Human Activity Recognition Using Smartphone Sensor Data

Aakash Sai Unnam
2019A7PS0196U

Asmi Sriwastawa
2019A7PS0147U

Safiya Ateeq
2019A7PS0183U

Abstract: Recognition of human action has grown in popularity among scientists in recent years. Because of its immediate applications in numerous industries, including healthcare and fitness, it is in the focus. Additionally, the widespread usage of smartphones nowadays makes it particularly simple to collect this kind of information from people in an unobtrusive and affordable manner without the use of additional devices. This paper aims at bringing light to human activity recognition using smartphone data and proposes an approach that uses BiLSTM to predict activities such as walking, laying, standing, walking upstairs/downstairs and sitting as performed by the user. The HAR Dataset from the UCI dataset repository is used. This dataset was compiled from 30 individuals who were wearing smartphones around their waists while engaging in the variety of activities mentioned above. Our proposed model obtained an overall accuracy of 81.3%. It was compared to various other Machine learning and Deep Learning approaches applied to the same dataset.

Keywords: Human Activity Recognition; LSTM; BiLSTM; Smartphones; Classification

1. Introduction

Human Activity Recognition or HAR is the system of interpreting human activity using artificial intelligence. Human activities can be interpreted as physical activities, gestures and behaviors sensed using a sensor. The data collected from sensors is translated into commands for the model to understand, analyze and execute human activity recognition.

The sensor data can be remotely recorded by using videos, radars or other wireless methods or directly by wearing sensor devices or using smartphones which contain accelerometers and gyroscopes.

In recent years, there has been a lot of interest regarding this topic, due to the availability of better data acquisition techniques and the low cost of power sensors which can even be embedded into smartphones. This growing interest can be useful in many real life situations such as HAR detection for the health and care of humans or in driverless cars to identify pedestrians.

Some of the main applications of human activity recognition are in medical applications, robotics, arts and entertainment, gesture and position analysis, smart surveillance, behavioral biometrics and sports.

Many HAR systems have been studied over the last few decades, with the researchers focusing on various activities in distinct application domains. Walking, sitting, running, exercising, and other activities are examples of such activities. In terms of duration and complexity, the activities can be divided into three categories: short activities, simple activities, and complex activities. Short activities consist of quick and short duration activities, including standing up from a sitting position. Walking and reading are examples of basic activities in the second category. The final one essentially combines progressions of fundamental tasks with contact with various things and people. These are more complex and long tasks such as attending a business meeting, partying, etc.

Physical human activity recognition is classified mainly into 3 categories namely vision-based, sensor-based and multimodal. The vision-based technique is restricted to the sensitivity of light in the environment and detects at a low range. The non-vision technique is primarily used for physical human activity recognition, which employs a wide range of sensors on wearable devices such as smartphones and smartwatches.

- Vision-based: Depth cameras and RGB video are used to record human behaviors.
- Sensor-based: This approach makes use of sensors such as wearable and ambient sensors that detect human activity.
- Multimodal: Both visual and sensor data are used to identify human activity.

Data collection for HAR is a daunting task. So many different variables are involved and that makes it difficult to gather proper data. The two biggest difficulties are handling the massive amounts of data that the devices can generate and secondly, not knowing how to connect this data to the corresponding physical movements such as walking, running, sitting, etc. Devices like smartwatches are gathering data in a limited number of activities and with specified device orientations. Given that each person uses their smartphone in a unique way, this is far from the optimal condition in terms of these orientations. As a result, even though some models perform well and get good results, the models' success may be biased.

IoT development is accelerating quickly in the field of HAR. The compatibility of IoT with wearable sensors, network objects, and traditional networks is the main driving force behind the evolution. For instance, body sensor nodes, which combine wireless detection networks and body sensor systems, are among the most crucial IoT technologies. Similarly, deep learning is increasing in popularity in human activity recognition. Deep learning is a type of neural network-based machine learning in which data is processed through successive layers to extract increasingly complex properties or features. It has become increasingly useful in extracting appropriate data from huge datasets. A typical machine learning technique requires manual feature extraction whereas the deep learning approach automates feature extraction and this has laid the foundation for impressive deep learning approaches to perform HAR.

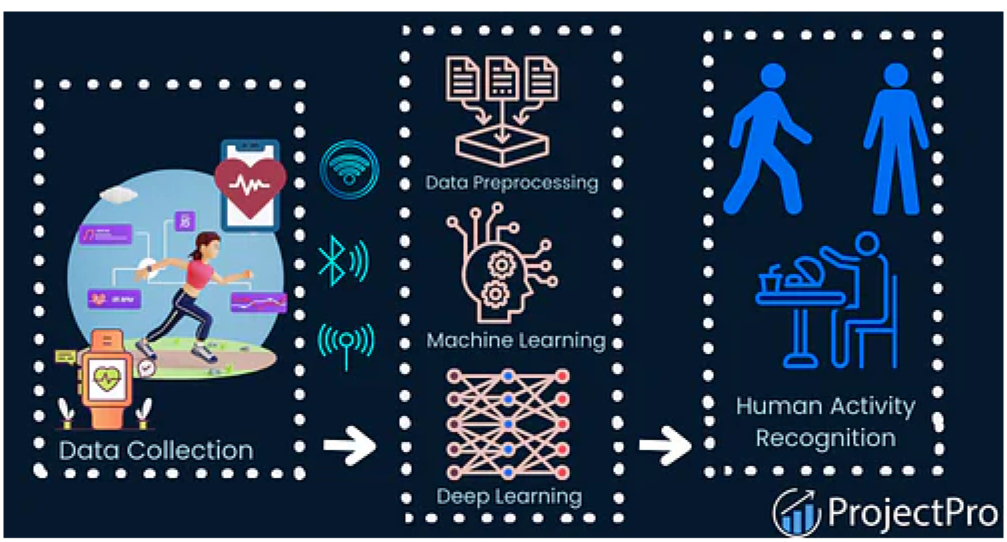

Figure 1. Process of Human Activity Recognition

2. Literature Review

In this section, we look into multiple implementations of human activity recognition algorithms. We have divided this section into 3 subsections. In section 2.1, research done with self-collected data is talked about. In section 2.2 studies involving deep learning techniques are looked into and in section 2.3, other significant works related to HAR are discussed.

In all the works discussed below, a general approach [8] is observed for HAR applications that involves data acquisition, data preprocessing, feature extraction and finally activity classification, in that order. The challenge in this field usually arises from individual differences in activity patterns that no dataset thus far has fully encompassed. As such, while surveying the literature we keep in mind that there should be a focus on free-living environments, algorithmic fairness, diversity in data used and scope for real world implementation when highlighting limitations.

2.1. Studies with Self-Collected Data

These papers collected data through android applications and used those readings to perform HAR. The first work delivers a more realistic dataset that is not dependent on the device's orientation and placement. Additionally, this study also proposes a support vector machine (SVM) model using the more realistic dataset provided by this study. For data collection, every person involved in the collection was required to set one of the four activities on the app before they start their session. The four activities were: Inactive, Active, Walking, and Driving. So, each session corresponds to one specific activity. The sensors used were accelerometer, gyroscope, magnetometer, and GPS. Data preprocessing and feature selection were performed. The features extracted were mean, variance, median absolute deviation, maximum, minimum, and interquartile range. Finally, the SVM model was employed. To obtain good results, SVM classifiers were applied which were intended to look out for the best kernel among RBF, polynomial, and linear SVM. It was noted that the best results corresponded to the RBF kernel. An f1-score of 64.14% and an accuracy of 67.22% were achieved. The model was also tested on other datasets which had only 2 sensors and the results obtained were bad compared to the proposed dataset with 4 sensors. The model proposed is relatively simple and is just a first approximation. Different frequencies of sensor data lead to gaps in measurements. This in turn led to uneven distribution of data for each activity. [1]

In the second work, a system for recognizing human physical activity based on information gathered from smartphone sensors was proposed. The suggested method calls for creating a classifier utilizing a smartphone's accelerometer, gyroscope, and gravity sensor. Data was first collected using an android application. The user could choose the activity, viz. walking, running, sitting, standing, and ascending/descending stairs. After the data was collected, features were extracted. The five features were Average, Average Absolute Difference, Standard Deviation, Average Resultant Acceleration, and Histogram. To train the classifier, this study used a multi-layer perceptron (MLP) due to its simplicity and resource consumption. The neural network outputs the probabilities corresponding to each activity. Its accuracy in the activities (excluding stairs) was above 92%. Some activities boasted a 94% accuracy whereas some were difficult to identify. The model struggled to differentiate between climbing and going down the stairs. The results were also compared with another external solution based on Naïve Bayes. The results of the proposed method were significantly better. [2]

In another work, for data collection, an Android-based app was used which fetched sensor data from the 3 sensors (Accelerometer, gyroscope, and magnetometer) along with the Google Fit API. But the high accuracy mode in the Google Fit app consumes a lot of the battery in a smartphone. [5]

A self-collected dataset was solely created by using a Samsung Galaxy S11 Android smartphone. Unfortunately, a limited number of dynamic activities were recorded and adding more activities would have made it a more comprehensive dataset. [14]

One particular study [12] used a custom dataset created using a Redmi smartphone, operated by four test subjects. Accelerometer and gyroscope data were collected at 20 Hz frequency and were used to address the problem of real time recognition of human activities using incremental learning algorithms. Six incremental learning algorithms – Incremental KNN (IKNN), Incremental Decision Trees, Incremental Random Forests, Incremental Adaboost, Incremental Naive Bayesian (INB) and Learn++ NSE – were embedded in a client-server architecture. Normalization and feature extraction was applied to the collected data before being fed into the chosen algorithm. Active learning was used to train the models over time. IKNN and INB were found to perform the best in terms of all classification metrics as well as convergence speed. Both require a small amount of labeled data to start performing well. They are also less sensitive to outliers. All incremental algorithms were also compared to their static/offline counterparts. KNN and NB showed the same performance in either case but other algorithms exhibited much higher performance in their static versions. It is mentioned that the real time system can be made lighter in terms of energy use, memory requirements and computational complexity for use in common smart devices.

2.2. Approaches using Deep Learning Techniques

A smartphone sensor-based approach for human activity recognition using Deep Belief Networks (DBN) was introduced. This system has 3 parts to it: sensing, extracting features and recognition. Triaxial accelerometers and gyroscope sensors were used in this study to fetch data.The four activities measured were standing, lying, sitting, and walking. After the data was collected, noise in the data was removed and statistical analysis was performed to find features. Dimension reduction was then performed using KPCA. Next, the neural network is formed. The proposed DBN consists of an input layer, 2 hidden layers and an output layer. To pre-train the algorithm, Restricted Boltzmann Machine (RBM) was used and then a backpropagation algorithm was run to adjust the parameters. Three experiments were run. Artificial Neural Network (ANN) was first tested and obtained an accuracy of 89.05%. Next, a multiclass SVM was applied and an accuracy of 94.12% was obtained. Finally, the proposed DBN method was used and the highest accuracy of 95.85% was obtained. [3]

The accelerometer, gyroscope, magnetometer, and Google Fit activity tracking module are combined in another paper to present a novel framework based on deep recurrent neural networks (DRNN). After the data was collected and preprocessed, the features were extracted and put into a matrix. This acts as the input for the proposed neural network. It is then trained on the dataset and the output of the DRNN predicts the activity performed. The predicted label is sent to the Agent-based analysis (ABA) along with the labels from the Google Fit API. ABA is important because it is used to check if the predicted label is accurate or not. The proposed framework was then compared to other existing studies and compared the precision, recall, and F1 score values. Without the ABA feature, an accuracy of 93.47% was seen but with the use of the ABA feature, an accuracy of 99.43 was achieved. It is also seen that Machine learning models do get good results, but deep learning models are better because you don’t need to worry about feature engineering. Also, when datasets are large, deep learning techniques perform better. [A5]

Another proposed approach was a lightweight and efficient model based on CNN and LSTM. It is compared with other deep learning models in terms of accuracy. A model's observations can be standardized beforehand for greater improvement and better outcomes. The Gaussian approach can be used to shift the distribution of each variable's values, which yields parameters. This is done on the dataset.The construction of a one-dimensional convolutional neural network model is done for the recognition of the human everyday activities dataset. The obtained set of observations is utilized to extract the features as well as map the internal features into various activities. Although the CNN has not been trained, training can be provided by back-propagating errors from the LSTM deep learning classifier. The performance metrics used in this study were accuracy, recall, precision, and F1 score. The experiment was also tested on an ML model using a SVM classifier and produced an accuracy of 89%. The proposed model gave an accuracy of 95%. [6]

Another study [13] also focused on a combined CNN-BiLSTM model for HAR applications. The aim was to capture both spatial and temporal aspects of mobile sensor data. WISDM and UCI-HAR datasets were used to generate input data which is fed to both CNN and Bi-LSTM models in parallel. Their outputs are concatenated and the fused features are used to predict the activity being performed. The hybrid model obtained very good accuracy for both datasets. To expand the study, CNN-BiLSTM was used as a feature extractor in an instance while SVM was used as a classifier. The results showed no changes for WISDM and slight improvement for UCI-HAR. In comparison to other models in the literature that have also worked on WISDM, this hybrid model performed better. Perhaps fusion of features at various levels of the CNN and BiLSTM models can be implemented, rather than just at the output layer. The features derived from the combined models can also be compared with current successfully implemented hand crafted features in the literature.

In this article, a deep neural network model is proposed which recognizes human physical activity in real-time, collected from tri-axial accelerometer data using a mobile. The model is created by combining feature extraction and convolutional layers. The Tensorflow Lite framework is used to export trained models in a mobile compatible format. This model takes 5 - 8 times less storage space and produces double the throughput of the current state-of-art implementation on the WISDM dataset and improves the classification performance. A classification accuracy of 94.18% is achieved on a 10-fold user independent cross validation of WISDM dataset. On testing the model with the Actitracker dataset, an accuracy of 79.12% is achieved. When implementing on a mobile, the storage is restricted so the size of the neural network always needs to be checked. [14]

Another approach to improve physical activity recognition is by creating a model that consists of two modules. In the first module, the model segments the acceleration signals into overlapped windows and from each window present in the frequency domain, the information is extracted. In the second module, using CNN, the performance activity at each window is detected. The first layer collects data from each sensor independently while the second layer classifies them into an activity. The main advantage is that it is easy to add new sensors without actually modifying the structure. After applying the post-processing techniques, the accuracy improved from 89.83% to 96.62%.Low range of variability of physical activities and subjects was its limitation. [15]

From the self collected data from the Samsung S11, 70% of the total data was used as training dataset and the remaining 30% was used as testing dataset. The training dataset is tested on both time and frequency domain. To check the performance of the model, two conventional machine learning techniques such as SVM and Ensemble Learning Machine and one DL of CNN are compared. The performance of the classifiers are not always the same. Machine learning algorithms such as SVM and ElM with additional handcrafted features had better performance compared to CNN in the case of public dataset. In the case of the custom dataset, CNN with additional handcrafted features had better performance. Feature fusion of handcrafted features resulted in better performance of the model. The proposed approach outperforms HAR systems based on state-of-art methods for both dataset in terms of accuracy, testing and training time. [17]

The traditional approaches based on hand crafted features are not capable of incremental learning as they use fixed mathematical formulas on input data. In this paper, a LSTM network is proposed to recognize six different human behaviors/activities based on smartphone data. The network has five LSTM cells connected end to end that are trained on sensor data. A single layer neural network is attached before the network to preprocess the data for the stacked LSTM network. To improve the generalization of the network a L2 regularizer is used in the cost function. The proposed network improves the average accuracy by 0.93% as compared to the closest state-of-the-art method without any manual feature engineering. [19]

A more recent development in deep learning for HAR has been the use of federated learning. One paper [11] proposed a model FedHAR that combines semi-supervised and federated learning to create a realistic global activity model for mobile HAR that also operates at a personalized level (using transfer learning) while respecting user privacy, all using minimal labeled data. The system basically consists of a global model (a deep neural network) hosted by a server which is initialized using public HAR datasets. A feedback loop is created between this server and all mobile devices in the network. A local model for every mobile is generated by copying the global model and adding personalized user features to it. In such a system both generalization and personalization are achieved over a period of time. It is seen that the F1 score of the model improves continuously as more data is received. When compared to other federated models that were trained and evaluated on fully labeled datasets in more controlled settings, FedHAR's performance was found to be on par. The advantage is that it can be applied to more realistic settings with regards to HAR applications. Even so, shared global model weights may still reveal sensitive information about participating users. Larger and more varied datasets can also be used for the initialization of the global model. The model remains unoptimized in terms of computational efficiency as well.

2.3. Other Related Works in HAR

This study has suggested a hybrid feature selection method with a filter and wrapper approach to curb the occurrence of 'curse of dimensionality'. For enhanced activity recognition, this technique employs a sequential floating forward search (SFFS) to extract the desired features. The sensor data used is from accelerometer and gyroscope. Various statistical features can be obtained from the sensor data. The proposed model is divided into 2 parts. In the first part, optimal features are selected using the hybrid model which works on the sequential floating forward search (SFFS) algorithm. In the second part, training and validation is done using the multiclass SVM. Classification accuracies for the sub optimal feature sets are calculated.The wrapper technique chooses the top feature out of the 10 features that are ideal for each sensor. The final optimal features are therefore chosen. These features are then validated through the validation process. With the feature selection model proposed in this paper, an accuracy of 96.61% is achieved whereas an accuracy of 90.84% is seen without feature selection. The drawback is that sensors, preferably body-mountable, might be used to provide continuous movement tracking in order to gain a better understanding of the activities performed. [4]

In this paper, SelfHAR, a semi-supervised model that efficiently learns to use unlabeled mobile sensing datasets to supplement limited labeled datasets is proposed. Semi-supervised learning techniques have been suggested to supplement situations when there is a dearth of labeled data by using unlabeled data, hence overcoming the inherent limits of labeled datasets.Utilizing unlabeled sensor data requires a method called En-Co-training that combines the co-training and ensemble learning paradigms. The approach employs an ensemble of many classifiers that have been trained repeatedly to predict the labels for a collection of unlabeled examples.A teacher model is first trained in the pipeline's teacher-student setup, which distills the knowledge of labeled data by labeling a sizable unlabeled dataset. To create a multi-task training dataset, high-confidence data points from the previous stage are chosen and then enhanced using signal processing. The ground truth labels from the training set are utilized to fine-tune the student model once it has been pre-trained to distinguish between signal modifications and the activities. The target dataset's unseen subset is then used for evaluation. SelfHAR and other HAR algorithms were compared for classification performance. Seven distinct labeled HAR datasets were compared using weighted F1 scores, and SelfHAR consistently beat the other methods. Overall, the findings indicated that SelfHAR outperformed supervised models in the majority of tasks across all datasets. The problem is that unlabeled datasets don't necessarily improve accuracy for all problems. It is only useful for certain problems. [7]

This alternate paper analyzed and compared several techniques to improve the robustness of a human activity recognition system by collecting data from smartwatches and smartphones used by various people. Cepstral-based technologies have reported to improve the feature extraction results. This paper also compared the features extracted using CNN with hand-crafted traditional features. CNN-MLP extracted good features even in a difficult environment. Only the time-based and frequency-based features are included in the baseline system's feature extraction module, and the random forest algorithm is used for classification. An improvement of 78.2% to 88.1% and 96.4% to 98.1% in accuracy was observed for smartwatch and smartphone data respectively. In the case of smartphones, modest gains were observed from the feature normalization techniques, but the same was ineffective for smartwatches due to its noisy data. [16]

Most of the methods present to recognize physical human activity do not take the height, weight, gender, etc., attributes into consideration. In this paper, a novel method is proposed which takes these factors into consideration and provides better accuracy in comparatively less time. Raw data is acquired from sensors such as the inbuilt accelerometer and gyroscope sensor modules of the smartphones. After the preprocessing of data collected, they are grouped based on the similarities between the subjects. A thorough analysis is done on the MotionSense(public) and custom dataset using the various machine learning algorithms. The proposed approach outperforms the traditional method with an accuracy of 99.8%. This approach trained the different models fast by having 110% more resource utilization compared to the traditional methods. This method managed to distinguish activities better than the traditional methods. [18]

Evidence has proved that regular recognition and monitoring of physical activity can reduce the risk of many diseases such as diabetes, obesity and cardiovascular disease. In this paper, the main goal is to investigate how gyroscope and accelerometer sensors impact human activity recognition and analysis using neural networks. Identifying activities such as walking upstairs and downstairs had an accuracy of 91% and 86% respectively. Unlike the above, the sitting activity had the lowest identification accuracy score of 72%. Whereas, lying down had the highest classification accuracy of 99%. Usage of smartphones has two main limitations: low battery capacity and low storage as for activity recognition to give the best results, continuous sensing from smartphones is required. [20]

This particular study [9] points out the importance of using appropriate cross validation methods to evaluate model performance for HAR applications. It begins by demonstrating that the commonly used validation procedure "k-folds cross-validation (k-CV)" is not a suitable method, especially in universal and hybrid HAR systems. Various algorithms are trained and tested with SHOAIB, WISDM and UCI-HAR datasets. Holdout, k-CV and leave-one-subject-out (LOSO) validation techniques are used to verify the stability of the test results. SHAP framework (Shapley additive explanations), based on game theory approach, is also used as a tool to visualize how dataset attributes are ranked by each cross validation strategy. In both holdout and k-CV strategies there occurs unnoticed overlap between train and test sets due to which model performance is overestimated. Thus, LOSO or any method that groups data by subject before splitting is a fairer evaluator, helping in building a more universal model. It is notable, however, that such methods require a lot of computational power and may limit applications of HAR systems to mobile/constrained devices or real time systems.

A separate study [10] instead chooses to highlight the advantages of dimensionality reduction techniques and checking for performance stability in HAR applications. Several univariate and multivariate methods are used for feature selection in five datasets. Following this, SMO and Random Forest are used for activity classification. It is observed that there is no significant degradation in performance with dimensionality reduction. 10-20% or even lesser of the dataset features are sufficient for accurate prediction. Relief based multivariate approaches, Chi Squared, and Information Gain gave the highest accuracies and also proved to be the most stable among all the algorithms tested. Although ranking based approaches for feature selection show good performance, they may be suboptimal in removing redundancy. The study also neglects hybrid approaches of feature selection.

3. Dataset Description

The dataset we used in this paper is from the UCI machine learning repository. It was built using recordings of 30 participants engaging in common tasks and postural changes while wearing a smartphone attached to their waist with inertial sensors. 30 people in the age range of 19 to 48 years participated in the trials. Six fundamental tasks were carried out by them: three static postures (laying, sitting, and standing) and three dynamic tasks (walking, walking downstairs and walking upstairs). Between the static postures in the trial, postural changes also took place. The person performed stand to sit, sit to stand, lie to sit, stand to lie, and lie to stand. Throughout the trial, each participant wore a smartphone around their waist. Using the device's integrated accelerometer and gyroscope, it recorded 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50 Hz.

This current dataset is an updated version of the previous dataset on UCI. Instead of the pre-processed signals from the smartphone sensors that were provided in version 1, this version offers the original unprocessed raw inertial signals from the sensors. This modification was made to allow for the creation of online tests using the raw data. Additionally, the activity labels have been changed to reflect postural transitions that were not included in the prior dataset version.

4. Proposed Workflow

This work consists of four modules: Data Acquisition, Pre-processing, Feature Extraction and Classification. The data acquisition module processes the sampling signals collected from the UCI HAR Dataset by wearable or external sensors such as smartphones or smartwatches and converts them into digital values that can be understood by the computer.

The preprocessing module transforms the raw data into well-formed datasets by reducing the noise from the data so that the data mining algorithms can produce accurate results. The feature extraction module selects the best features and removes the redundant data to improve the performance of the model.

The next and final module is the classification module where we use the data to classify it into activities such as standing, sitting, sleeping, and walking upstairs and downstairs.

From the literature survey, neural networks have proven to be the perfect algorithm for determining a person's physical activity. This is due to the fact that they can recognize the patterns behind the data and can be implemented easily. The classification module will classify the data into the activities using a deep learning architecture known as Bidirectional Long Short Term Memory (BiLSTM). The proposed method first processes the input from every sensor independently and then combines the information to classify appropriately.

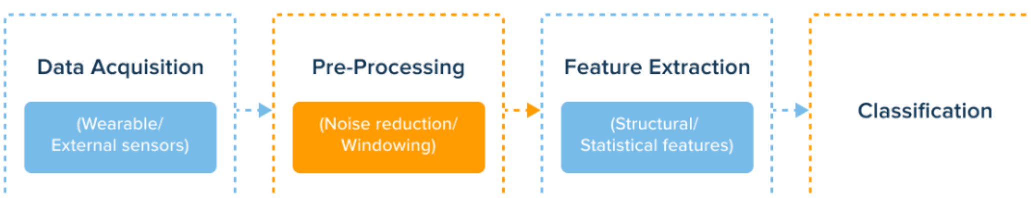

Figure 2. Proposed Workflow

5. Implementation Details

In this section, the environment we used for implementing the proposed system is discussed along with additional information on the experimental data used.

5.1 Experimental Setup

All work related to the data such as EDA, applying pre-processing techniques, performing feature extraction and building the model was executed on Google Colab. It is a free to use hosted Jupyter notebook service that provides free access to computer resources, including GPUs, and requires no setup to use. Colab is particularly well suited to machine learning, data analysis, and enables anyone to create and execute Python code through the browser.

5.2 Experimental Data

The dataset as described in section 3 has 30 subjects each performing 6 activities. The data was collected using both the accelerometer and gyroscope sensors present in a mobile phone. The resulting dataset was divided into two sets at random, with 30% of the volunteers chosen to create test data and 70% of the participants chosen to create training data. From the training set, data pertaining to one participant was extracted to be used as validation data.

5.3 Methodology

This section covers the implementation of the model which consists of the three main steps: (1) Data pre-processing, (2) Feature extraction and (3) classification.

5.3.1 Data Pre-Processing

To start off, we get the dataset from the UCI Machine Learning Repository. The data is found to be in a text file, so we load it into a dataframe which can be used throughout the tenure of processing the data. After fetching the data, we go on to getting them ready to be used later in the model. The dataset was checked for any duplicates or null values as a precautionary measure. They could significantly affect our results. The features of the dataset are then normalized in the range of [-1, 1] to achieve better results.

Now we decided to do sampling. In the dataset each of the six activities performed have a different number of datapoints. As seen in Figure 3, activities which have fewer datapoints might not be represented as well as the activities with more datapoints. So, we took 986 samples (the number of samples corresponding to the least well represented activity) from each activity so that they will be equally represented.

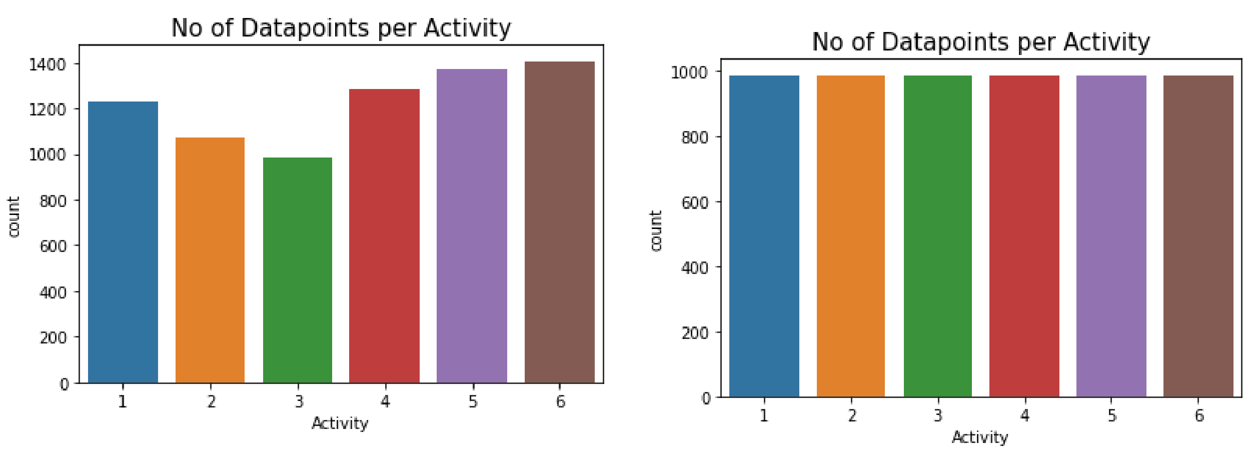

Figure 3. No. of datapoints per activity before and after sampling

5.3.2 Feature Extraction

For selecting the best features for the model, we use the fisher score method. Fisher score is a supervised filter-based method where the best features are selected based on the weight/score assigned to the features. This method was used due to its higher performance accuracy and reduced time-space complexity compared to wrapper or embedded methods. Using the fisher score method, the originally present 563 features are reduced to 280 features for the train, test and validation dataset.

5.3.3 Classification

The classification model is implemented using the BiLSTM algorithm. A Bidirectional Long Short-Term Memory OR BiLSTM is a sequence processing model consisting of 2 LSTMs: one takes input in forward direction while the other takes input in the backward direction in order to preserve the past and future information. BilSTM makes the model capable of learning local features and their long term dependencies in sequential data.

The model is trained over 25 epochs using Adam optimizer and categorical crossentropy loss function. The accuracy and loss of the training and validation dataset are measured to determine the performance of the model.

6. Results and Discussion

This section shows the results and evaluation metrics obtained from the BiLSTM model and provides a detailed discussion of those results.

The proposed model performed quite well with an overall accuracy of 81.30%. To evaluate the performance of the model, a classification report was generated as seen in Figure 4.

	precision	recall	f1-score	support
0	0.79	0.76	0.77	496
1	0.98	0.53	0.69	471
2	0.59	0.92	0.72	420
3	0.77	0.92	0.84	491
4	0.93	0.76	0.83	532
5	0.99	0.98	0.98	537
accuracy			0.81	2947
macro avg	0.84	0.81	0.81	2947
weighted avg	0.85	0.81	0.81	2947

Figure 4. Classification report

Both the accuracy and loss graph are shown below in Figure 5. In case of the loss graph, our model's loss exhibits step-like, repeating activity. It might be that our model observed recurrent behavior in the input data it received.

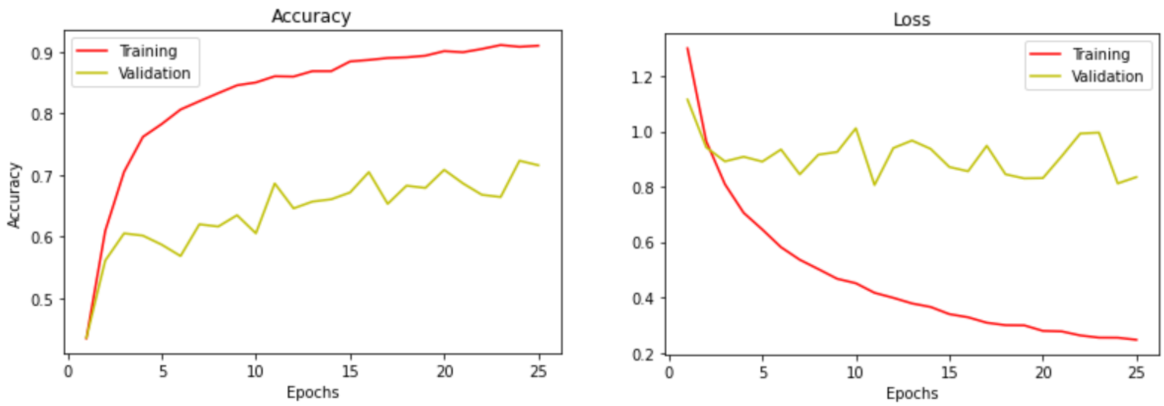

Figure 5. Accuracy and loss graphs

Now, looking at the confusion matrix in Figure 6, it is observed that the model manages to accurately predict class 5 (laying). But it looks like the model is confused between class 1(walking upstairs) and class 2(walking downstairs). Similarly, there is some confusion between class 4(standing) and class 3(sitting). In these cases, the model is unable to accurately differentiate between the classes.

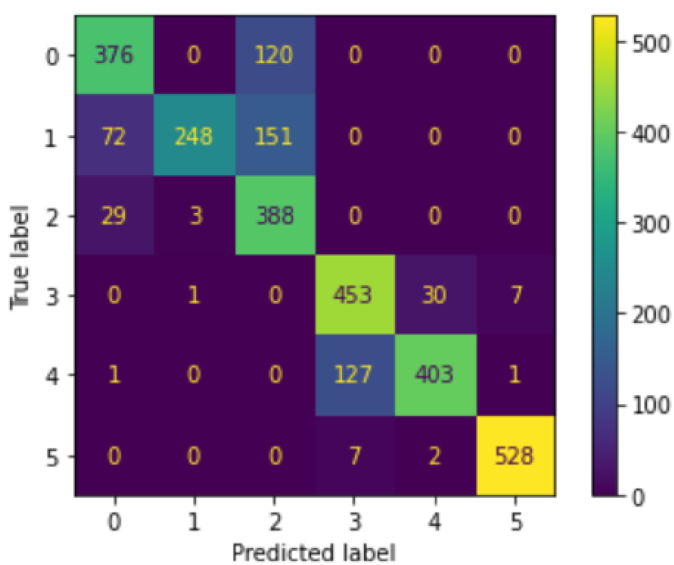

Figure 6. Confusion Matrix

Various other models from other works are compared against our model in Table 1 given below. We think that our results are reasonable given the challenge at hand, and can easily be improved through further experimentation with data preprocessing techniques.

Table 1: Comparison to related works

Model	Overall Accuracy
Our proposed model (BiLSTM)	81.3 %
SelfHAR [7]	91.35 %
MLP [20]	92%
LSTM [19]	93%
SVM (Rbf kernel) [1]	74.34 %

7. Conclusion

In this paper, we proposed a BiLSTM model to detect human activity using data from smartphone sensors. The data in general was challenging to work with as it consisted of 3 axis data from each of the two smartphone sensors. Through this paper we hope to show that human activity recognition is possible not only by using computer vision but by also using smartphone sensor data and predicting the activity. This problem with smartphone data is an interesting one and it can definitely be solved by better datasets. The datasets available right now are limited in number and another important factor is the placement of the smartphone during data retrieval. Once these problems are solved, models will most likely perform better. We also anticipate that future research will be able to clear up the confusions that the proposed model now causes among some of the chosen activities. This would make it possible to develop a system that could accurately identify someone's movements or actions, regardless of how they place their smartphone or any physical quirks they may have. The ability to track or anticipate a certain person's activities could be quite intriguing for many businesses around the world.

8. References

1. D. Garcia-Gonzalez, D. Rivero, E. Fernandez-Blanco, and M. R. Luaces, "A Public Domain Dataset for Real-Life Human Activity Recognition Using Smartphone Sensors," *Sensors*, vol. 20, no. 8, p. 2200, Apr. 2020, doi: 10.3390/s20082200.

2. R.-A. Voicu, C. Dobre, L. Bajenaru, and R.-I. Ciobanu, "Human Physical Activity Recognition Using Smartphone Sensors," *Sensors*, vol. 19, no. 3, p. 458, Jan. 2019, doi: 10.3390/s19030458.

3. M. M. Hassan, Md. Z. Uddin, A. Mohamed, and A. Almogren, "A robust human activity recognition system using smartphone sensors and deep learning," *Future Generation Computer Systems*, vol. 81, pp. 307–313, Apr. 2018, doi: 10.1016/j.future.2017.11.029.

4. N. Ahmed, J. I. Rafiq, and M. R. Islam, "Enhanced Human Activity Recognition Based on Smartphone Sensor Data Using Hybrid Feature Selection Model," *Sensors*, vol. 20, no. 1, p. 317, Jan. 2020, doi: 10.3390/s20010317.

5. A. R. Javed, R. Faheem, M. Asim, T. Baker, and M. O. Beg, "A smartphone sensors-based personalized human activity recognition system for sustainable smart cities," *Sustainable Cities and Society*, vol. 71, p. 102970, Aug. 2021, doi: 10.1016/j.scs.2021.102970.

6. Ankita, S. Rani, H. Babbar, S. Coleman, A. Singh, and H. M. Aljahdali, "An Efficient and Lightweight Deep Learning Model for Human Activity Recognition Using Smartphones," *Sensors*, vol. 21, no. 11, p. 3845, Jun. 2021, doi: 10.3390/s21113845.

7. C. I. Tang, I. Perez-Pozuelo, D. Spathis, S. Brage, N. Wareham, and C. Mascolo, "SelfHAR: Improving Human Activity Recognition through Self-training with Unlabeled Data," *arXiv:2102.06073 [cs]*, Feb. 2021, doi: 10.1145/3448112.

8. M. Straczkiewicz, P. James, and J.-P. Onnela, "A systematic review of smartphone-based human activity recognition methods for health research," *npj Digital Medicine*, vol. 4, no. 1, Oct. 2021, doi: 10.1038/s41746-021-00514-4.

9. H. Bragança, J. G. Colonna, H. A. B. F. Oliveira, and E. Souto, "How Validation Methodology Influences Human Activity Recognition Mobile Systems," *Sensors*, vol. 22, no. 6, p. 2360, Mar. 2022, doi: 10.3390/s22062360.

10. M. M. Manca, B. Pes, and D. Riboni, "Exploiting Feature Selection in Human Activity Recognition: Methodological Insights and Empirical Results Using Mobile Sensor Data," *IEEE Access*, vol. 10, pp. 64043–64058, 2022, doi: 10.1109/ACCESS.2022.3183228.

11. R. Presotto, G. Civitarese, and C. Bettini, "Semi-supervised and personalized federated activity recognition based on active learning and label propagation," *Personal and Ubiquitous Computing*, Jun. 2022, doi: 10.1007/s00779-022-01688-8.

12. M. Vakili and M. Rezaei, "Incremental Learning Techniques for Online Human Activity Recognition." Accessed: Oct. 17, 2022. [Online]. Available: https://arxiv.org/pdf/2109.09435.pdf

13. O. Nafea, W. Abdul, G. Muhammad, and M. Alsulaiman, "Sensor-Based Human Activity Recognition with Spatio-Temporal Deep Learning," *Sensors*, vol. 21, no. 6, p. 2141, Mar. 2021, doi: 10.3390/s21062141.

14. K. Peppas, A. C. Tsolakis, S. Krinidis, and D. Tzovaras, "Real-Time Physical Activity Recognition on Smart Mobile Devices Using Convolutional Neural Networks," *Applied Sciences*, vol. 10, no. 23, p. 8482, Nov. 2020, doi: 10.3390/app10238482.

15. M. Gil-Martín, R. San-Segundo, F. Fernández-Martínez, and J. Ferreiros-López, “Improving physical activity recognition using a new deep learning architecture and post-processing techniques,” *Engineering Applications of Artificial Intelligence*, vol. 92, p. 103679, Jun. 2020, doi: 10.1016/j.engappai.2020.103679.

16. R. San-Segundo, H. Blunck, J. Moreno-Pimentel, A. Stisen, and M. Gil-Martín, “Robust Human Activity Recognition using smartwatches and smartphones,” *Engineering Applications of Artificial Intelligence*, vol. 72, pp. 190–202, Jun. 2018, doi: 10.1016/j.engappai.2018.04.002.

17. D. Thakur and S. Biswas, “Feature fusion using deep learning for smartphone based human activity recognition,” *International Journal of Information Technology*, vol. 13, no. 4, pp. 1615–1624, Jun. 2021, doi: 10.1007/s41870-021-00719-6.

18. N. A. Choudhury, S. Moulik, and D. S. Roy, “Physique-based Human Activity Recognition using Ensemble Learning and Smartphone Sensors,” *IEEE Sensors Journal*, pp. 1–1, 2021, doi: 10.1109/jsen.2021.3077563.

19. M. Ullah, H. Ullah, S. D. Khan, and F. A. Cheikh, “Stacked Lstm Network for Human Activity Recognition Using Smartphone Data,” *IEEE Xplore*, Oct. 01, 2019. https://ieeexplore.ieee.org/document/8946180

20. G. Ogbuabor and R. La, “Human Activity Recognition for Healthcare using Smartphones,” *Proceedings of the 2018 10th International Conference on Machine Learning and Computing*, Feb. 2018, doi: 10.1145/3195106.3195157.