❒ 927

A wearable device for machine learning based elderly's activity tracking and indoor location system

Nour Eddin Tabbakha, Chee-Pun Ooi, Wooi-Haw Tan, Yi Fei Tan
Faculty of Engineering, Multimedia University, Cyberjaya, Selangor, Malaysia

Article Info

ABSTRACT

The number of older people is increasing in many countries. By 2030, it is estimated that 15% of the overall population will be comprised of people aged 65 and above. Hence, the monitoring and tracking of elder activities to ensure they live an active life has become a major research topic in recent years. In this work, an elderly sub-activity tracking system is developed to detect the sub-activity of the elderly based on their physical activities and indoor location. The physical activities tracking system and indoor location system is combined in this project to enhance the context of the elderly activities (i.e. sub-activities as defined in this project). An indoor location system is developed by using bluetooth low energy (BLE) beacon and BLE scanners to measure the received signal strength indicator (RSSI) signal to detect the location of the elderly. The activity tracking is carried out via a waist wearable device worn by the elderly. Random forest and support vector machine (SVM) are used as machine learning classifiers to predict the activity and indoor location with an accuracy of 95.03% and 86.58%, respectively. The data from activity tracking and indoor location sub-systems will then be combined to derive the sub-activity and push to an online internet of things (IoT) platform for remote monitoring and notification.

Corresponding Author:

Nour Eddin Tabbakha
Faculty of Engineering, Multimedia University
Persiaran Multimedia, 63100 Cyberjaya, Selangor, Malaysia
Email: nourfdss11@gmail.com

1. INTRODUCTION

The number of older adults is increasing at a very high rate worldwide. According to the United Nation's projection, people are now living longer and the number of the elderly is increasing in all the countries. This number represents 8.5% or 617.1 million of the total population in 2015. Moreover, the number of older people is expected to become about 1 billion in 2030 and 1.6 billion by 2050 [1, 2]. Older adults who are 65 years and above are prone to injuries from several activities such as sleeping or sitting in the wrong location or over-exercise, and occasional accidents such as minor falls. However, these injuries can be avoided with proper and regular physical activities [3]. A study conducted by the Canadian Institute for Health Information has also shown that nearly 90% of people over the age of 65 prefer to stay at home even if they need day-to-day assistance or ongoing health care [4].

This paper proposes the design and development of a sub-activity recognition smart system (SRSS) by analysing the data from activity tracking and indoor location sub-systems. A wearable device based on Linkit 7697 and MPU6050 sensor to be worn on the waist of the elderly has been developed to send the data of the accelerometer and gyroscope to the activity tracking system. The activity tracking system will help to predict the physical activity which can be divided into sitting, standing, lying, falling or walking. Indoor

location system has also been developed to detect the location of the elderly by measuring the RSSI signal of the BLE beacon from the BLE scanners. Indoor location system will help to predict the location of the elderly in a home environment such as living room, kitchen, bedroom, office or toilet.

The proposed SRSS system combines the results from both the indoor location and activity tracking to derive sub-activity such as lying on the sofa in the living room implies relaxing while lying on the bed in the bedroom implies sleeping. The activity tracking dataset has been created from 20 volunteers performing each activity namely sitting, standing, walking, falling and lying. Conversely, the indoor location dataset has been created by only one volunteer as the RSSI values measured will not be affected by the gender, size and the height of the volunteer. The data collected has gone through the feature engineering process to extract features that could help the machine learning model to learn more about the problem and increase the accuracy of the prediction. In this project, support vector machine (SVM) classifier gave the highest accuracy for indoor location and random forest classifier performed the best for activity tracking.

2. LITERATURE REVIEW

The first studies on human activity recognition (HAR) were carried out in the late 1990s [5]. Nevertheless, many new challenges have encouraged the development of new technologies in realistic conditions to improve accuracy. Some of these challenges include: (1) choosing the measurement attributes; (2) building a portable, non-starter and inexpensive system for data acquisition and (3) designing the feature extraction [6]. Evenson reviewed that the wearable that wore on the waist, hip or lower body area is more reliable and only a sensor is sufficient to observe a good accuracy [7].

Zhou used a deep convolutional neural network (CNN) to detect physical activity using smartphone sensors. The data were collected from an accelerometer and a gyroscope sensor transmitted through a Bluetooth signal to a computer where it is stored and analysed, the overall accuracy is 98% [8]. Lee collected data from five students who performed 2377 seconds of running, 3588 seconds of walking and 3934 seconds of staying still. The accelerometer data generated from the sensor was recorded at 1Hz in the smartphone, CNN was used as a classifier with an accuracy of 92.71% [9]. HAR systems can provide information about the activity performed by the elderly such as walking, sitting and standing, but this information might be insufficient to track activities such as toileting, eating, relaxing and sleeping [10]. To solve this issue, an indoor location system will be added to track the location of the physical activities done by the elderly, which will give us a better understanding of the elderly behaviour in a home environment, i.e. sitting in the toilet can be interpreted as toileting.

A broad range of services like indoor location has been developed through the proliferation of smartphones and other wireless devices over the last few years. Therefore, indoor location is the process of obtaining the position of an indoor system or consumer location. Indoor location using BLE technology can be implemented in two ways, first by placing many fixed BLE beacons around the indoor area and BLE scanner attached to the subject, so the BLE scanner carried by the subject will scan for nearby BLE beacon and then estimate the location of the subject based on the RSSI signal. The second way is by placing many fixed BLE scanners and the BLE beacon attached to the subject. The system proposed by [11] placed 15 BLE beacons distributed randomly on different types of objects such as a table, TV remote control and trash baskets. In the system done by [12], the authors detected the indoor location by placing five BLE beacon in the indoor location, and smartphone as BLE scanner carried by the subject and divided the home environment as five separate sections. The authors in [13] implemented the indoor location system by placing fixed BLE scanners in the building and BLE beacon carried by the subject, the distance estimation done by using statistical equation between the BLE scanners and BLE beacon. This approach is adopted in this project as the BLE beacon can be embedded into the wearable device easily. The indoor location can be obtained by mathematical method or by using a machine learning method. The mathematical method can be obtained by using RSSI and path-loss propagation model, this method can be used to obtain an accurate location, but it requires many reference points to reduce the noise which will be costly [14]. The machine learning method can be done by training the machine learning model with RSSI values measured in a room during the offline path. Then, the trained model can be used to predict the room location once new RSSI values have been received. The advantage of using a machine learning method is its adaptability to the new indoor environment [15].

The physical activity and indoor location systems are an essential approach to detect the activities and indoor location. But implementing one system alone is not enough to detect a specific behaviour, because the elderly needs a system that can track their activities and their location at the same time to predict their behaviour. Nienhold *et al.* [16] proposed an idea to study the daily activities and location of the elderly person by placing motion sensors in each room, and a wearable placed on arm, leg and breast for activities tracking with an accuracy rate of over 83%. The limitations of this approach were low accuracy, many

sensors and the indoor location detected by using motion sensors which were difficult to install and costly. The system proposed by [17] combined the activities and the indoor locations to track the patient's daily life, and the system is accessible by the caregiver or family members to see if any abnormal behaviour happened. Sie *et al.* [18] proposed a way to combine the result from two systems which were activity tracking and indoor location then saved them into activity log. K-nearest neighbours (K-NN) was used as a classification algorithm with an accuracy of 95%. Even though the accuracy obtained was high and the indoor location was accurate but the actions studied were limited. The authors in [19] also proposed a method to study the actions, namely sleeping, resting, dressing, undressing, eating, toileting and taking a shower. The limitation of this paper was the very small activity tracking dataset. The author in [20] developed the ActiveLife system, which tracked the daily life of patients such as lying and being in the bedroom implied that the person was sleeping and standing in the kitchen implied that the person was cooking [20]. From the literature review, it was apparent that many strategies were used in tracking activities and indoor location. Some researcher combined the outcomes of the activity tracking and indoor location systems to predict the current posture of the elderly inside the home environment. The researcher presented a good accuracy for activity tracking and indoor location but the accuracy associate with more sensors and many features. In this paper, the sub-activities have been studied by integrating the results from activity tracking and indoor location using fewer sensors and reducing the number of training features.

3. PROPOSED SYSTEM ARCHITECTURE AND WORKFLOW

As shown in Figure 1, the system consists of a wearable device that integrates an accelerometer and a gyroscope, BLE beacon and Wi-Fi module. Other components include BLE scanners, message queuing telemetry transport (MQTT) Server, Processing Machine and IoT platform. The details of the system architecture shown in Figure 1 are described as:

a. A wearable device that consists of an accelerometer and a gyroscope to collect the data of the acceleration and angular velocity of the body. It is also equipped with BLE beacon, which communicates with the BLE scanners to help the detection of the wearable device location. Finally, a Wi-Fi microchip is embedded to transmit the data to a processing machine via an MQTT broker.

b. BLE scanners based on Raspberry Pi 3 to keep looking for BLE beacon (wearable device). A BLE scanner is placed in each room to send the RSSI data to a processing machine (laptop) to be analysed for the indoor location.

c. MQTT broker that connects all devices.

d. Processing machine (laptop), which receives the data from the wearable devices and BLE scanners to apply the machine learning model to predict the type of physical activity and the indoor location.

e. IoT platform i.e. online website to present and display the result of the activity tracking and indoor location.
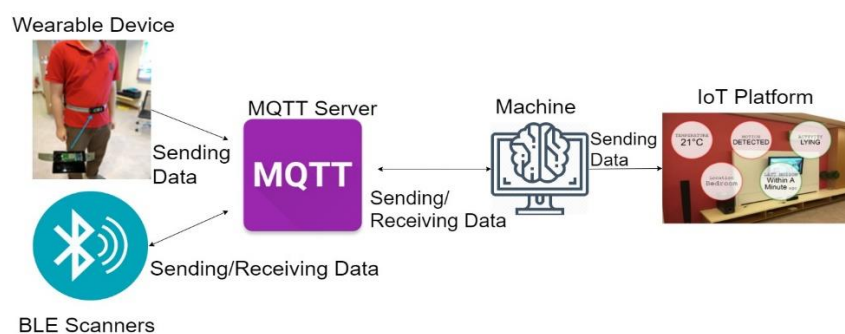

Figure 1. The system architecture

The flow chart presented in Figure 2 explains the process of the machine learning-based system modelling proposed in this paper. It starts with raw data collected from the wearable device for the physical activities and from the BLE scanners for Indoor location. These two datasets will be used to train the machine learning classifiers to classify the type of activity and the indoor location. Next, features engineering is applied to transform the raw data collected into useful features to interpret the problem in this work. Then, the datasets created will be split into two parts as training dataset (75%) and testing dataset (25%). The training dataset will be fitted into a machine learning model with hyperparameters optimisation to test which

hyperparameters work the best. Later, the model will be evaluated using the testing part of the dataset to measure accuracy. The resulting machine learning model will be implemented in the final system for prediction. When receiving new data, the features will be calculated on-line, and the prediction will take place in real-time.
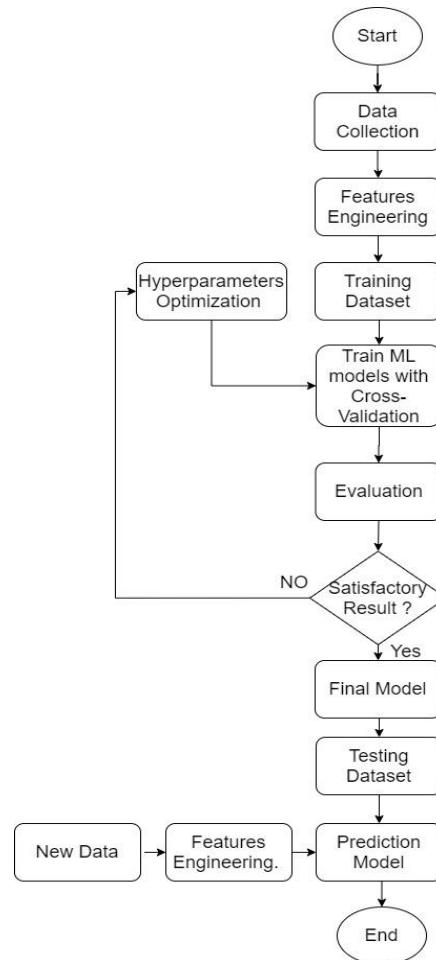

Figure 2. ML-based system modelling flow chart

3.1. Creating the dataset

This section describes the process of creating two datasets from the wearable device and BLE scanners for physical activity tracking and indoor location, respectively. These datasets are used to train the machine learning models, which will be described in the subsequent sections. The activity tracking and indoor location datasets were collected from 20 volunteers (10 males and 10 females) at the digital home and lifestyle center (DHALC) lab. The process of collecting the activities dataset consists of:

a. Linkit7697 reads the data of the accelerometer and gyroscope from the MPU6050 at a sampling rate of 20 Hz, Kalman filter is used to filter the noise from the MPU6050.
b. The data of accelerometer and gyroscope are sent wirelessly to the processing machine via MQTT broker.
c. A program running on the processing machine subscribed to the activity topic on the MQTT broker records these data and saves it into a CSV file. Table 1 shows an example of the raw data stored in the CSV, which contains: time, accelerometer X-axis, Y-axis, Z-axis, and Gyroscope X-axis, Y-axis, Z-axis, and the name of the activity performed by the volunteer.
d. The data was collected from 20 volunteers performing five activities namely "walking, standing, sitting, falling and lying". The volunteers were asked to perform each activity in many ways and behave like the elderly. Then, all the files are saved in one single dataset named as Activity Tracking Raw Data for further analysis.

Table 1. Activity tracking raw data

Time	Accelerometer Data	Gyroscope Data	Activity
5:23:11	X1, Y1, Z1	X1, Y1, Z1	Sitting
5:24:11	X2, Y2, Z2	X2, Y2, Z2	Standing
5:25:11	X3, Y3, Z3	X3, Y3, Z3	Walking
5:26:11	X4, Y4, Z4	X4, Y4, Z4	Lying
5:27:11	X5, Y5, Z5	X5, Y5, Z5	Falling

The indoor locations dataset is recorded in the DHALC Lab. DHALC consists of 5 rooms namely living room (10m×5m), kitchen (4.5m×4.5m), bedroom (5m×4.5m), Office (6m×4m), Foyer (4m×3.8m), and Toilet (1.6m ×0.95m). The red crosses in the floor plan of DHALC shown in Figure 3 indicate the placements of the BLE scanners.
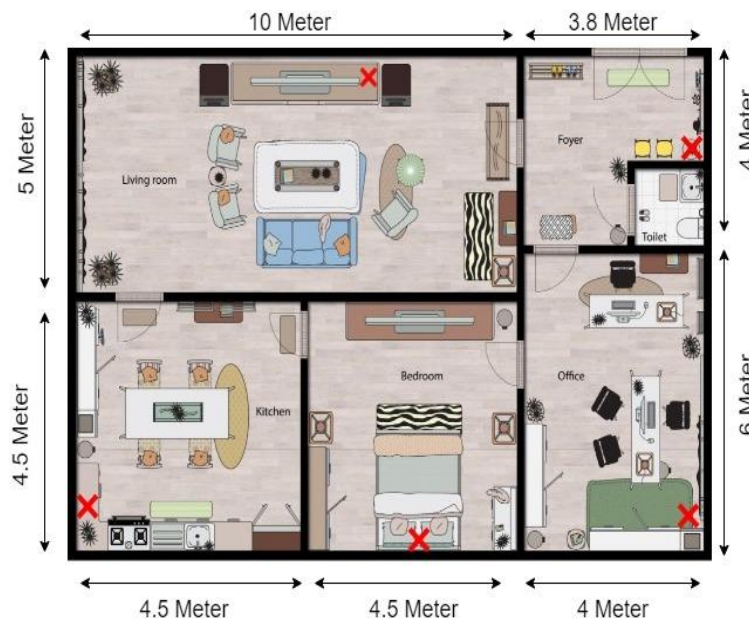

Figure 3. The placements of the BLE scanners are indicated by the red crosses

The process of collecting the indoor locations dataset consists of:
a. The BLE scanners in kitchen, bedroom, office and toilet are configured and a threshold is set to avoid any overrun from any BLE scanner because the sizes of the rooms are small. The threshold is obtained by calculating the average of 50 RSSI values in each room.
b. The data collected from one volunteer wearing the BLE beacon (on a wearable device) walks around each room for five hours.
c. The machine will run a Python script that sends a scanning request every 5 seconds through the MQTT broker to all the BLE scanners to scan for a BLE beacon at the same time. BLE scanners will reply to the topic with its scanner ID (the name of the room where the scanner exists) and the current RSSI value of the BLE beacon. The machine will record the data in a CSV file, as shown in Table 2. If an RSSI is equal to -120, this means that the BLE beacon is very far from the BLE scanner or not found. All the files are saved in one single dataset name as Indoor Location Raw Data for further analysis.

Table 2. Indoor location raw data

Living Room	Bed Room	Kitchen	Office	Toilet	Location
-80	-120	-120	-65	-50	Toilet
-45	-70	-80	-120	-55	Living
-65	-80	-55	-120	-120	Kitchen
-90	-65	-85	-95	-120	Bedroom
-120	-95	-120	-70	-75	Office

3.2. Feature engineering

Feature engineering is essential in analysing datasets of both activity tracking and indoor location. Raw data acquired by sensors are not enough to get high accuracy in machine learning stages. Feature engineering is the process of using the cleaned dataset to extract features that help machine learning models to learn more about the problem and increase the accuracy of the prediction. This process includes feature extraction, feature normalisation, and feature selection.

3.2.1. Feature extraction

Feature extraction is a way of generating and extracting more data from the raw data, and it is different from one application to another [21]. Feature extraction could be a mathematical expression applied to raw data such as mean (average), standard division, min, max, median, variance and energy. The activity tracking dataset has six columns (three accelerometer measures X, Y and Z-axis and three gyroscope measures X, Y, and Z-axis), while the indoor location dataset has five measures of RSSI from five places (bedroom, living room, kitchen, toilet, and office). In this paper, standard division, mean/average, maximum, minimum values and median are used as the mathematical expressions to extract new features from the raw data. The values of all these parameters were determined with an overlapping sliding window of size 5 with a step size of 1. After the feature extraction is completed, new datasets for activity tracking and indoor location are created and named as activity tracking dataset (ATD) and indoor location dataset (ILD), respectively.

3.2.2. Feature normalisation

Feature normalisation is used to normalise the features into one structure, which will make it easy for machine learning to learn about the problem. Min-Max and Z-Score normalisation (Standardisation) is the most common algorithms in data normalisation [22]. Z-score normalisation will be used to normalise the data because the data has a normal distribution and Z-score is much less affected by outliers [23]. Z-score first subtracts the features with mean value and then divides them by standard deviation so that the resulting distribution has a unit variance.

3.2.3. Feature selection

Feature selection is a process to choose which features have more influence on the machine learning model accuracy and to reduce the dimensions of the features by removing nonrelated features without losing any vital information. Reducing the number of features will also help in reducing the time of training, memory and overfitting. In this paper, two methods of feature selection are used, namely feature correlation and feature importance.

Feature correlation measures the relationship between the input features and output classes and how strongly they are related to each other. The correlation coefficient ranges from -1 to 1. If the correlation shows a strong relationship between two features means that one of these features can be removed to reduce the number of features.

Feature importance is used as the second method for features selection in this system. Features selection is a method to select the most important features based on random forest classifier to reduce the number of the features by removing all the features with low importance to the prediction output [24]. Features importance provide a score that indicates how important or useful each feature to the prediction output by using a random forest classifier. After features selection is applied on the dataset, all the changes are stored into a new dataset named as activity tracking dataset after correlation and features importance (ATD-C-FI) and indoor location after correlation and features importance (ILD-C-FI).

3.3. Machine learning modelling

In this work, supervised machine learning is used since the datasets for activity tracking and indoor locations are collected and labelled. 75% of the dataset is used for training, and the remaining 25% is used for testing. Different classification algorithms have been evaluated, namely, logistic regression, random forest, decision trees, K-NN, SVM, and stochastic gradient descent (SGD). The selection of machine learning algorithms is done by selecting the most used algorithms in the literature review. To avoid overfitting and underfitting problems, a 10-fold cross-validation technique is used in this system. The cross-validation technique trains and tests the models 10-times on a different subset from the training dataset. Besides, tuning the hyperparameters of the ML models and checking the performance of the machine learning models are also done to find the best set of values which give the best accuracy for each classifier [25]. Grid search is used for hyperparameters optimisation, which is a very efficient way of testing different hyperparameters and finding the optimal hyperparameters of the machine learning model [26]. Grid search allows us to define many hyperparameters and automatically test all the hyperparameters by using a different combination of hyperparameters and report the best score. The selected machine learning model is implemented and tested

again by asking 10 new volunteers to perform each activity for 20 mins and record the experimental results to compare the accuracy between the simulation results and experimental results. All machine learning classifiers are implemented and tested using the Scikit-Learn package [27].

3.4. Sub-activity detection

An algorithm is developed to integrate the results of activities tracking and indoor location to derive the sub-activities. The derived sub-activities could be different due to the location, for example lying in the bedroom may imply sleeping and lying in the living room may imply relaxing, but lying in the toilet may indicate an emergency. Figure 4 shows the flow chart of the action recognition system, the results from the activity tracking and indoor location system are sent to the data analysis algorithm to derive the sub-activity. Subsequently, the concluded sub-activity is published to the IoT platform for remote monitoring and notification. The sub-activities for the elderly play a vital role to ensure that the elderly continue to live an active life. As mentioned by [28] active living means the elderly being active by doing some indoor activity such as walking, cooking, dressing and bathing, instead of sitting idle which will have a huge impact on their balance, flexibility coordination, endurance, mental health, cognitive function and muscle strength. The American College of Sports Medicine (ACSM) and the American Heart Association (AHA) recommends the elderly to have a mix of moderate and vigorous aerobic activity every week. The ACSM and AHA also suggest carrying out strength exercises regularly to work out all the major muscles such as legs, hips, back, abdomen, chest, shoulders and arms. In view of this, the detection of sub-activities is able to avoid long hours of sitting or idling by alerting the elderly to do some light activities when they have been sitting for too long [29]. The detected sub-activity will also be recorded and reported to a caregiver for them to analyse their routine actions.
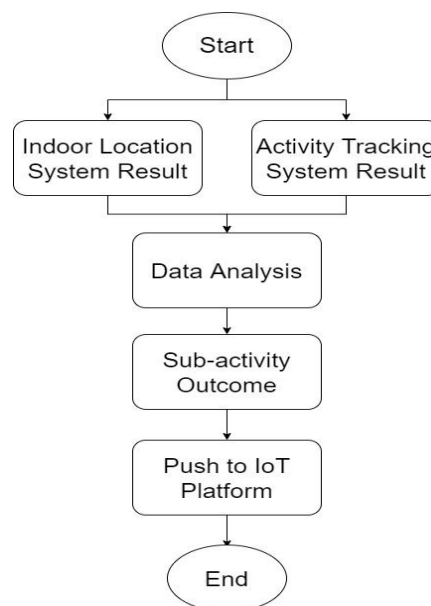

Figure 4. Flow chart of the sub-activity recognition system

4.   RESULTS AND DISCUSSION

The raw data from the activity dataset contains seven columns that represent the sensory data from the 3D-accelerometer and the 3D-gyroscope collected from the waist wearable device and labels of the activity performed. The dataset of the indoor location contains six columns that represent the RSSI reading from Bedroom, living room, kitchen, office, toilet BLE scanners and label of the room where the BLE beacon is located. Table 3 summarises the dimension of the raw data collected for activity tracking and indoor location sub-system.

Table 3. The dimension of the raw data collected

Dataset name	Row	Columns
Activity Tracking	2880000	7
Indoor Location	90000	6

4.1. Feature engineering

Feature extraction is implemented on the raw data to extract min(min), max(max), mean/average(avg), median(med) and standard deviation(std). The total number of features for the activity tracking dataset are 37 columns, including accelerometer (3 features) and Gyroscope (3 features) readings, the statistical features (30 features), and the label (1 column). The total features for Indoor location datasets are 31 columns, including the reading of the RSSI signal from each room (5 features), the statistical features for each of these readings (25 columns), and the label of the room (1 column). Table 4. 3 summarises the dimensions of the features after feature extraction.

Table 4. The dimension of the features after feature engineering

Dataset name	Rows	Columns
ATD	2880000	37
ILD	90000	31

4.2. Feature selection

Feature selection checks all the features in the datasets to see which features contribute the most in prediction output and to remove all the irrelevant or redundant features in the dataset to improve the accuracy and the performance of the machine learning models. The first feature selection technique implemented in this project is Features Correlation, which measures the linear relationship between two features. Based on Pearson's correlation coefficient, high degree correlation lies between ±0.5 and ±1. If the correlation coefficient for a pair of features lies in between ±0.5 and ±1, then one of the features can be eliminated [30]. By doing this, some features will be removed and thus will reduce the total number of features.

After implementing correlation coefficient between the features in the activity dataset, the features extracted from avg, med, min and max are highly correlated with ax, ay, az, gx, gy and gz, so keeping one of them is enough. In this case, the average is kept while the rest of the features are removed. Secondly, the features extracted from standard deviation are highly correlated with each other. Thus, by keeping one feature is enough where std-ax is kept and the rest are removed. Figure 5 shows the correlation matrix after removing the highly correlated features. The colours in Figure 5 range from red (strong positive correlation=+1) to blue (strong negative correlation=-1).

After implementing the correlation between all the features in the indoor location dataset. It shows that the features extracted from avg, med, min and max are highly correlated for bedroom, kitchen, living, office and toilet. So, the features med, min and max are removed and the avg features is kept. Figure 6 shows only the important features after removing the highly correlated features and the correlation coefficient are recalculated. Table 5 shows the final dimension of the features after feature correlation implemented on activity tracking and indoor location dataset.
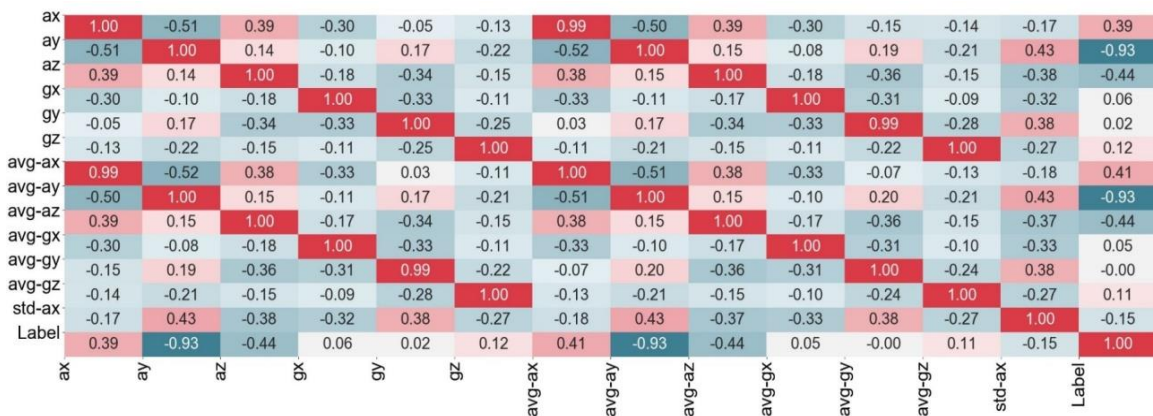

Figure 5. Correlation matrix after removing the correlated features in ATD

Table 5. The dimension of the features after features correlation

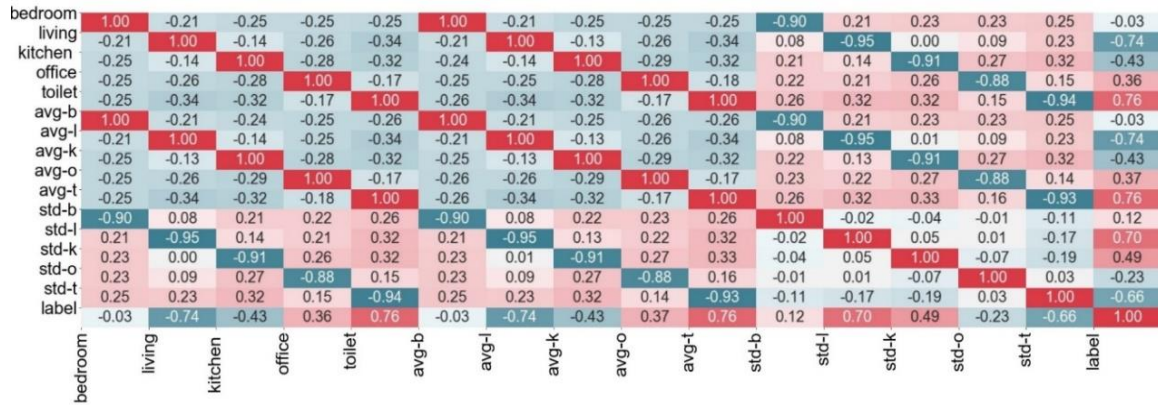| Dataset name	Row	Columns
ATD-C	2880000	14
ILD-C	90000	16

Figure 6. Correlation matrix after removing the correlated features in ILD

Next, feature importance is applied on the dataset created after features correlation is done. In this process, the features selection is selected based on a threshold value and in this work, the threshold is set to 0.01. All the features with an important score below 0.01 will be removed. Figure 7(a) shows the importance of the features implemented on the ATD-C, the features; gx, gz, avg-gx and avg-gz with a low score (less than 0.01) are removed. Figure 7(a) also illustrates that the gyroscope data are less important than the accelerometer data because the activity detection does not depend much on the angular velocity. Figure 7(b) shows the importance of the features implemented on the ILD-C, the features namely living and std-o with a low score (less than 0.01) are removed.

Table 6 shows the final dimension of the features after implementing the feature importance on activity tracking and indoor location dataset. The total number of columns reduced from 68 (original) to 30 (after correlation coefficient) and further reduced to 24 (after feature importance).
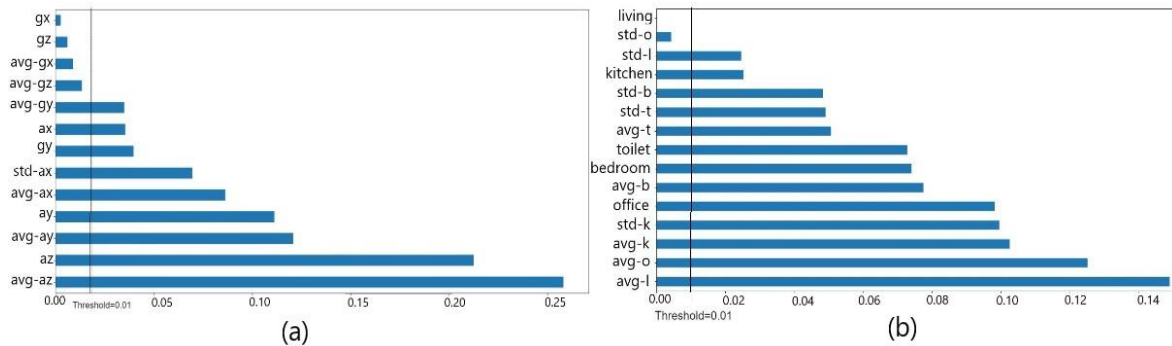

Figure 7. Features importance, (a) For ATD-C, (b) For ILD-C

Table 6. The dimension of the features after features importance

Dataset name	Row	Columns
ATD-C-FI	2880000	10
ILD-C-FI	90000	14

4.3. Classification result

As a result of the feature engineering stage, there are eight datasets, four for activity tracking dataset namely Activity Tracking Raw Data, ATD, ATD-C, ATD-C-FI and four for indoor location namely Indoor Location Raw Data, ILD, ILD-C, ILD-C-FI. These datasets will be used for classification to check the performance for each machine learning techniques. Table 7 compares the result of each classifier trained by using the different activity tracking dataset. The accuracy of logistic regression and SGD models improved when more features extracted from the raw data. Originally, logistic regression gives an accuracy of 64.90% when trained and tested on the raw data dataset. But when the logistic regression is trained and tested on

ATD, the accuracy improved to 87.32%. SGD classifier shows that the accuracy gives 68.30% when trained and tested with raw data dataset but when the model is trained and tested again with ATD, the accuracy improved to 89.11%. This shows that the number of the features improve the accuracy of linear classifiers but not the number of the samples. The accuracy results of Random forest, Decision Trees and K-NN are similar and showing not much improvement even with more features extracted, which shows that the number of samples improves the accuracy more than the number of features. This experiment helps to choose the dataset which gives good accuracy with fewer features to reduce the processing time. So ATD-C-FI dataset gives a promising result with fewer features and it is chosen to train and test the machine learning classifiers with new volunteers (unseen data). Table 8 shows the evaluation results on indoor location datasets by the various classifiers. As shown in Table 8, the result shows good accuracy for all the classifiers by using a different dataset. So, indoor location raw data dataset will be chosen to be trained and tested again with a new volunteer to check which classifier will give a good result.

Table 7. Evaluation results on activity tracking datasets

Dataset name	Logistic Regression	Random Forest	Decision Trees	K-NN	SGD
Raw Data	64.90%	99.79%	99.62%	99.85%	68.30%
ATD	87.32%	99.85%	99.64%	99.24%	89.11%
ATD-C	86.28%	99.85%	99.68%	99.73%	84.25%
ATD-C-FI	85.05%	99.83%	99.68%	99.72%	81.83%

Table 8. Evaluation results on indoor location datasets

Dataset name	Logistic Regression	Random Forest	Decision Trees	K-NN	SVM
Raw Data	99.72%	99.80%	99.76%	99.76%	99.45%
ATD	99.89%	99.92%	99.76%	99.72%	99.96%
ATD-C	99.87%	99.84%	99.72%	99.92%	99.76%
ATD-C-FI	99.61%	99.96%	99.84%	99.88%	99.76%

Table 9 shows the experimental results for activity tracking system trained with ATD-C-FI dataset and tested with new volunteers. Random Forest gives the best performance among other classifiers, which is chosen for the final system implementation. Table 10 shows the experiment result for indoor location system trained with indoor location dataset and tested with new data. SVM gives the best result among other classifiers which will be chosen for the final system implementation.

Table 9. Experimental result for activity tracking

ML Classifier	Precision [%]	Recall [%]	F1 Score [%]	Accuracy [%]
Random Forest	95.00	94.87	94.31	95.03
Decision Tree	90.63	86.23	85.50	86.23
SGD	87.43	84.51	83.73	84.51
Logistic Regression	90.01	85.94	85.13	85.94
K-NN	93.20	91.30	91.14	91.30

Table 10. Experimental result for indoor location

ML Classifier	Precision [%]	Recall [%]	F1 Score [%]	Accuracy [%]
Random Forest	84.45	83.52	82.71	83.72
Decision Tree	85.31	84.65	84.41	84.84
SVM	87.51	85.91	85.72	86.58
Logistic Regression	86.14	83.22	82.42	85.66
K-NN	88.11	85.14	84.71	79.00

Indeed, combining the indoor location and activity tracking systems would give a better understanding of the elderly sub-activity. Although in most cases, the results from activity tracking are sufficient, but in some cases, the same activity at different locations of the home might have a different interpretation. As shown in Table 11, standing in the living room and office area may imply "standing", whereas standing in the kitchen may imply "cooking" and standing in the bedroom may imply "changing cloth".

Table 11. Sub-activity outcome

Activity	Indoor Location	Sub-activity
Sitting	Living room	Relaxing
Lying	Living room	Relaxing
Walking	Living room	Moving around
Falling	Living room	Emergency
Standing	Living room	Standing
Standing	Kitchen	Cooking
Sitting	Kitchen	Eating
Walking	Kitchen	Moving around
Lying	Kitchen	Emergency
Falling	Kitchen	Emergency
Lying	Bedroom	Sleeping
Standing	Bedroom	Changing cloth
Sitting	Bedroom	Relaxing
Walking	Bedroom	Moving around
Falling	Bedroom	Emergency
Sitting	Office	Working
Lying	Office	Emergency
Walking	Office	Moving around
Falling	Office	Emergency
Standing	Office	Standing
Sitting	Toilet	Defecating
Lying	Toilet	Emergency
Walking	Toilet	Moving around
Falling	Toilet	Emergency
Standing	Toilet	Standing

5. CONCLUSION

This research work focused on the design and development of a sub-activity recognition system for the elderly by combining activity tracking and indoor location. The proposed work is representing the main structure that provides a real-time platform for the sub-activity recognition system for older adults by using a wearable device and BLE scanners. The validation accuracy rates obtained using the proposed technique are 86.58% and 95.03% for indoor location tracking and activity recognition, respectively. The system still needs more work on how to use the results to detect any abnormal behaviour for the elderly. The system can also be further improved by analysing the data to avoid any future symptoms such as high blood pressure, help to manage weight and reduce the risk of heart disease due to lack of physical activity.

REFERENCES

[1] Samsudin, S. Abdullah, N. Ismail, N. S. A, Hassan, K., & Ahmad U. A, "Growing old at own home: A study among pre-elderly adults in Peninsular Malaysia," *Economics & Sociology*, vol. 12, no. 1, pp. 115-124, 2019.
[2] Wan He, Daniel Goodkind, and Paul Kowa, "An aging world: 2015 International Population Reports," *United States Cencus Bureau*, 2016.
[3] UN DESA, "World Population Prospects 2019: Highlights," *UN Department of Economic and Social Affairs*, 2019.
[4] "Health care in Canada, 2011: A focus on seniors and aging," *Canadian Institute for Health Information,* 2011.
[5] F. Foerster, M. Smeja, J. Fahrenberg, "Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring," *Computers in Human Behavior*, vol 15, no. 5, pp. 571-583, Sep 1999.
[6] E. Kim, S. Helal and D. Cook, "Human Activity Recognition and Pattern Discovery," in *IEEE Pervasive Computing*, vol. 9, no. 1, pp. 48-53, Jan.-March 2010.
[7] Kelly R. Evenson, Michelle M. Goto and Robert D. Furberg, "Systematic review of the validity and reliability of consumer-wearable activity trackers,". *International Journal of Behavioral Nutrition and Physical Activity*, vol. 12, no. 159, pp. 1-22, 2015.
[8] Baoding Zhou, Jun Yang, Qingquan Li, "Smartphone-based activity recognition for indoor localization using a convolutional neural network," *Sensors*, vol. 19, pp. 1-15, 2019.
[9] Song-Mi Lee, Sang Min Yoon and Heeryon Cho, "Human activity recognition from accelerometer data using Convolutional Neural Network," *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)*, Jeju, pp. 131-134, 2017.
[10] Michelle E. Mlinac, Michelle C. Feng, "Assessment of activities of daily living, self-care, and independence," *Archives of Clinical Neuropsychology*, vol. 31, no. 6, pp. 506-516, 2016.

[11] A. R. Jiménez, F. Seco, P. Peltola and M. Espinilla, "Location of Persons Using Binary Sensors and BLE Beacons for Ambient Assitive Living," *2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Nantes*, pp. 206-212, 2018.

[12] N. Duong and T. Dinh, "Indoor Localization with lightweight RSS Fingerprint using BLE iBeacon on iOS platform," *2019 19th International Symposium on Communications and Information Technologies (ISCIT)*, Ho Chi Minh City, Vietnam, pp. 91-95, 2019.

[13] Kenta Urano, Katsuhiko Kaji, Kei Hiroi, Nobuo Kawaguch, "A location estimation method using mobile BLE tags with tandem scanners," In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, pp. 577-586, 2017.

[14] F. Seco, A. R. Jimenez, C. Prieto, J. Roa and K. Koutsou, "A survey of mathematical methods for indoor localization," *2009 IEEE International Symposium on Intelligent Signal Processing, Budapest*, pp. 9-14, 2009.

[15] F. Zafari, A. Gkelias and K. K. Leung, "A Survey of Indoor Localization Systems and Technologies," in *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2568-2599, thirdquarter 2019.

[16] D. Nienhold, R. Dornberger and S. Korkut, "Sensor-Based Tracking and Big Data Processing of Patient Activities in Ambient Assisted Living," *2016 IEEE International Conference on Healthcare Informatics (ICHI)*, Chicago, IL, pp. 473-482, 2016.

[17] W. Zhao, T. Qiu and X. Luo, "Automatic User Authentication for Privacy-Aware Human Activity Tracking Using Bluetooth Beacons," *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Miyazaki, Japan,* pp. 3609-3613, 2018.

[18] J. Sie, S. Yang, Z. Hong, C. Liu, J. Chen and S. C. Li, "Integrating Cloud Computing, Internet-of-Things (IoT), and Community to Support Long-Term Care and Lost Elderly Searching," *2016 International Computer Symposium (ICS)*, Chiayi, pp. 452-457, 2016.

[19] A. Fleury, M. Vacher and N. Noury, "SVM-Based Multimodal Classification of Activities of Daily Living in Health Smart Homes: Sensors, Algorithms, and First Experimental Results," in *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 274-283, March 2010.

[20] N. W. Tsang, U. M. Qureshi, K. Lam, J. K. Ng, S. Han and I. Papavasileiou, "Tracking Indoor Activities of Patients with Mild Cognitive Impairment Using Motion Sensors," *2017 IEEE 31st International Conference on Advanced Information Networking and Applications (AINA)*, Taipei, pp. 431-438, 2017.

[21] S. J. Preece, J. Y. Goulermas, L. P. J. Kenney and D. Howard, "A Comparison of Feature Extraction Methods for the Classification of Dynamic Activities From Accelerometer Data," in *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 3, pp. 871-879, March 2009.

[22] S. Gopal Krishna Patro, Kishore Kumar Sahu, "Normalization: A preprocessing stage," *arXiv preprint arXiv:1503.06462*, 2015.

[23] Al Shalabi, L., Shaaban Z., & Kasasbeh B, "Data mining: A preprocessing engine," *Journal of Computer Science*, vol. 2, no. 9, pp. 735-739, 2006.

[24] Menze B H, Kelm B M, Masuch R, Himmelreich U, Bachert P, Petrich W, & Hamprecht F. A, "A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data," *BMC bioinformatics*, vol. 10, no. 213, 2009.

[25] Moradi B, Aghapour M, & Shirbandi A, "Compare of Machine Learning And Deep Learning Approaches for Human Activity Recognition" *EasyChair*, no. 1945, 2019.

[26] La Valle, S. M. Branicky, & Lindemann, S. R, "On the relationship between classical grid search and probabilistic roadmaps," *The International Journal of Robotics Research*, vol. 23, no. 7-8, pp. 673-692, 2004.

[27] Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, "Scikit-learn: Machine learning in Python," *Journal of machine Learning research*, vol. 12, pp. 2825-2830, 2011.

[28] Collette M, Godin G, Bradet R, Gionet NJ, "Active living in communities: understanding the intention to take up physical activity as an everyday way of life," *Canadian journal of public health= Revue canadienne de sante publique*, vol. 85, no. 6, pp. 418-421, 1994.

[29] Wangui Anthony Muchiri, Oloo Micky Olutende, Issah W. Kweyu, Esther Vurigwa, "Meaning of Physical Activities for the Elderly: A Review," *American Journal of Sports Science and Medicine*, vol. 6, no. 3, pp. 79-83, 2018.

[30] Sedgwick, "Pearson's correlation coefficient," *Bmj*, no. 345, 2012.

BIOGRAPHIES OF AUTHORS

Nour Eddin Tabbakha was born in Damascus, Syria, in 1994. He received a bachelor's degree in Computer Engineering from Multimedia University, Malaysia. He is preparing his master thesis on the indoor positioning and activity tracking for assisted living. His research interests include machine learning, deep learning, data science, embedded systems and Internet of Things (IoT).

Dr. Ooi Chee Pun obtained his Ph.D. in Electrical Engineering from University of Malaya, Malaysia, in year 2010. Dr. Ooi current position is a senior lecturer with the Faculty of Engineering in Multimedia University. He is a chartered engineer registered with Engineering Council (the British regulatory body for Engineers), member of IET. His research areas are in FPGA based embedded systems, embedded systems

Dr. Tan Wooi Haw received his M.Sc. in Electronics from Queen's University of Belfast, UK and a Ph.D. in Engineering from Multimedia University. He is currently a senior lecturer at Multimedia University. Dr. Tan's areas of expertise include image processing, embedded system design, Internet of Things (IoT), machine learning and deep learning.

Dr. Tan Yi Fei obtained her B.Sc(Hons), M.Sc and PhD from University of Malaya(UM). She is currently a lecturer at the Faculty of Engineering in Multimedia University(MMU), Cyberjaya, Malaysia. Her research interests include machine learning, image processing, big data analytic and queueing theory.