# Introduction

The traditional user authentication methods (passwords, fingerprints, face locks, passcodes received by text or email) have many drawbacks such as time-consuming, only at entry-points, and more importantly, susceptible to attacks including side-channel, shoulder surfing, social engineering, etc.
As a result, there's an increasing need to build an continuous user authentication system as an additional support to improve the security.

This project is mainly based on research "An empirical evaluation of activities and classifiers for user identification on smartphones" by C. Tang and V. V. Phoha. They tried 10 different classifiers to classify users based on accelerometer and gyroscope data captured by mobile devices.
They claimed that k-NN produced a consistent and outstanding performance (almost 100% accuracy) under dynamic behaviors (walking, upstairs, downstairs), static behaviors (sitting, standing, lying), transitions between static behaviors, and under unlabeled activities.

## Deficiencies of their work

1. The data set (#users, #samples especially in transition) isn't big enough to make sure that this system can be deployed in real-world scenarios where there might be thousands of users, some of which might have very similar patterns.

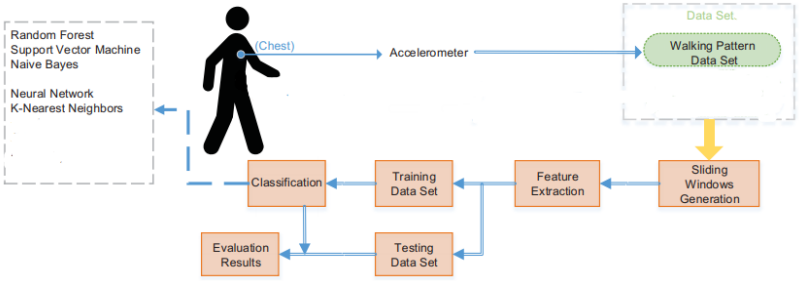
2. Although K-NN gives an outstanding performance on these datasets, the time complexity of computing could be insanely high. We might need to try approximation algorithms for k-NN.

3. There could me many other scenarios that this research didn't consider: put on a table, carried on a vehicle, hold by hands, etc., under which (especially on a table) the k-NN classifier would probably not work. We might need to try combining this system with touch-based systems, etc.

4. This research simply uses default or very simple parameters for models, which could influence a lot on their performances. Tuning is an art.

## Our Aim

In our project, we focus on deficiencies 2 (k-NN time complexity) and 4 (parameter tuning) of their research and try to improve their system by using k-NN approximation algorithms and tuning the parameters on Walking Pattern Data Set from UCI Machine Learning Repository. Our design is similar to theirs.




# Data Preprocessing

The Walking Pattern Data Set from UCI Machine Learning Repository contains accelerometer time series data from 22 users. It's easy to tell that there's an obvious discrepancy between different users' walking patterns from our first glance at the following x-axis accelerometer data.

We partition the raw data into fixed-width (100 samples) sliding windows with a 50% overlap, and then extract 9 features (the mean values, standard deviations, and median absolute deviations of x, y, and z axis) from every window.

We remove users 3, 5, 16, & 19 as they have too few samples (csv file < 5kb) from our data set.

From the correlation matrix we conclude that our feature extraction is appropriate.

## Model

In the study from Tang et al., 10 classification models (Random Forest, SVM, Naïve Bayes, J48, Neural Network, k-NN, RPart, JRip, Bagging, and AdaBoost) were used. However, the parameters were set arbitrarily or simply by default.

In this project, we try *Logistic Regression, Decision Trees, Discriminant Analysis, Naïve Bayes, SVM, and k-NN (without or with approximation)* with parameter tuning.

## Evaluating Results

To compare the performance of different classifiers, we calculate the accuracy for each model.

Also, to check if the classifier is sensitive and specific, we calculate the sensitivity and specificity for each user for each model.

## Future Work

Other researches by Phoha et al. has considered phone usage contexts by training context identification model and combing continuous phone movement pattern model with touch-based models, etc.

With the increasing demand to try much larger data sets to see if the model still works consistently, however, there's no large enough dataset available publicly. We might need to do a larger experiment by advertising the app collecting movement patterns publicly to collect sufficient data for study.

Although we started parameter tuning and k-NN approximation, there's still a long way to go on this path.

## References

[1] A. Serwadda, V. V. Phoha and Z. Wang, "Which verifiers work?: A benchmark evaluation of touch-based authentication algorithms," 2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS), Arlington, VA, 2013, pp. 1-8.

[2] C. Tang and V. V. Phoha, "An empirical evaluation of activities and classifiers for user identification on smartphones," 2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS), Niagara Falls, NY, 2016, pp. 1-8.

[3] Kumar, Rajesh, Partha Pratim Kundu, Diksha Shukla, and Vir V. Phoha. "Continuous user authentication via unlabeled phone movement patterns." In 2017 IEEE International Joint Conference on Biometrics (IJCB), pp. 177-184. IEEE, 2017.

[4] Google and Wikipedia

[5] Walking Activity Data Set from UCI Machine Learning Repository

https://archive.ics.uci.edu/ml/datasets/User+Identification+From+Walking+Activity