ISSN 2409-2665
Journal of Logistics, Informatics and Service Science
Vol. 9 (2022) No.4, pp.129-148
DOI:10.33168/LISS.2022.0410

Activities of Daily Living Recognition Using Deep Learning Approaches

Lim Chin Hong, Connie Tee, Michael Kah Ong Goh

Faculty of Information Science & Technology, Multimedia University, Melaka, Malaysia

michael.goh@mmu.edu.my (Corresponding Author)

Abstract. Alzheimer's disease has become a prevalent disease faced by the elderly in Malaysia. Studies believe that early symptoms of the disease can be detected via activities of daily living. Activities of daily living is a term that collectively refer to the basic or fundamental activities performed independently to care for oneself. In this paper, a deep learning approach is presented for activities of daily living recognition to classify daily life activities such as drinking from the cup, eating at a table, reading the book, using the telephone, and walking. A number of long short-term memory (LSTM) variants have been tested in this study. Experiments results demonstrate a promising accuracy of 94% can be achieved using the public Toyota Smarthome dataset.

Keywords: Deep Learning, LSTM, Toyota Smarthome Dataset, Classification, Daily life activities

1. Introduction

Ageing population is one of the critical problems faced by the community worldwide. Alzheimer's disease is considered the most common type of dementia that affects more than 50 million people globally. According to World Health Organization (WHO), there are almost 10 million new Alzheimer's cases every year. The most significant known risk factor that causes Alzheimer's is due to ageing. People who face Alzheimer's are commonly people who are 65 and older. Alzheimer's will worsen over time, which will cause the individuals to face more severe memory loss, behaviour changes, and difficulty carrying activities of daily living (ADL).

In addition, more and more seniors prefer to stay alone when their children work in a separate cities and do not live with them. The symptoms for the dementia disease appear gradually. It takes years for a patient to discover the disease. Very often, the symptoms are mistakenly treated as signs of old age. As a result, the disease will only be discovered when it reaches a serious stage, which is very hard to cure by then. The seniors who stay alone, especially, are vulnerable to this disease as it is hard for them to realise the gradual changes in their daily activities. Sometimes, they just simply ignore the symptoms as they do not want to face such disease and deny the facts.

What is an activity of daily life? Cooking, reading, and sweeping the floor are examples of ADL, which refer to the fundamental skills required to take care of oneself independently. Studies have shown that early symptoms of dementia and Alzheimer's disease can be detected through ADL. The method is proposed by Patel and Shah (2019). This is because the daily movement of a person is controlled by the brain. If there are some degenerative functions in the brain, the movement and coordination of a person will be affected. Therefore, monitoring of the ADL could serve as an early indicator for potential cognitive diseases. Early preventive measure can then be taken to slow if not cure the diseases.

The focus of the study is to classify basic ADLs of the elderly by using a deep learning approach. Common ADL like drinking water from the cup, eating at a table, reading books, using the telephone, and walking are included. Videos from CCTV cameras are used as the input to the system. A human pose estimation method is applied to extract skeletal data from the human joints to represent the pose of the subject. After that, the sequential pose data are fed to the LSTM architectures to perform classification. A number of LSTM variants have been explored in this research. Experimental results show that an accuracy of up to 94% could be achieved by the proposed approach.

2. Related Works

In the earlier days, research focused on using conventional machine learning

approaches such as Support Vector Machines (SVM) and Logistic Regression (LR) to perform ADL recognition. However, these methods usually require extensive empirical trials to determine the best combinations of parameters/settings to achieve a good recognition performance. With the advent of deep learning technology, human activity recognition has advanced significantly. Unlike the conventional methods, deep networks can perform classification tasks in an end-to-end manner. Moreover, deep learning methods usually simplify the experimental process, by combining two or three steps like feature extraction and classification in ADL recognition.

In 2019, supervised algorithms for detecting human activities in an Ambient Assisted Living Environment were investigated. The method is proposed by Patel and Shah (2019). Activities like standing, sitting, and walking were included in a dataset known as the UCI HAR dataset. Nine classifiers were tested, including both machine learning and deep learning algorithms. Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), K-Nearest Neighbours (KNN), Support Vector Machine (SVM), and Gradient Boosting were among the machine learning algorithms used, while deep learning algorithms included Artificial Neural Network (ANN), Recurrent Neural Network (RNN), and Long-Short Term Memory Network (LSTM). Besides, data pre-processing such as normalization, missing data identification, and other cleansing processes were also performed to enhance the training process. In the study, LSTM yielded the best classification accuracy of 0.92, followed by 0.86 for RNN and 0.78 for ANN for deep learning approaches. On the other hand, logistic regression reported the highest classification accuracy for conventional machine learning algorithms at 0.94, whereas the accuracies were 0.85, 0.92, 0.83, 0.90, and 0.92 for Decision Tree, Random Forest, KNN, SVM and Gradient Boosting, respectively. The paper concluded that the LSTM network was more efficient and dependable when it came to recognising human behaviours in an ambient assisted living setting. Meanwhile, although conventional machine learning algorithm such as logistic regression could provide good result with an overall accuracy of 96 percent, it would confuse the walking upstairs and walking downstairs activities and mistaken them as walking.

In 2020, a study on IMU-based human activity recognition using deep traditional machine learning was proposed. The method is proposed by Hou (2020). Two public datasets were used in the paper which were the USC-HAD and WISDM datasets. The USC-HAD dataset contained twelve activities including walking forward, standing, sleeping, and many more. On the other hand, activities like walking, ascending stairs, descending stairs, sitting, jogging, standing were contained in the WISDM dataset. WISDM is a larger dataset as compared to USC-HAD. Three traditional machine leaning algorithms which included KNN, SVM, and Random Forests, were compared to the Deep Learning (DL) methods like Convolutional Neural Network (CNN) and LSTM. For traditional machine learning methods, Random Forests gained the best results with 82.7 percent of test accuracy in the WISDM dataset; while SVM and KNN achieved 77 percent and 67.8 percent of test accuracy. Random Forest also

reported the best result for the USC-HAD dataset with 67.9 percent of test accuracy as compared to SVM and KNN which only obtained test accuracies of 52.5 percent and 52.1 percent, respectively. The deep learning methods were also applied to the USCHAD and WISDM datasets. The CNN algorithm was tested with 1 to 3 Conv2D layer(s) in the WISDM dataset. CNN with 1 Conv2D layer gained the best result of 86 percent accuracy as compared to 83 percent and 84 percent for 1 Conv2D layer and 2 Conv2D layers, respectively. For the UCS-HAD dataset, CNN with 1 Conv2D layer gained the best result at 59.2 percent accuracy, while CNN with 2 Conv2D layers and 3 Conv2D layers obtained 45.4 percent and 36.3 percent of accuracies. LSTM was also tested with three layers just like CNN. The study showed that two LSTM layers would be the more suitable for the WISDM dataset with an 81% of accuracy as compared to 80% each for 1 LSTM layer and 2 LSTM layers. For the USC-HAD dataset, one LSTM layer obtained an accuracy of 42.9 percent as compared to 38.3percent and 24.6 percent for two LSTM layers and three LSTM layers. The study found that traditional machine learning methods were more suitable for small scale datasets while deep learning methods like CNN and LSTM are more suitable for datasets with large-scale characteristics.

A smart system for identifying daily human activities based on wrist IMU sensors was presented. The method was proposed by Ayman, Attalah and Shaban (2020). Two datasets named Handy, which included nine activities, and PAMAP2, which included 12 activities, were used in this study. Activities like cleaning window, dusting a table, eating, drinking soup and others were included in the Handy dataset. Meanwhile, activities like ironing, lying, vacuum cleaning and others were included in the PAMAP2 dataset. Following that, 5 of 9 people were asked to participate in six more optional activities, including folding laundry, driving a vehicle, playing football, watching TV, cleaning the home, and working on a computer. To segment the data from the two datasets, a one-second sliding window of size 1 sec with a 50 percent overlap between two successive frames was utilised. The paper proposed a feature extraction technique to extract time-domain features from the dataset. Three machine learning techniques namely Random Forests (RF), Bagged Decision Tree (DT), and SVM were performed. After combining data from all three sensors from the Handy dataset, the RF classifier achieved 98.87 percent accuracy, which was the highest as compared to SVM and DT. For the PAMAP2 dataset, SVM obtained the highest result of 98.1 percent, after combining the data from all three sensors. In summary, this study showed that data from sensor fusion could lead to a more promising result.

Also, in 2020, the Single Triaxial Accelerometer-Gyroscope Classification for Human Activity Recognition has been proposed. The method is proposed by Minarmo, Kusuma, Wibowo, Akbi and Jawas (2020). Standing, sitting, walking, and three more activities are included in the UCI HAR dataset. There are no feature selection and extraction methods used, which means all the features given in the public data in this paper. After that, the classifiers that were used were TML methods

such as DT, RF, Extra Trees Classifier (XT), KNN, LR, Support Vector Classification (SVC), and Ensemble Vote Classifier (ECLF). The classifiers have been applied to both model selection and testing models. Among all of these classifiers, LR produced the best results in the train datasets that were consistent with the normal distribution of each class. Regarding the testing model, LR also obtained the best accuracy, 98.40 percent, while DT, RF, XT, KNN, SVC, ECLF gained 93.44 percent, 96.73 percent, 96.68 percent, 96.21 percent, 93.86 percent, and 97.60 percent. In conclusion, LR gains the best result and is known as good with modern machine learning methods.

In 2021, the implementation of a Machine Learning Algorithm for Human Activity Recognition has been proposed. The method is proposed by Vijayvargiya, Kumari, Gupta and Kumar (2021). WISDM dataset includes six activities used in this paper. Standing, sitting, walking, and three more activities are included in this dataset. Besides, segmentation and feature extraction are also applied in this paper for data pre-processing. First, signal segmentation is used because some signals are irregular in the data. After that, feature extraction is used to improve the outcome of Machine Learning (ML) models when dealing with a large dataset. The Machine Learning model (ML) used in this paper are KNN, DT, RF, SVM) Bagging classifier, Gradient boosting classifier, and Linear Discriminant Analysis (LDA) classifier. By using 5-fold cross-validation, performance metrics for all classifications are computed for all classifiers. In conclusion, among all ML models, RF obtained the best result with a 92.71 percent accuracy and a standard deviation of 1.60 compared to 86.64 percent, 89.76 percent, 89.65 percent, 92.48 percent, 92.54percent, 78.55 percent, 89.07 percent, 92.48 percent, along with standard deviation of 1.75, 2.09, 1.53, 1.69, 1.57, 0.74, 1.29 and 1.29 for LDA, decision tree, Gradient boosting classifier, bagging classifier, KNN, SVM with linear, radial basis function (RBF), and polynomial kernel respectively. SVM with polynomial kernel gain the best result, 92.48 percent, among linear, RBF, and polynomial kernel.

Another human activity recognition method using batch normalization deep LSTM recurrent networks from inertial sensors was proposed. The method is proposed by Zebin, Sperrin, Peek and Casson (2018). Walking upstairs, sitting, lying, and three more activities were included in the dataset acquired from a waist-mounted inertial sensor. The research grouped 128 samples per sequence with six channels to become a ‘Frame’ and be utilized by the LSTM cell. Furthermore, each data row's 127 prior samples were grouped to function as a memory for the present data. The model was used in conjunction with the sequential model as well as the Dense, LSTM, Dropout, and Batch Normalization layers. The initial LSTM RNN layer, layer one, contained 30 neurons that were trained using the preceding 128 data points. At the second layer, another LSTM layer with 30 neurons was utilised to enhance the time dependence for forecasting the next value. To classify the data, a fully connected hidden dense layer of 15 neurons and a dense output layer of 6 neurons with the soft-max function were turned to a classifier to classify the data. The performance of

LSTM obtained a good result with an average precision of over 95% when classifying activities like walking level, walking up, and walking down. Because static behaviours like sitting and standing had fewer temporal correlations and repetitive components, they accounted for the majority of misclassifications. Aside from that, LSTM with Batch Normalization (BN) outperformed the generic LSTM model with a 92 percent class-wise accuracy. The LSTM with BN also performed faster than the generic LSTM, which only needed 20 epochs rather than 80 epochs to achieve a 98 percent training accuracy. Overall, the model achieved a 92 percent average accuracy for the six daily-life activities when utilising a raw accelerometer and gyroscope as input.

A deep neural networks for activity recognition using multisensor data in a smart home was presented. The method is proposed by Park, Jang and Yang (2018). The MIT dataset which contained 295 activities including toileting, preparing dinner and doing laundry was used in the paper. The Residual LSTM/ Gate Recurrent Unit (GRU) was compared with an artificial neural network and an LSTM/ GRU model. The Residual-RNN structure was built using two bi-directional hidden layers, each with 256 nodes. After that, 0.01 learning rate and batch size as 3 were applied in the two hidden layers in the model. A 0.5 drop rate was applied for better efficient computation to the fully connected layer. The Residual-LSTM/GRU obtained the best result as compared to the ANN and LSTM/GRU models. With a 90 percent ratio of the training set, the Residual-LSTM/GRU model obtained the highest accuracy at 90.85 percent and 89.52 percent as compared to 71.21 percent accuracy for ANN, 71.20 percent accuracy for LSTM, and 71.17 percent accuracy for GRU. The Residual LSTM model likewise outperformed the Residual-GRU model in terms of error loss rate, but the Residual-GRU model had a faster performance speed since LSTM contained one more gate. In summary, the Residual-LSTM/GRU proposed in the paper has achieved the highest accuracy as compared to the other models because the Residual-LSTM/GRU utilised a short-cut path during the process.

On the other hand, a human activity recognition method using Multi-Head CNN followed by LSTM was proposed. The method is proposed by Ahmad, Kazmi and Ali (2019). The UCI dataset which contained 30 people from 19 to 48 years old was used in the study. Walking upstairs, sitting, standing, and three more activities were included in the dataset. Standardisation and normalisation were applied as the pre-processing techniques, with the normalisation approach being used to reduce the dataset to zero mean and unit variance. The study used a multi-head CNN architecture followed by an LSTM model to recognise human physical activities. In the study, two models were tested: one was single-head CNN, and the other was multi-head CNN. The first model was composed of a single CNN, an LSTM layer with 128 units, and a fully connected layer with two Dense layers, one with 1000 units and one with 6 units. The second model was composed of three CNN architectures that were linked in parallel, one for total acceleration, one for body acceleration, and one for body

gyroscope. Following that, the multi-head CNN output were combined and utilised as the input to the LSTM layer, which was followed by two Dense layers, one with 1000 units and one with six units. In general, a single CNN-LSTM reported an accuracy of 94.1 percent accuracy, while the multi-head CNN-LSTM obtained a result of 95.76 percent accuracy.

In 2020, an LSTM-CNN architecture for human activity recognition was proposed. The methos is proposed by Xia, Huang and Wang (2019). Three sets of datasets that are used in this paper are UCIHAR, WISDM, and OPPORTUNITY. Standing, lying, walking, and three more activities are included in the UCI-HAR, and this dataset has 30 participants ranging in age from 19 to 48 years old. After that, upstairs, downstairs, jogging and three more activities are included in the WISDM datasets, and this dataset has the largest samples among all datasets used in this paper. Finally, open door, close door, clean table and 14 more activities are included in the OPPORTUNITY dataset. This paper performed three data pre-processing steps: linear interpolation that can fill the missing values of the datasets, scaling and normalization that can normalize the output to 0 and 1, and segmentation that can separate the data in order to get additional samples. The objective of data pre-processing is to provide a certain dimension of data to the proposed network in order to improve the model's accuracy. This paper used an LSTMCNN model, which contains eight layers. The pre-processed data will first feed into two layers of LSTM with 32 neurons each. Two convolutional layers are the following layer after LSTM, where 64 and 128 filters are contained inside the first and second convolutional layers. A max-pooling layer is also included between the two convolutional layers. The model's last three layers are the global average pooling layer (GAP) and one batch normalisation layer, followed by one dense layer with the SoftMax classifier as the output layer. The model's overall accuracy for the UCI-HAR dataset is 95.80 percent, 95.75 percent for the WISDM dataset, and 92.63 percent for the OPPORTUNITY dataset. In the UCI-HAR dataset, performance was quite bad in sitting and standing, which might be attributed to the two activities being identical from a sensor standpoint. Finally, the approach used in this paper had an overall accuracy of 87.58 percent.

In 2021, a multichannel CNN-LSTM network for daily activity recognition using smartwatch sensor data has been proposed. The method is proposed by Mekruksavanich and Jitpattanakul (2021). The daily human activity (DHA) utilised was given by Kokorin University in the Republic of Korea, and there are 11 activities in this dataset includes reading, eating, cooking, and many more. CNN was employed to generate the feature maps from the input data for the multichannel CNN-LSTM architecture used in this study. The flattened feature maps were then fed into the LSTM layers. Lastly, at the complete linked layer, the performance of every channel was merged, and the labels were predicted. This paper has conducted three experiments. The first experiment utilises the DHA dataset to undertake a 10-fold

cross-validation procedure and the recognition performance of two fundamental deep learning models, CNN and LSTM, with results of 91.022 percent and 84.347 percent, respectively. The multichannel CNN-LSTM was then employed in the second experiment to train with window sizes of 3, 5, 10, 30, and 60 seconds, with the 10 seconds window size achieving the greatest accuracy of 94.55 percent. Finally, the third experiment added the position data to the multichannel CNN-LSTM model that could increase efficiency. After adding the position data, the accuracy improved to 96.87 percent, which is the maximum accuracy by using the window size of 10. In conclusion, the recognition system suggested in this study using the multichannel CNN-LSTM network in this paper outperformed other basic deep learning networks with a high accuracy of 96.87 percent.

3. Proposed Solutions

3.1. Dataset

The Toyota Smarthome dataset developed by Toyota Motor Europe (TME) is used in this study. The method is proposed by (Das, Koperski, Minciullo, Garattoni, Bremond and Francesca (2019). This contains 16115 videos from 18 elderly subjects each for RGB data, skeleton data, and depth data. Besides, the skeleton data comes in two versions which are V1.1 and V1.2. The V1.1 skeleton data was the original data, while V1.2 was the skeleton data that are refined. In this project, the skeleton data V1.2 was used, and the dataset contains activities like eating at a table, getting up, lying down, and many more. Some samples activities contained in the Toyota Smarthome dataset are illustrated in Figure 1. The nature of 2d pose data is that it is the key points from the image in term of pixel level. The location of the key point is represented as X and Y. After that, the nature of 3D data is that the object is transformed into a 3D image from a 2D image by additional z-axis to the prediction of the output. Some samples for the CHU dataset are also illustrated in Figure 2.

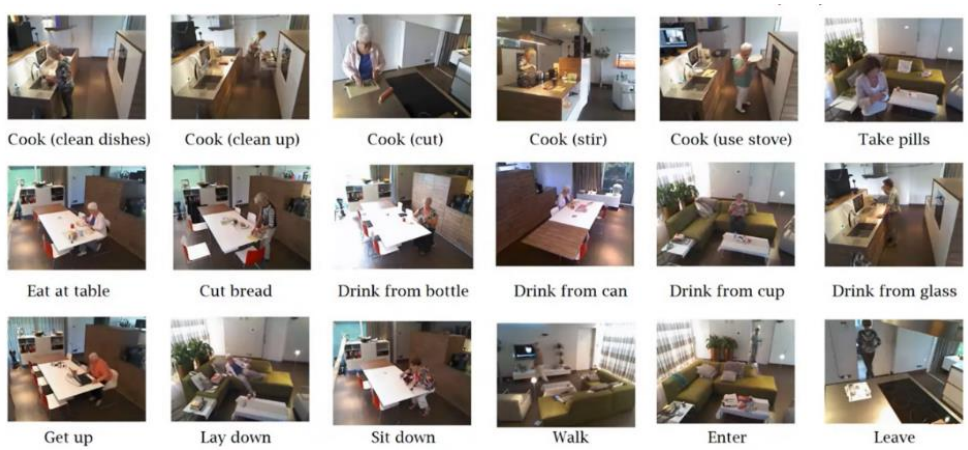

Fig. 1: Sample activities contained in the Toyota Smarthome dataset

3.2. Variants of LSTM

A number of variations of LSTM are explored in this study, including Recurrent Neural Network, Long Short-Term Memory, Bidirectional Long Short-Term Memory and Gated Recurrent Unit. The details of each method are delineated in the subsequent section. LSTM and its variants are used because the models' contextual dependency in the temporal domain are quite effectively.

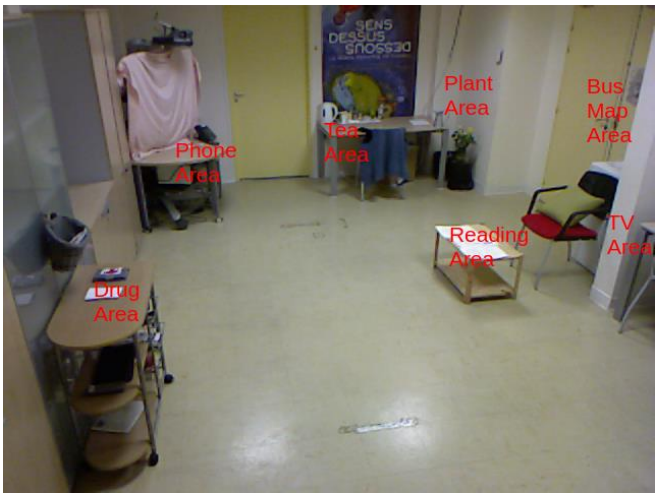

Fig. 2: Sample image of the CHU activities dataset.

3.3. Recurrent Neural Network (RNN)

RNN is a variation of neural networks that is more helpful in modelling sequence data. The method is proposed by Donges (2021). RNN is the first algorithm that remembers its input using internal memory, and it was initially created in the 1980s. Therefore, it is suitable for sequential data in machine learning problems. Furthermore, RNN is utilized by Apple's Siri and Google's voice search since it is the one and only one with internal memory, and it is recognized as the most promising algorithm.

3.4. Long Short-Term Memory (LSTM)

LSTM can be considered as a recurrent neural network extension that extends memory, allowing them to learn order dependency in sequence prediction tasks. The method is proposed by Donges (2021). As a result, it is suitable for activities that have a long-time duration in between. LSTM is a better version of RNN because LSTM operates like computer memory, and it can remember the inputs for a long period. LSTM can also read, write, and erase data from its memory in the same way that a computer can.

3.5. Gated Recurrent Unit

GRU aims to overcome the vanishing gradient issues that a standard RNN encounters.

The method is proposed by Kostadinov (2019). Furthermore, GRU may be seen as a version of LSTM since they are both built similarly and provide equally excellent results in certain instances. GRU employs the update and resets gates to circumvent the vanishing gradient issue of a normal RNN. They are simply two vectors that determine which content should be sent to the output. After that, they are special in that they can be designed to keep knowledge from the past from being washed away over time, as well as to reject data that is unrelated to the forecast.

3.6. Bidirectional Long Short-Term Memory (Bi-LSTM)

Bidirectional LSTM, also known as Bi-LSTM, is a type of LSTM that enhances the efficiency of the algorithm on sequence classification issues. The method is proposed by Verma (2021). In Bi-LSTM, it differs from conventional LSTM in that input flows in two ways. Using a standard LSTM, we can only make input flow in one direction, maybe either backwards or just forwards. However, by employing bi-directional, we may allow the same information to flow in both directions, maintaining both the future and the past.

3.7. Fusion of 2D and 3D Pose Data

The Toyota Smarthome dataset was recorded in an apartment with 7 Kinect v1 cameras. The dataset contains 18 subjects range 60 – 80 years old doing common daily living activities. On top of that, the dataset has a resolution of 640 x 480, and it offers three modalities: RGB, Depth, and 3D Skeleton. The skeleton data contained two versions, V1.1 and V1.2. The V1.1skeleton data are extracted using the LCRNet and output to JSON format, while the V1.2 is refined using SSTA methods, which makes it tackle the occlusion, truncation, and low-resolution issues in the pose-estimation stage.

Both 2D and 3D pose data are utilised in this study. The pose information are stored in a CSV file. There are altogether 26 and 39 features for 2D and 3D poses, respectively. After the data is being read from the CSV file and assigned to a data frame. Then, the 2D and 3D data frames were concatenated to form a single feature vector so that it could be input into the deep learning model such as LSTM. The initial sizes for pose2dand pose3d data are 1x26 and 1x39, respectively. After fusion, the final feature size is 1x65.

4. Experiments and Discussions

4.1. Evaluation using pose2d data

We first perform tests for the pose2d data. The model uses one LSTM with 16 units when it comes to hyperparameters. The LSTM layer followed a dropout layer with a 0.5 dropout rate. Following that, two dense layers were created, the first dense layer has 32 units with an activation function ‘relu’ while the second dense layer is used as an output layer with the activation function ‘softmax’ that are suitable for multi-class

classification was used to classify the five daily life activities. Finally, the model is compiled with a loss function called 'categorical crossentropy' that can be used in multi-class classification tasks and 'RMSProp' as an optimizer. 'RMSProp' is intended to speed up the optimization process, for example, by reducing the number of function evaluations necessary to achieve the optima or increasing the optimisation method's capabilities, resulting in a better result.

Model fitting is a measurement of how well a machine learning model generalized to similar data to that on which it was trained. When it comes to model fitting, the model was trained with 100 epochs, batch size of 64 and shuffle was set to false. The shuffle of the model was set to false because we did not want to reshuffle the dataset every time. The data used in this experiment contain 250 pieces of data for each activity where 200 data for training and 50 data for testing. The first experiment was tested with different variations of LSTM with the hyperparameter mentioned above. The accuracy for pose2d data for different variations of LSTM in the Toyota Smarthome dataset for experiment 1 is shown in Table 1.

Table. 1: Experiment results for pose 2D data

LSTM Variation	Results for pose 2d data (%)	Loss for pose2d data
Bidirectional LSTM	92.50	0.7698
GRU	85.00	0.7268
Normal LSTM	93.75	0.5230

4.2. Evaluation using pose3d data

The same experimental setting as pose2d data is used to evaluate the pose3d data. Table 2 presents the experimental results for pose3d data. We observe that the LSTM approach yields the best accuracy when dealing with the 3d data.

Table. 2: Experiment results for pose 3D data

LSTM Variation	Results for pose 3d data (%)	Loss for pose3d data
Bidirectional LSTM	95.00	0.1345
GRU	84.17	1.1004
Normal LSTM	97.08	0.2082

4.3. Evaluation using fused data

The fusion of pose2d and pose3d data in the Toyota Smarthome dataset is tested in this section. The settings for this experiment were the same as the previous

experiments. The performances of the different LSTM variants are shown in Table 3. We can observe from the table that both conventional LSTM and bidirectional LSTM yield promising results. The loss and accuracy graphs for the best performing method, i.e. conventional LSTM is depicted in Figure 3, together with the corresponding architecture given in Figure 4.

Table. 3: Experiment results for fused data

LSTM Variation	Results for fusion data (%)	Loss for fusion data
Bidirectional LSTM	94.17	0.2306
GRU	89.58	0.6901
Normal LSTM	94.58	0.2706

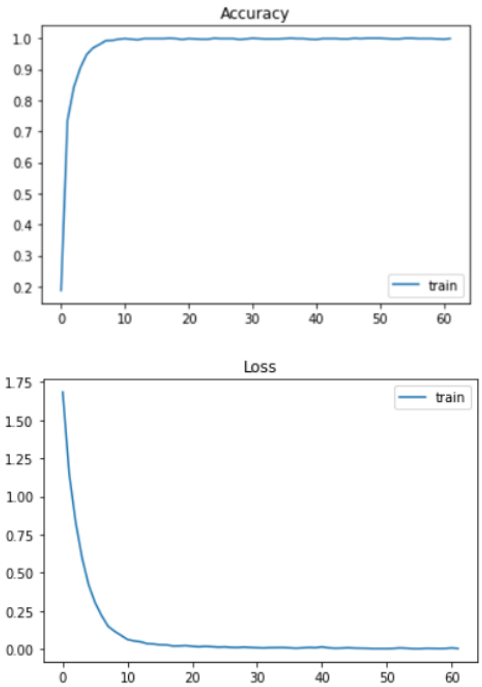

Fig. 3: Loss and accuracy graph for best results in fusion data

4.4. Evaluation using the CHU dataset

For the first experiment, the initial hyperparameter for the CHU dataset was the same as the Toyota Smarthome dataset. Therefore, the model is first using one LSTM with 16 units. After that, the LSTM layer was then followed by a dropout layer with a 0.5 dropout rate. Following that, two dense layers were created, the first dense layer has 32 units with an activation function 'relu' while the second dense layer is used as an output layer with the activation function 'softmax' that are suitable for multi-class classification was used to classify the five daily life activities. Finally, the model is compiled with a loss function called 'categorical crossentropy' that can be used in multi-class classification tasks and 'rmsprop' as an optimizer.

For model fitting, the model was trained with 100 epochs, batch size of 64 and shuffle was set to false. The shuffle of the model was set to false because we did not want to reshuffle the dataset every time. The data used in this experiment contain 250 pieces of data for each activity where 200 data for training and 50 data for testing. The first experiment was tested with column data using the hyperparameter mentioned above and the results are presented in Table 4.

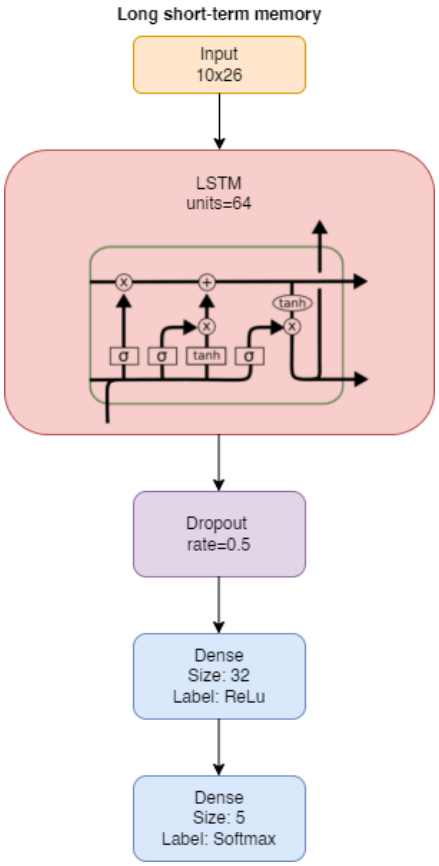

Fig. 4: Model Architecture for best result in pose2d data

Table. 4: Experiment results for pose2D data in CHU dataset

LSTM Variation	Results for pose 2d data (%)	Loss for pose2d data
Bidirectional LSTM	82.92	2.6450
GRU	81.67	1.9082
Normal LSTM	83.33	1.7000

The 3d data in the CHU dataset was also tested. The settings for this experiment were the same as with the experiment for 2d data. It was tested with different variations of LSTM and the experimental results are recorded in Table 5.

Table. 5: Experiment results for pose3D data in CHU dataset

LSTM Variation	Results for pose 3d data (%)	Loss for pose3d data
Bidirectional LSTM	89.58	1.8185
GRU	87.50	1.5826
Normal LSTM	88.33	1.8612

In addition, the fusion data in the CHU dataset was also evaluated. The settings for this experiment were the same as the previous experiments for 2d and 3d data. It was tested with different variations of LSTM and the empirical results are presented in Table 6. The loss graphs for the best performing methods, together with the corresponding model are illustrated in Figure 4 and Figure 5, respectively.

Table. 6: Experiment results for fused data in CHU dataset

LSTM Variation	Results for pose 3d data (%)	Loss for pose3d data
Bidirectional LSTM	98.26	0.0455
GRU	95.65	0.1043
Normal LSTM	97.39	0.0668

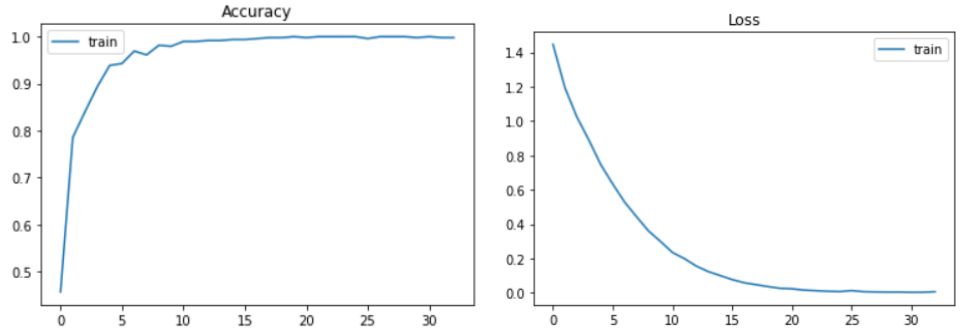

Fig. 5: Loss and accuracy graph for best results in fusion data in CHU dataset

4.5. Discussion

From all the experiment that has been conducted using the Toyota Smarthome dataset and CHU dataset, there are some important and interesting findings from the experimental results. In these experiments, different variations of LSTM have been tested, and different hyperparameter tuning has been made, including the units for one LSTM layer, number of patience in early stopping, optimizer, dropout rate, and size of training data. In addition, the CHU dataset was tested with one more experiment, which is testing whether using all attributes given except the time feature or selecting certain attributes will get a better result or accuracy.

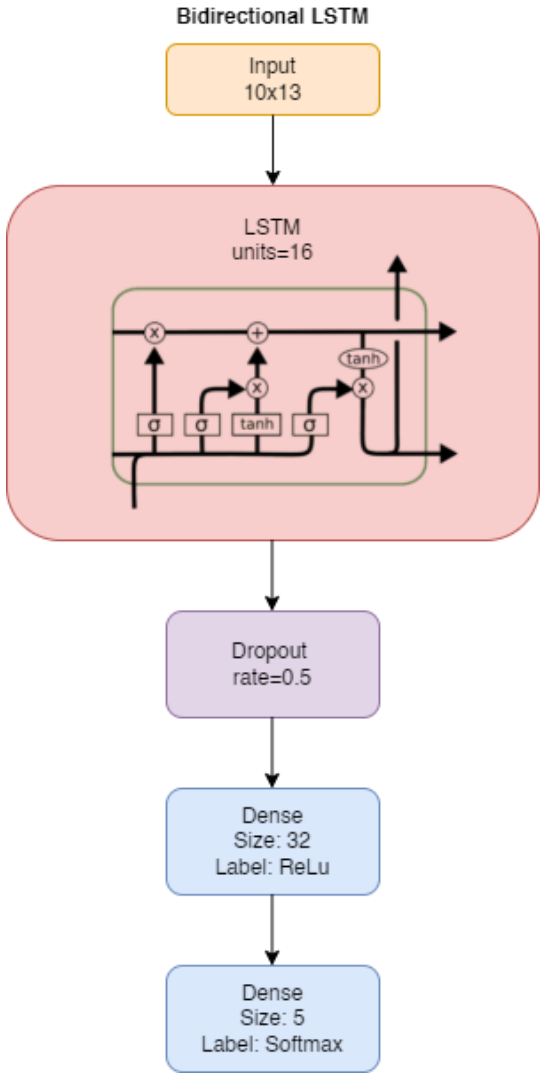

Fig. 6: Model Architecture for best results in pose2D data in CHU dataset

First and foremost, it can observe that the pose2d data from the Toyota Smarthome dataset gets the highest accuracy of 94.17% with the settings of normal LSTM, 64 units for one LSTM layer, number of patience 5 in early stopping, Adam as an optimizer, 0.5 dropout rate, and 200 sizes of training data for each activity in the dataset. After that, the pose3d data gets the highest accuracy of 99.17% with normal LSTM, 64 units for one LSTM layer, number of patience 2 in early stopping, Adam as an optimizer, 0.1 dropout rate, and 200 sizes of training data for each activity in the dataset. Finally, the fusion data gets the highest accuracy of 96.67% with normal LSTM, 64 units for one LSTM layer, number of patience 6 in early stopping, Adam as an optimizer, 0.5 dropout rate, and 200 training data sizes for each activity

in the dataset. The interesting findings in these experiments are that the commonalities for all pose2d, pose3d and fusion data in the Toyota Smarthome dataset used normal LSTM, 64 units of LSTM, Adam optimizer and 200 training data in order to get their best results. The difference between these experiments to get the highest accuracy is between the number of patience in early stopping and the dropout rate.

From here, it is shown that the pose3d data in the Toyota Smarthome dataset gets the highest accuracy which is 99.17%, among all the experiments that have been conducted for the Toyota Smarthome dataset and CHU dataset. In comparison, the lowest accuracy observed is the pose2d data in the Toyota Smarthome dataset, which gets an accuracy of 94.17%. This may be caused by the pose3d data containing more attributes than pose2d data, where pose3d data has 39 attributes while pose2d data only contains 26 attributes. Therefore, the fusion data that fuses the pose2d and pose3d data gets an accuracy between them might be caused by the data that was being fused having some valuable attributes, but some attributes might be not that useful.

Furthermore, it can be observed that the best results for all 3 types of data, which are pose2d, pose3d and fusion data using 64 units for one LSTM layer inside the experiment for the Toyota Smarthome dataset to get the best result. This may be caused by the training data being used is not so big, so 64 units in the LSTM layer is sufficient to get the optimal performance.

When it comes to the CHU dataset, the 2d data gets the highest accuracy of 98.89% with the settings of using all attributes except time feature, normal LSTM, 80 sizes of training data for each activity in the dataset, 16 units for one LSTM layer, number of patience 2 in early stopping, Rmsprop as an optimizer, and 0.5 dropout rate. Following that, the 3d data gets the highest accuracy of 96.52% with the settings of using all attributes except time feature, bidirectional LSTM, 100 sizes of training data for each activity in the dataset, 256 units for on LSTM layer, number of patience 3 in early stopping, Rmsprop as an optimizer, and 0.1 dropout rate. Finally, the fusion data which fuse the 2d and 3d data gets the accuracy of 98.26% with the settings of using all attributes except time feature, bidirectional LSTM, 100 sizes of training data for each activity in the dataset, 32 units for one LSTM layer, number of patience 2 in early stopping, Rmsprop as an optimizer, and 0.3 dropout rate. The interesting findings from the experiment in the CHU dataset are that they used all attributes to perform better because all attributes consist of more details about the data. Moreover, all these experiments also used Rmsprop as an optimizer to get the best result.

Finally, the experiments also show that the 3d data and fusion data in the CHU dataset gets their best performance by using Bidirectional LSTM. This may be caused by the characteristic of Bidirectional LSTM that will run the input provided in two ways, one from past to future and one from future to past, and because of running

backwards in Bidirectional LSTM can preserve information from the future and using two hidden states combined it is able in any point in time to preserve information from both past and future.

5. Conclusion

This study proposes an activities of daily life recognition system using machine learning approaches. The Toyota Smarthome and CHU datasets have been assessed using different variations of the LSTM models. Empirical evidences show that machine learning approaches are promising directions to perform activities of daily life recognition with high accuracy rates. Comprehensive evaluations have been performed with different variations of LSTM, increasing/decreasing the number of units in the LSTM layer, and performing a number of hyperparameter tuning. The proposed model reported an accuracy of 94.17% for pose2d data, 99.17% for pose3d data and 96.67% for fusion data from all the experiments conducted in the Toyota Smarthome dataset.

Acknowledgements

This project is supported by the Fundamental Research Grant Scheme (Grant no. FRGS/1/2020/ICT02/MMU/02/5). The authors would also like to thank the providers of the Toyota Smarthome Dataset for sharing their databases.

References

Ahmad, W., Kazmi, B. M., & Ali, H. (2019). Human Activity Recognition using Multi-Head CNN followed by LSTM. 15th International Conference on Emerging Technologies, ICET 2019, 6–11. https://doi.org/10.1109/ICET48972.2019.8994412

Ayman, A., Attalah, O., & Shaban, H. (2020). Smart system for recognizing daily human activities based on wrist IMU sensors. 2019 International Conference on Advances in the Emerging Computing Technologies, AECT 2019, 0–5. https://doi.org/10.1109/AECT47998.2020.9194154

Das, S., Dai, R., Koperski, M., Minciullo, L., Garattoni, L., Bremond, F., & Francesca, G. (2019). Toyota smarthome: Real-world activities of daily living. Proceedings of the IEEE International Conference on Computer Vision, 2019-Octob, 833–842. https://doi.org/10.1109/ICCV.2019.00092

Donges, N. (2021, July 29). A guide to Rnn: UNDERSTANDING recurrent neural networks and LSTM NETWORKS. Built In. Retrieved September 13, 2021, from https://builtin.com/data-science/recurrent-neural-networks-and-lstm.

Hou, C. (2020). A study on IMU-based human activity recognition using deep learning and traditional machine learning. 2020 5th International Conference on Computer and Communication Systems, ICCCS 2020, 225–234. https://doi.org/10.1109/ICCCS49078.2020.9118506

Kostadinov, S. (2019, November 10). Understanding GRU Networks - Towards Data Science. Medium. https://towardsdatascience.com/understanding-gru-networks-2ef37df6c9be

Mekruksavanich, S., & Jitpattanakul, A. (2021). A Multichannel CNN-LSTM Network for Daily Activity Recognition using Smartwatch Sensor Data. 2021 Joint 6th International Conference on Digital Arts, Media and Technology with 4th ECTI Northern Section Conference on Electrical, Electronics,Computer and Telecommunication Engineering, ECTI DAMT and NCON 2021, 277–280. https://doi.org/10.1109/ECTIDAMTNCON51128.2021.9425769

Minarno, A. E., Kusuma, W. A., Wibowo, H., Akbi, D. R., & Jawas, N. (2020). Single Triaxial Accelerometer-Gyroscope Classification for Human Activity Recognition. 2020 8th International Conference on Information and Communication Technology, ICoICT 2020. https://doi.org/10.1109/ICoICT49345.2020.9166329

Park, J., Jang, K., & Yang, S. B. (2018). Deep neural networks for activity recognition with multi-sensor data in a smart home. IEEE World Forum on Internet of Things, WF-IoT 2018 - Proceedings, 2018-Janua, 155–160. https://doi.org/10.1109/WF-IoT.2018.8355147

Patel, A. D., & Shah, J. H. (2019). Performance analysis of supervised machine learning algorithms to recognize human activity in ambient assisted living environment. 2019 IEEE 16th India Council International Conference, INDICON 2019 - Symposium Proceedings, 2019–2022. https://doi.org/10.1109/INDICON47234.2019.9030353

Verma, Y. (2021, November 20). Complete Guide To Bidirectional LSTM (With Python Codes). Analytics India Magazine. https://analyticsindiamag.com/complete-guide-to-bidirectional-lstm-with-python-codes/

Vijayvargiya, A., Kumari, N., Gupta, P., & Kumar, R. (2021). Implementation of machine learning algorithms for Human Activity Recognition. 2021 3rd International Conference on Signal Processing and Communication, ICPSC 2021, May, 440–444. https://doi.org/10.1109/ICSPC51351.2021.9451802

Xia, K., Huang, J., & Wang, H. (2020). LSTM-CNN Architecture for Human Activity Recognition. IEEE Access, 8, 56855–56866. https://doi.org/10.1109/ACCESS.2020.2982225

Zebin, T., Sperrin, M., Peek, N., & Casson, A. J. (2018). Human activity recognition from inertial sensor time-series using batch normalized deep LSTM recurrent networks. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2018-July, 1–4. https://doi.org/10.1109/EMBC.2018.8513115