations, because D1 dataset contains less useful information resulting in misclassification. Similarly, system trained with D3 dataset has highest accuracy than D1 and D2 in cross-validation evaluation because it is trained with 20 different types of noise which make the system robust to noisy data resulting in decrease of rate of misclassification.

B. Comparative study

The proposed model is compared with the existing methods to validate its performance. Table IV shows the comparative analysis of proposed methodology with K-means clustering, LDA and SVM. It can be seen from the Table that our proposed methodology outperformed with the same evaluation method used in the existing methods and also with inter-subject and Hybrid evaluation, our proposed model is giving better accuracy than existing methods. These analysis results show that our proposed methodology is robust to interclass variability (Patient independent-generalized system), noise resistant (reliable), and provides good accuracy than all three conventional methods. It also has low computation cost than existing methods as feature extraction computation is excluded in this methodology. These unique features proved robustness of our proposed methodology.

TABLE III. Classification Accuracy Analysis

Evaluation Method	Accuracy		
	Basic Pre-processing	*Advanced Pre-processing*	*Synthetic data*
Cross-Validation	90.6	92.5	99.8
Inter-subject	85.9	87.5	87.8
Hybrid	89.0	89.0	90.0

TABLE IV. Comparative Analysis with Conventional Methods (Cross-validation Evaluation)

Study	Method	Accuracy (%)
Biswas [9] (2015)	LDA	45
Biswas [9]	SVM	53.75
Biswas [9]	K-means clustering	87.5
Our study	CNN	**99.8**

V. Conclusion

In this work, we have studied the robustness of deep learning framework for predicting arm movements performed in daily activity using a wrist worn tri-axial accelerometer. The evaluation of different pre-processing steps, training with noisy data and comparative analysis with conventional methods (SVM, LDA and K-means clustering) demonstrated that: 1) CNN has great potential in handling the feature engineering process; 2) produces high accuracy if design parameters are define in an efficient way 3) able to classify daily living activities in real-time and practical scenarios. Overall, this work contributed towards proposing a, generalized, low-computational cost and noise resistant model for recognizing arm movements using a wrist-worn sensor based on deep learning framework. This proposed work can be further extended towards evaluating more number of subjects and also towards people suffering from neurodegenerative disease.

References

[1] Chen, Liming, et al. "Sensor-based activity recognition." IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 42.6 (2012): 790-808.

[2] Gu, Tao, et al. "A pattern mining approach to sensor-based human activity recognition." IEEE Transactions on Knowledge and Data Engineering 23.9 (2011): 1359-1372.

[3] Zhang, Mi, and Alexander A. Sawchuk. "Human daily activity recognition with sparse representation using wearable sensors." *IEEE journal of Biomedical and Health Informatics* 17.3 (2013): 553-560.

[4] Zheng, Yonglei, et al. "Physical Activity Recognition from Accelerometer Data Using a Multi-Scale Ensemble Method." *IAAI*. 2013.

[5] Maurer, Uwe, et al. "Activity recognition and monitoring using multiple sensors on different body positions." *Wearable and Implantable Body Sensor Networks, 2006. BSN 2006. International Workshop on*. IEEE, 2006.

[6] Lara, Oscar D., and Miguel A. Labrador. "A survey on human activity recognition using wearable sensors." *IEEE Communications Surveys and Tutorials* 15.3 (2013): 1192-1209.

[7] Zhang, Mi, and Alexander A. Sawchuk. "Motion primitive-based human activity recognition using a bag-of-features approach." *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*. ACM, 2012.

[8] Hammerla, Nils Y., Shane Halloran, and Thomas Ploetz. "Deep, convolutional, and recurrent models for human activity recognition using wearables." arXiv preprint arXiv:1604.08880 (2016).

[9] Biswas, Dwaipayan, et al. "Recognizing upper limb movements with wrist worn inertial sensors using k-means clustering classification." *Human movement science* 40 (2015): 59-76.

[10] Pullini, Antonio, et al. "A Heterogeneous Multi-Core System-on-Chip for Energy Efficient Brain Inspired Computing." *IEEE Transactions on Circuits and Systems II: Express Briefs* (2017).