<https://www.tandfonline.com/doi/full/10.1080/24751839.2021.1987706>

* Pirttikangas et al. 2006 – pioneer research
  + “Tested a model that used several multilayer perceptron and k-nearest neighbor’s algorithms to recognize 17 activities to achieve an overall accuracy and 90.61%”
* Casale et al. 2011 –
  + “Used a wearable device and applied a random forest classification algorithm” (like in the penguin’s blog post) “to model five distinct activities (walking, climbing stairs, talking to a person, standing, and working on the computer).”
    - 90% accuracy
* Ahmed and Loutfi 2013 –
  + Using “case-based reasoning, support vector machines (SVMs) and neural networks (NN) to achieve an overall accuracy of 0.86, 0.62, and 0.59 respectively.”
* Brophy et al. (2018) –
  + “Proposed a hybrid convolutional neural network and an SVM model with an accuracy of 92.3% for four activities (walking and running on a treadmill, low and high resistance bike exercise).
* Ryoo et al. (2018) –
  + “Proposed a backscattering activity recognition network of tags, which comprises a network of passive RF tags capable of recognizing daily human activities with an average error of 6%.”
* Boukhechba et al. (2019) –
  + “a DL NN was developed for five activities (standing, walking, jogging, jumping, and sitting). The F1 score was 0.86.”
* Another publicly available data set – can download as CSV.
  + Raw data from sensors in RespiBAN and Empatica E4 devices
* “Most of the misclassified cases are related to the activity ‘zero’ which is the *transition activity*. There are also a few other misclassified cases, but most of them are related to activity zero. To improve the classification further, a variable window length has been employed and the *windows shift segmentation* approach is proposed in this paper.”
  + “The length of each segment cannot be a random value.”
  + “A window length of around 3s when the performed activity is “slow walking”.
  + “The cadence of an average person walking is within 90-130 steps/min and at least a full walking cycle (two steps) is preferred on each window.”
  + “Very small window size could produce several segments ‘without any activity at all’ if the performed activity is a soft one. One the other hand, big window size could produce segments with high overlapping unless a very high offset is selected.”

Human Activity Recognition Using K-Nearest Neighbor Machine Learning Algorithm by Saeed Mohsen, Ahmed Elkaseer, and Steffen G. Scholz

<https://www.researchgate.net/publication/220866565_Feature_Selection_and_Activity_Recognition_from_Wearable_Sensors>

* “This paper presents a k-nearest neighbor (KNN) algorithm for classification of human activities, namely Laying, Downstairs walking, Sitting, Upstairs walking, Standing, and Walking.”
* “The results show that the KNN algorithm provides a high performance in the classification of human activities. The weighted average precision, F1-score, and the area under the micro-average precision-recall curve for the KNN are 90.96%, 90.46%, 90.37%, and 96.5%, respectively, while the area under the ROC curve is 100%.”
* “In literature, a number of machine algorithms have been presented for classification of human activities”, but it is arguable that KNN algorithm has shown high potential to address the accuracy issue in HAR.
* “Implemented using Scikit-learn framework. It consists of five steps: the first step is pre-processing using filters to remove the noise from the utilized dataset, the second step is a feature extraction of the dataset… The third step is the selection of a specific feature, the fourth step is the training on the dataset, and the fifth step is the testing/classification for the activities.”
  + As part of step 2:
    - “561 features are extracted with fixed windows of 2.56s.”
    - “The dataset features are ax, ay, and az, which represent the accelerations in the x-axis, y-axis, z-axis respectively.”
    - “Also, another extracted feature of the data set are gx, gy­, and gz, which represent the gyroscope angles.”
* A mechanism of the KNN algorithm:
  + “It assumes two categories or classes “A and B” and a new data point *x*1 which colored blue. So, the KNN algorithm can detect the data point in which category will be lied. This algorithm uses neighbors’ number of 20 and it is based on the Euclidean distance (*dE*), which is computed between a two categories “points”, where *dE* is represented in [the equation]: ”.
  + “In this algorithm, a neighborhood function is used to classify the activities. The best selection of k depends on the dataset. So, the largest k reduces the noise applied to a classification.”
  + They also used the UCI-HAR dataset.

Diagram

Description automatically generated

<https://www.researchgate.net/figure/3-axes-accelerometer-readings-sitting-laying-down-standing-standing-up-and-walking-a_fig2_348163122>

Chart, box and whisker chart

Description automatically generated

1. Sitting
2. Laying down
3. Standing
4. Standing up and walking