Complex Nurse Care Activity Recognition Using Statistical Features

Promit Basak
Electrical and Electronic Engineering
University of Dhaka
Dhaka, Bangladesh
basakpromit@gmail.com

Shahamat Mustavi Tasin
Electrical and Electronic Engineering
University of Dhaka
Dhaka, Bangladesh
tasin.mustavi@gmail.com

Malisha Islam Tapotee
Electrical and Electronic Engineering
University of Dhaka
Dhaka, Bangladesh
malishatapotee@gmail.com

Md. Mamun Sheikh
Electrical and Electronic Engineering
University of Dhaka
Dhaka, Bangladesh
mamunsheikhdueee52@gmail.com

A.H.M. Nazmus Sakib
Electrical and Electronic Engineering
University of Dhaka
Dhaka, Bangladesh
nazmussakib2970@gmail.com

Sriman Bidhan Baray
Electrical and Electronic Engineering
University of Dhaka
Dhaka, Bangladesh
bidhanbaray@gmail.com

M.A.R. Ahad
Electrical and Electronic Engineering, University of Dhaka, Bangladesh
Department of Media Intelligent, Osaka University, Japan
atiqahad@du.ac.bd

ABSTRACT

Human activity recognition has important applications in healthcare, human-computer interactions and other arenas. The direct interaction between the nurse and patient can play a pivotal role in healthcare. Recognizing various activities of nurses can improve healthcare in many ways. However, it is a very daunting task due to the complexities of the activities. "The 2nd Nurse Care Activity Recognition Challenge Using Lab and Field Data" provides sensor-based accelerometer data to predict 12 activities conducted by the nurses in both the lab and real-life settings. The main difficulty of this dataset is to process the raw data because of a high imbalance among different classes. Besides, all activities have not been performed by all subjects. Our team, 'Team Apophis' has processed the data by filtering noise, applying windowing technique on time and frequency domain to extract various features from lab and field data distinctly. After merging lab and field data, the 10-fold cross-validation technique has been applied to find out the model of best performance. We have obtained a promising accuracy of 65% with an F1 score of 40% on this challenging dataset by using the Random Forest classifier.

UbiComp/ISWC '20 Adjunct, September 12–16, 2020, Virtual Event, Mexico

ACM ISBN 978-1-4503-8076-8/20/09...$15.00
https://doi.org/10.1145/3410530.3414338

CCS CONCEPTS

• Computing method → Feature extraction → Statistical features; • Learning paradigms → Supervised → Classification; • Algorithm → Random Forest.

KEYWORDS

Activity recognition; Nurse care; Healthcare; Statistical features; Random Forest

ACM Reference Format:

Promit Basak, Shahamat Mustavi Tasin, Malisha Islam Tapotee, Md. Mamun Sheikh, A.H.M. Nazmus Sakib, Sriman Bidhan Baray, and M.A.R. Ahad. 2020. Complex Nurse Care Activity Recognition Using Statistical Features. In *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers (UbiComp/ISWC '20 Adjunct), September 12--16, 2020, Virtual Event, Mexico. ACM New York, New York, 6 pages.* https://doi.org/10.1145/3410530.3414338

1 INTRODUCTION

Human Activity Recognition (HAR) has been an important research topic for quite some time. Over the years, there has been a huge development in microelectronics and computer systems, introducing key features like high computational power, small size, and low cost [9]. This has caused a rise in activity recognition using wearable sensors and many kinds of research were performed [17,18]. However, recognition of activities such as nurse care activities is not much explored due to different kinds of complications. The complications include lack of availability of open datasets, collection of proper data, complex body movements, etc. Moreover, such data require huge cleaning process due to the presence of redundant, ambiguous, and unusable data. In this paper, we "Team Apophis" provide an approach to the "The 2nd Nurse Care Activity Recognition Challenge Using Lab and Field Data".

Nurse care activity recognition can benefit healthcare in several ways. Feedback of the activities can be provided to the caregiver, which can be utilized for better methods of execution. This can improve nurse training, leading to a better understanding and improvements in medical care [19]. As a result, unnecessary activities and excessive work can be reduced. At the same time, these approaches are beneficial to patients because the overall care process is optimized, thus resulting in shorter hospitalization times and lower costs [6].

However, there are a number of difficulties dealing with such activity recognition. For instance, nurses perform these activities to a patient rather than performing on themselves. This creates a wide range of possibilities of performing the activities differently by different nurses depending on the situation [8]. In addition, class imbalance is often observed in such data. And problems like non-uniform sampling rate, presence of redundant data, absence of precise timestamp, etc., make preparation of the data more difficult.

Previously, in HAR a number of approaches had been taken depending on the dataset and the type of data. For example, in case of data collected from the accelerometers of the smart-phones, Iterative Dichotomiser 3 Decision Tree (ID3 DT) was used among machine learning techniques. From deep learning techniques, Artificial Neural Network (ANN) was used [7]. It is noteworthy that the "Nurse Care Activity Recognition Challenge" also took place in 2019, where the top 2 teams were selected, Team IITDU and Team TDU-DSML. In this challenge, data from motion capture, Meditag, and accelerometer were collected and used for activity recognition [12]. For training and classification, Team IITDU used an ensemble of K-Nearest Neighbors (KNN) classifiers on different types of features extracted from the given sensors [5] and Team TDU-DSML proposed an activity recognition algorithm that uses an ST-GCN to process 3D motion capture data for that challenge [2].

The paper is outlined as follows: Section 2 covers the dataset. In Section 3, we describe our method to process the raw sensor data, which means data-preprocessing and feature extraction methods. Then we evaluate suitable machine learning models to recognize the complex nurse care activities in Section 4. And from the experimental results, we select the model that performs the best with the training data. Finally, we discuss the results in Section 5, and conclude the paper in Section 6 along with possible future improvements.

2 CHALLENGE DATASET

The dataset (Heiseikai data, nurse care activity dataset) for "The 2nd Nurse Care Activity Recognition Challenge Using Lab and Field Data" [13] is used in this paper which consists of care activities that nurses perform in the care facilities. The activities are categorized in 3 principle types **(**Help in Mobility, Assistance in Transfer, Position change**)**, and further divided in the following way:

A. **Help in Mobility**
  1. Guide (From the front)
  2. Partial assistance
  3. Walker
  4. Wheelchair

B. **Assistance in Transfer**
  5. All assistance
  6. Partial assistance (From the front)
  7. Partial assistance (From the side)
  8. Partial assistance (From the back)

C. **Position Change**
  9. To supine position
  9. To Right lying position
  10. To Left lying position
  11. Lower body lifting
  12. Horizontal movement

It is noteworthy that there are two different activities ("To supine position" and "To Right lying position") labeled the same in the challenge description [14] and also in the provided dataset. We have considered both activities identical. The data collection process took place in two different environments - in a lab experiment and in real field. The lab-data have been collected in the Smart Life Care Unit of the Kyushu Institute of Technology in Japan. In this experiment, 2 professional nurses participated as subjects. The accelerometer sensor in the mobile phone attached in the right arm using an armband has been used to collect the data. The field data has been collected in a similar way but in a care facility in Japan, which contains data collected from 47 nurses. For training and testing, data of 6 nurses and 3 nurses have been provided respectively. The resulting training data contains data collected from a total of 8 nurses. For both field and lab data, the sampling rate is 60Hz. Additionally, the training data contains unlabeled data from 3 more nurses which were unusable. The test data contains field data collected from 3 nurses.

3 METHOD

In this section, proposed method for the "The 2nd Nurse Care Activity Recognition Challenge Using Lab and Field Data" [15] has been described.

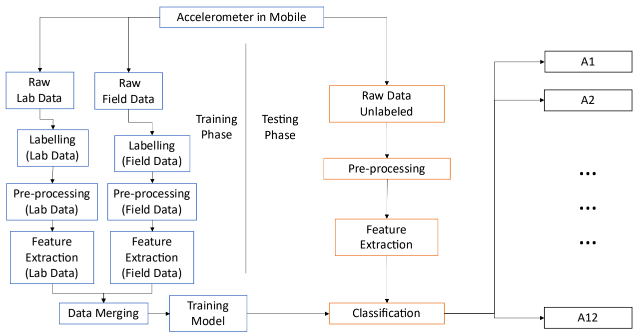

Figure 1: Basic structure of Nurse Care Activity Recognition

Figure 1 describes the overview of the implementation of our process. The total process is done in 3 steps: Pre-processing, Feature Extraction, and Classification. After getting the sensor data, both train and test data have been processed according to the figure. In the training set, field and lab data have been merged before training the model. The twelve activities conducted by the nurses have been presented as A1, A2 and so on.

3.1 PRE-PROCESSING

The training data is divided into two segments: Field Data and Lab Data. Raw data is actually accelerometer-based data presenting 3 axes (X, Y, and Z) without labeling. The data was then labeled after removing the duplicates from the given raw data and converting all the date-time into the same time zone. It is worth mentioning that several data were found unusable as there was not enough data to label all the accelerometer readings. Also, some raw lab data have the same starting and finishing time which is incompatible. So, 59 sec. has been added at finishing time on those raw lab data. The datasets were not sampled uniformly. Without resampling, it is not possible to remove noise from the raw data. That is why we have resampled the data at 20 Hz. Here, median filter has been used to remove outliers. Also, to reduce the noise level, low pass filter has been used at 1 Hz and 3 Hz frequency for field and lab data respectively. Then, the raw data has been analyzed in time and frequency domain.

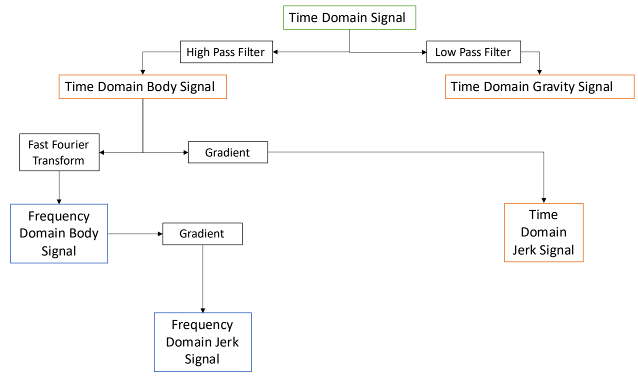

Figure 2: Processing of time domain signal

Figure 2 demonstrates how high pass and low pass filters have been used respectively to obtain time domain body and gravity signal. In case of field data, the corner frequency for both filters is 0.1 Hz where for lab data it is 0.3 Hz. Time domain jerk signal has been obtained by applying gradient on time domain body signal which implies the rate of change of acceleration. Fast Fourier Transform (FFT) has been introduced to gain the frequency domain signals from time domain. Again, applying gradient on frequency domain body signal, frequency domain jerk signal has been obtained. Then, all the five signals have been sampled at 15 sec non-overlapping sliding windows. In addition, for activities other than 4, 5, and 7, signals have been sampled again at 5s and 10s windows and have been added to the existing data to reduce the imbalance.

3.2 FEATURE EXTRACTION

The three time-domain signals and two frequency-domain signals have been studied to extract features from each window where each signal consists of three-axis values. Standard measures taken in HAR literature indicates that statistical features may play an important role in recognizing complex activities [1]. In this time series data, mean and median have been used to measure the central tendency. Moreover, standard deviation and inter-quartile range have been applied to identify the variability of data. And to measure the randomness of data, entropy has been introduced. Then, Correlation Coefficient has been used which gives better understanding to find out the similarities between axes.

In frequency domain signal, skewness and kurtosis features have been used to measure the symmetry. Also, Max Indices feature in frequency domain signal has been introduced to find out the index of dominant frequency. Magnitudes of all domain signals have been obtained by using the following equation:

$$A = \sqrt{A_x^2 + A_y^2 + A_z^2}$$

where, A represents the magnitude and A_x, A_y, A_z are the values of that signal corresponding with the axis. The extracted features which have been used in training and testing are tabulated in the Appendix section. Finally, a total of 218 features has been extracted for all 3-axes from both time and frequency domain signals.

3.3 CLASSIFICATION

The training data has a total 25525 samples and 218 features after merging the field and lab data. Then, all the features were normalized. Using this data, we have trained different machine learning models and evaluated them. At first, 10-fold cross validation has been applied to identify the best model to train. For high imbalance in class contributions, Decision tree algorithms have been popular choices [4]. On the other hand, previously K-Nearest Neighbors have been found effective for Nurse Care activities [5]. So, XGBoost classifier [3], KNN classifier [11] and Random Forest classifier [10] have been chosen for training and evaluation. Among these models Random Forest performed best. Therefore, to predict the activities of the test data we have used this classifier on training data.

4 RESULTS

In this section, we have discussed our experimentation on training data using 3 different models: K-Nearest Neighbor classifier (KNN), XGBoost classifier (XGB), and Random Forest classifier (RF). Table 1 contains the result obtained from these models using 10-fold cross-validation.

Table 1: Performance comparison of the three models

Model	Accuracy (%)	F1 score (%)
KNN	53	30
RF	65	40
XGB	60	33

From the Table, we can see that RF provides comparatively better results than KNN and XGB with 65% accuracy and 40% F1 score. That is why we have chosen the RF model for recognizing the activities from test data. Further, using 75% of the training data used as the experimental training set, we trained the RF model and then tested the model on the remaining 25% data. Table 2 contains the classification report obtained from the experimental test data where A1 means activity ID 1, A2 means activity ID 2, and so on. And the columns corresponding to each class represents the precision, recall, F1 score, and the support for each activity. Here, the accuracy column contains the overall accuracy while classifying the twelve activities in the experimental test data.

Table 2: Activity-wise performance of RF model (in %)

<table>
<tr><th>ID</th><th>Precision</th><th>Recall</th><th>F1 score</th><th>Support</th><th>Accuracy</th></tr>
<tr><td>A1</td><td>68</td><td>73</td><td>70</td><td>453</td><td rowspan="14">71</td></tr>
<tr><td>A2</td><td>40</td><td>33</td><td>37</td><td>138</td></tr>
<tr><td>A3</td><td>46</td><td>53</td><td>49</td><td>413</td></tr>
<tr><td>A4</td><td>70</td><td>86</td><td>77</td><td>775</td></tr>
<tr><td>A5</td><td>95</td><td>85</td><td>89</td><td>1144</td></tr>
<tr><td>A6</td><td>29</td><td>16</td><td>21</td><td>112</td></tr>
<tr><td>A7</td><td>72</td><td>59</td><td>64</td><td>420</td></tr>
<tr><td>A8</td><td>29</td><td>14</td><td>19</td><td>95</td></tr>
<tr><td>A9</td><td>56</td><td>42</td><td>48</td><td>159</td></tr>
<tr><td>A10</td><td>34</td><td>15</td><td>20</td><td>117</td></tr>
<tr><td>A11</td><td>11</td><td>04</td><td>06</td><td>95</td></tr>
<tr><td>A12</td><td>73</td><td>88</td><td>80</td><td>1090</td></tr>
<tr><td>Macro avg.</td><td>52</td><td>47</td><td>48</td><td rowspan="2"></td></tr>
<tr><td>Weighted avg.</td><td>69</td><td>71</td><td>69</td></tr>
</table>

From the Table 2, it is evident that activities A6, A8, A10, and A11 have generated poor results due to poor support, Figure 3 showing accuracies and F1 scores of each class also demonstrate the same. On the other hand, classes with a higher number of instances or support like A1, A4, A5, and A12 have generated significantly better results. This shows that these activities are comparatively more distinguishable and our model is expected to perform well for these activities.

However, in the confusion matrix shown in Figure 4, the activities A6, A8, A10, and A11 have been very frequently confused with activity A12. This shows that our model has struggled to identify these activities. Also, other activities except A4, A5, and A7 were also confused a number of times with A12. This is because the number of instances of these 3 activities were high. So, these activities were more distinguishable and hence, were less likely to be confused with A12.

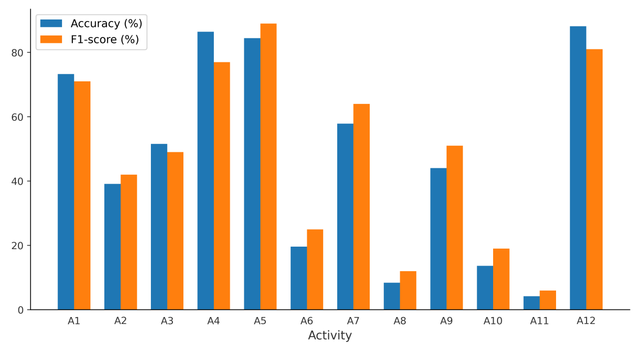

Figure 3: Accuracies and F1 scores obtained for different activities using RF

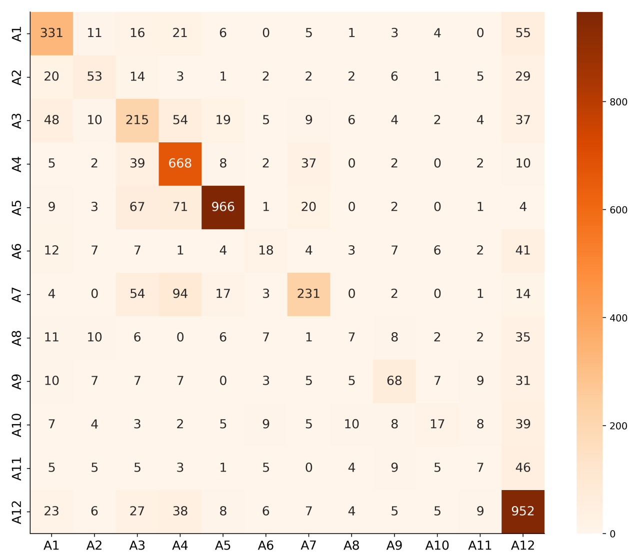

Figure 4: Confusion matrix obtained using RF

5 DISCUSSION

In the training data, activities A4 and A5 contained a much higher number of instances than that of other activities. For this reason, to reduce imbalance from data, some methods have been introduced. In this regard, we have at first removed outliers from the data corresponding to these activities using Clustering Based Local Outlier Factor (CBLOF). Then to perform under-sampling, we have randomly removed data from A4 and A5. After that, to increase the number of instances of other activities, we have again extracted features from the obtained signals using 5 sec. and 10 sec. non-overlapping windows. Finally, we have added these data using multiple windowing to the previously obtained training data. We have used this less imbalanced data to train our model. This has significantly improved our model but reduced the overall accuracy due to under-sampling. We have also tried Synthetic Minority Oversampling Technique (SMOTE) to balance the data on the experimental training set but it reduced the F1 score on the experimental test set from 48% to 43%. Also, using SMOTE, the accuracy also reduced from 71% to 65% in the experimental test set. That is why we have chosen the multiple-windowing technique over SMOTE.

There have been some additional issues while dealing with lab data. Firstly, in the data containing activity labels, the time ranges were given in minutes, causing the presence of arbitrary data. Also, in some cases, the starting and finishing time of some activities were the same. We had to add some seconds to the finishing time of such cases. This has generated even more arbitration. Because of this, many of the resulting features contained incorrect data, causing more misclassifications.

It is noteworthy that there is no data for activities A8 and A11 in the field data. So, we trained the model excluding these 2 classes and obtained 81% accuracy and 79% F1 score only using the field data. After adding lab data, the accuracy and F1 score have dropped to 69% and 48% respectively. Thus, merging lab data with field data has lowered the performance of our model.

6 CONCLUSION

In this paper, we have pre-processed the raw sensor data and extracted statistical features to train different models. Among them, the Random Forest classifier performed the best. Some important observations have been noted which should be addressed to get better outcomes. A better result might be obtained if we had applied an overlapping sliding window instead of a non-overlapping one. Removal of arbitrary data from lab data can also improve the model. In the future, we want to work on those issues and extract other important features. Furthermore, we want to implement other Decision Tree based approaches such as CART, ID3, C4.5/C5 etc. on this data. The recognition result for the testing dataset will be presented in the summary paper of the challenge [16].

REFERENCES

[1] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz. *A Public Domain Dataset for Human Activity Recognition Using Smartphones.*

[2] Xin Cao, Wataru Kudo, Chihiro Ito , Masaki Shuzo, and Eisaku Maeda. 2019. Activity recognition using ST-GCN with 3D motion data. *UbiComp/ISWC 2019- - Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers* (2019).

[3] Tianqi Chen and Carlos Ernesto Guestrin. XGBoost: A Scalable Tree Boosting System. DOI: https://doi.org/10.1145/2939672.2939785.

[4] David A. Cieslak and Nitesh V. Chawla. 2008. Learning decision trees for unbalanced data. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).* 5211 LNAI, PART 1 (2008), 241–256. DOI: https://doi.org/10.1007/978-3-540-87479-9_34

[5] Data Description - Second Nurse Care Activity Recognition Challenge: *https://abc-research.github.io/nurse2020/data/*. Accessed: 2020-07-10.

[6] Md. Eusha Kadir, Pritom Saha Akash, Sadia Sharmin, Amin Ahsan Ali, and Mohammad Shoyaib. 2019. Can a simple approach identify complex nurse care activity? *UbiComp/ISWC 2019- - Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers.* September (2019), 736–740. DOI: https://doi.org/10.1145/3341162.3344859

[7] Sozo Inoue, Naonori Ueda, Yasunobu Nohara, and Naoki Nakashima. 2016. Understanding Nursing Activities with Long-term Mobile Activity Recognition with Big Dataset. *Proceedings of the ISCIE International Symposium on Stochastic Systems Theory and its Applications.* (2016). DOI: https://doi.org/10.5687/sss.2016.1

[8] Charmi Jobanputra, Jatna Bavishi, and Nishant Doshi. 2019. Human activity recognition: A survey. *Procedia Computer Science* (2019).

[9] Paula Lago, Sayeda Shamma Alia, Shingo Takeda, Tittaya Mairittha, Nattaya Mairittha, Farina Faiz, Yusuke Nishimura, Kohei Adachi, Tsuyoshi Okita, François Charpillet, and Sozo Inoue. 2019. Nurse care activity recognition challenge: Summary and results. *UbiComp/ISWC 2019- - Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers* (2019).

[10] Óscar D. Lara and Miguel A. Labrador. 2013. A survey on human activity recognition using wearable sensors. *IEEE Communications Surveys and Tutorials.* (2013). DOI: https://doi.org/10.1109/SURV.2012.110112.00192.

[11] Sozo Inoue, Sayeda Shamma Alia, Paula Lago, "Nurse Care Activities Datasets: In laboratory and in real field", IEEE Dataport, 2020. [Online]. Available: http://dx.doi.org/10.21227/jem3-ap07. Accessed: Jul. 17, 2020.

[12] Nurse Care Activity Recognition Challenge | IEEE DataPort: *https://ieee-dataport.org/competitions/nurse-care-activity-recognition-challenge.* Accessed: 2020-07-26.

[13] Yu L. Pavlov. 2019. Random forests. *Random Forests.* (2019), 1–122. DOI: https://doi.org/10.1201/9780367816377-11.

[14] Second Nurse Care Activity Recognition Challenge: *https://abc-research.github.io/nurse2020/.* Accessed: 2020-07-10.

[15] Zhongheng Zhang. 2016. Introduction to machine learning: K-nearest neighbors. *Annals of Translational Medicine.* 4, 11 (2016), 1–7. DOI: https://doi.org/10.21037/atm.2016.03.37.

[16] Sayeda Shamma Alia, Paula Lago, Kohei Adachi, Tahera Hossain, Hiroki Goto, Tsuyoshi Okita and Sozo Inoue, "Summary of the 2nd Nurse Care Activity Recognition Challenge Using Lab and Field Data", Proc. HASCA 2020.

[17] Md Atiqur Rahman Ahad, Anindya Das Antar, Masud Ahmed, "IoT Sensor-Based Activity Recognition - Human Activity Recognition", Springer Nature, 2020.

[18] Md Atiqur Rahman Ahad, "Vision and Sensor based Human Activity Recognition: Challenges Ahead", Chapter 2, in "Advancements in Instrumentation and Control in Applied System Applications", IGI Global, pp. 17-35, 2020.

[19] Md Atiqur Rahman Ahad, Anindya Das Antar, and Omar Shahid, "Vision-based Action Understanding for Assistive Healthcare: A Short Review", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, USA, pp. 1-11, 2019.

APPENDIX

Table 3: Miscellaneous information about the experiment

Used sensor modalities	3 axis accelerometer
Features used	Mean value, magnitude, standard deviation, median, maximum, minimum, signal magnitude area, energy measure (sum of the squares divided by the number of values), inter-quartile range, signal entropy, correlation coefficient between two axes, skewness of the frequency domain signals, kurtosis of the frequency domain signals, index of the frequency component with the largest magnitude
Programming language	Python 3
Libraries used	NumPy, Pandas, SciPy, Scikit-learn, xgboost, matplotlib, seaborn
Window size	15 seconds non-overlapping
Training and testing time	Training set (approx. 50 million data units): Approximately, 350 minutes for Data processing, Feature Extraction, Training, and Prediction.
	Test set (approx. 40 million data units): Approximately, 220 minutes for Data processing, Feature Extraction, and Prediction.
Machine specification (RAM, CPU)	RAM 12GB, Disk 107 GB CPU: Intel Xeon @2.20 GHz, Cores: 2