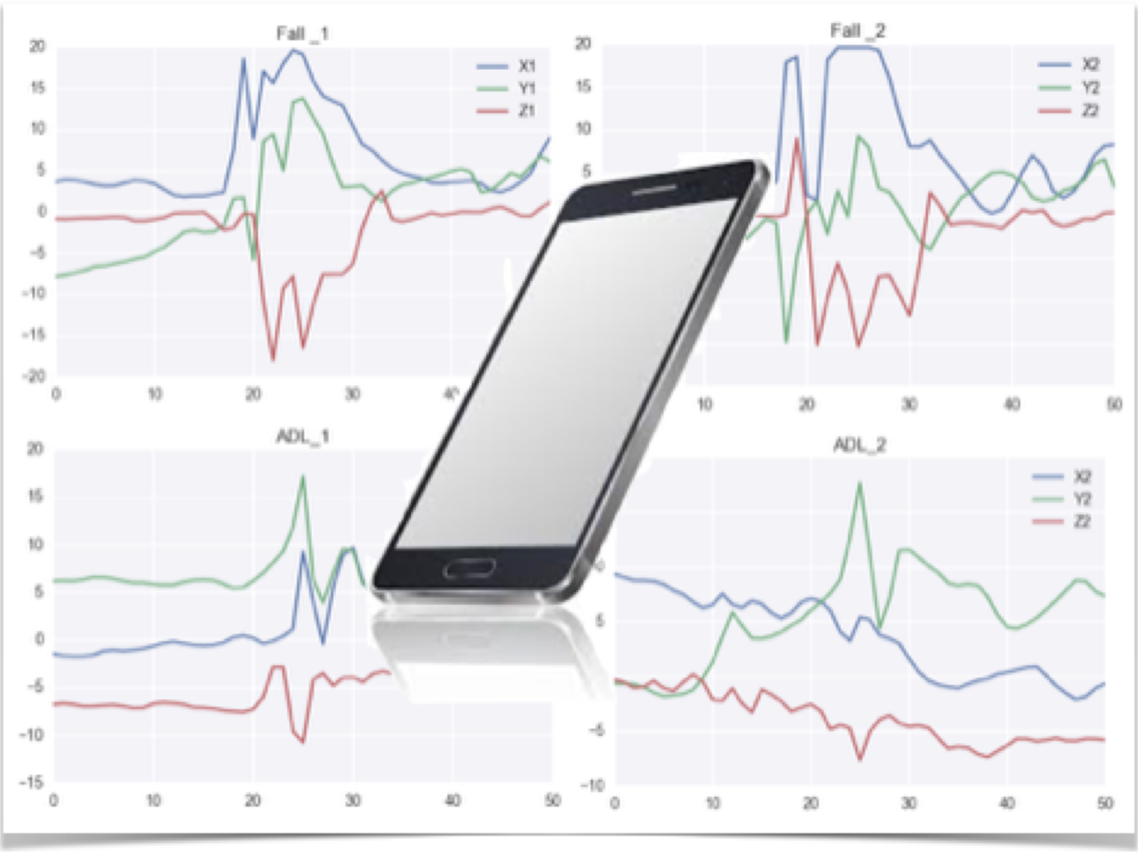
ASSISTIVE TECHNOLOGIES

Smartphone-based Fall Detection System



Capstone project – Milestone Report

Prepared by: Anna Zamaraeva

27 June 2017

Proposal number: 000-1234

Table of Contents

EXECUTIVE SUMMARY

Overview

Falls are a major health risk that impacts the quality of life of many people. Young people engaging in extreme leisure activities, disabled persons and elderly people, patients recovering from injuries - are all subject to a higher risk of accidental falls. When a fall occurs, a prompt notification would help to provide quick help and reduce potential injuries.

With modern technologies, a smartphone can serve as a detection system for the falls. Smartphones are wide spread, do not introduce any additional costs, can be used in any place and are accepted by people as a part of everyday life. Smartphones have inbuilt sensors: an accelerometer and a gyroscope, data from which can be used for falls detection.

An effective smart-phone based system can provide multiple benefits: (1) automatic notification of occurred falls; (2) promptness in order to provide quick assistance and (3) communication capabilities in order to alert the caregivers. The system can be further evolved into a commercial smartphone app which can be used in several application domains: assisted living settings, care-providing and by general users.

Goal

The goal of the project is to train a machine-learning based algorithm to classify human activities and to detect the falls using patterns recorded by smartphone accelerometers.

Results and Recommendations

Results and recommendations.

PROJECT DATA

Human Activities Accelerometer Datasets

Recently several accelerometer datasets for human activities have been collected by researchers worldwide and made publicly available. These datasets can be broadly divided by two criteria: a type of sensor used in data acquisition (a smartphone or a specialised device) and inclusion of falls into a dataset.

For the purpose of the project, the datasets were evaluated in terms of the following requirements: (a) accelerometer data captured by smartphones located in the pockets of study participants, rather than by specialised device; (b) a wide range of human activities and a rich variety of study participants; and (c) a large number of simulated falls in a dataset.

After the analysis, UniMiB SHAR - A New Dataset for Human Activity Recognition Using Acceleration Data from Smartphones, was selected for the project. This dataset was made

available by researchers of University of Milano-Bicohcca at
http://www.sal.disco.unimib.it/technologies/unimib-shar/
The dataset is a labelled, rich and complete collection of acceleration patterns, and serves as a good base for conducting data experiments.

Overview of the project dataset

The UniMiB SHAR dataset contains triaxial acceleration data captured by smartphones during a wide range of activities. Recorded daily activities and simulated falls are performed by a large number of subjects varying in age, gender and physical characteristics. UniMiB SHAR dataset contains total 7,013 activities samples performed by 30 subjects, mostly females (24), of ages ranging from 18 to 60 years.

The dataset is composed of

- 5,314 samples of activities of daily living (ADL): walking, jumping, running , sitting down, standing up, lying down, getting up, going up/down;
- 1,699 simulated falls of 8 different types.

The dataset is imbalanced with falls being a minority class and ADL is a majority class. So this might require rebalancing techniques.

{ToDo} How the data was collected and what preprocessing applied….

The samples are a timeseries of triaxial accelerometer readings. Each timeseries is 1 sec long sampled with 50Hz frequency. This gives 51 observations of acceleration in 3 dimensions (total 153 values).

Figures 1 and 2 demonstrate sample data for various daily activities and fall types.

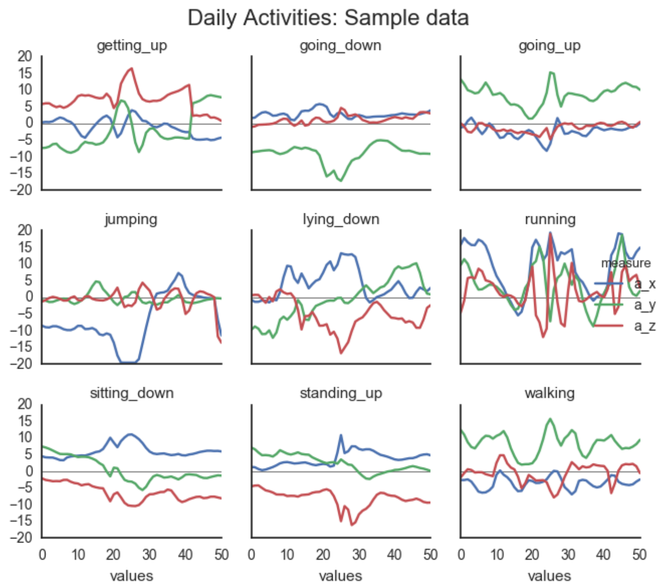


Figure 1: Daily Activities - Sample data

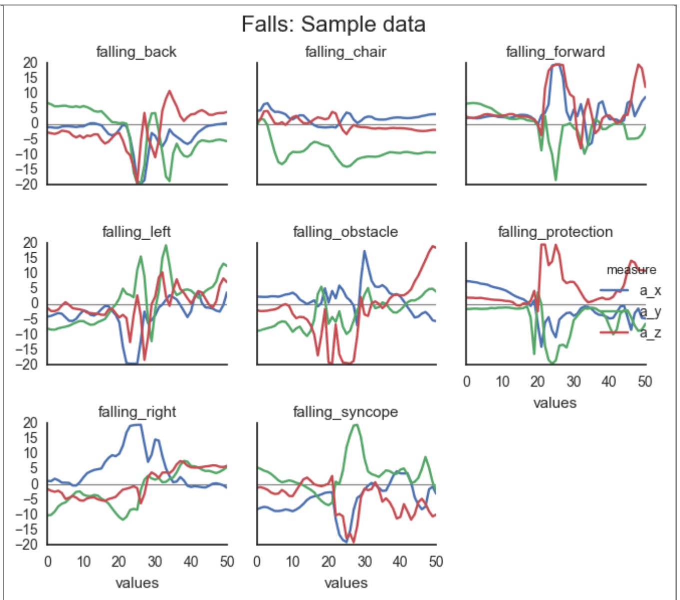


Figure 2: Falls - Sample Data

Extracted Feature Sets

As described above, the raw data is a time series of acceleration projections into 3 dimensions against the fixed walls of a smartphone:

$$Triax\_Accel(t_i) = \{a_x(t_i), a_y(t_i), a_z(t_i)\}, \quad \text{where } i = (0, 1, .., 50)$$

Intuitively one would think that the acceleration magnitude can be a significant feature. So we created a time series of acceleration magnitude, as

$$Magn(t_i) = \sqrt{a_x(t_i)^2 + a_y(t_i)^2 + a_z(t_i)^2}, \text{ where } i = (0, 1, .., 50)$$

Also, given that we are looking for some sharp changes in the movement, we calculated a jerk measure. In physics, jerk, also known as jolt, or surge, is the rate of change of acceleration; that is, the derivative of acceleration with respect to time.

So we calculated a triaxial jerk using acceleration projections, as

$$Jerk_k(t_i) = a_k(t_i) - a_k(t_{i-1}), \quad \text{where } k = (x, y, z)$$

And also a jerk in total acceleration, as

$$Jerk_{Magn}(t_i) = Magn(t_i) - Magn(t_{i-1})$$

We also calculated summary statistics for acceleration and jerk timeseries, using mean, standard deviation, minimum and maximum of the timeseries.

This gave us 4 feature sets to work with. Table 1 below is summarising the sets, ordered by a dimensionality of the feature space from the smallest to the largest.

Table 1: ***Feature Sets Summary***

	Set name	Description	N feat
1	Magnitude_Stats	Summary statistics for magnitudes of acceleration and jerk	8
2	TriaxAccel_Stats	Summary statistics for triaxial acceleration and jerk	24
3	Magnitude_Timeseries	Time series of acceleration magnitude	51
4	TriaxAccel_Timeseries	Time series of triaxial accelaration (raw data)	153

EXPLORATORY DATA ANALYSIS

Clustering of data

To build an understanding of data, we built an hierarchical clustering using varios feature sets. Each activity is represented by its mean sample, serving as a class representative. Figure 3 shows the resulting dendrograms.

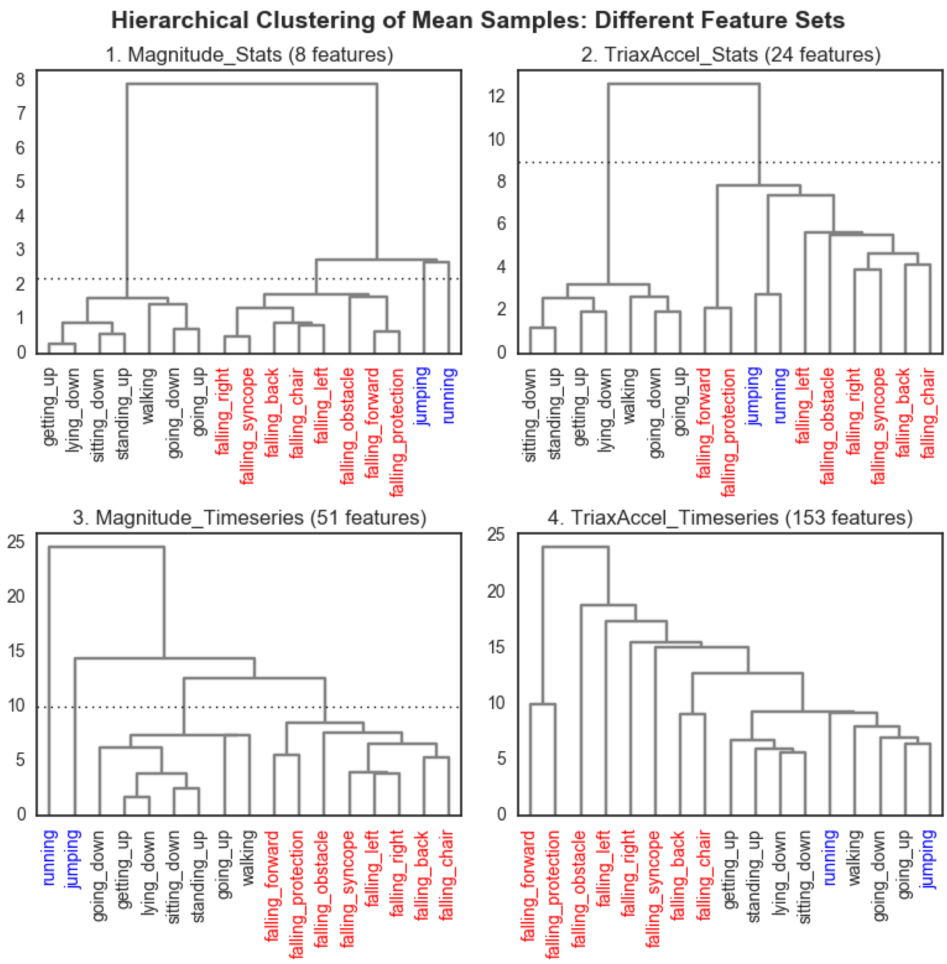


Figure 3: Clustering of activities data

{To Do} Discussion of dendrograms... Running and Jumping as separate classes; Falls and daily activities grouped together; Triaxial Acceleration set has falls recombining only at the top, probably because of high dimensionality

Separability in PCA Space

Based on the dendrograms, we look into separating the data into 4 categories: Running, Jumping, Daily Activities, Falls.

Again based on each feature set, we project the data into first two principal components and plot the corresponding groups in Figure 4.

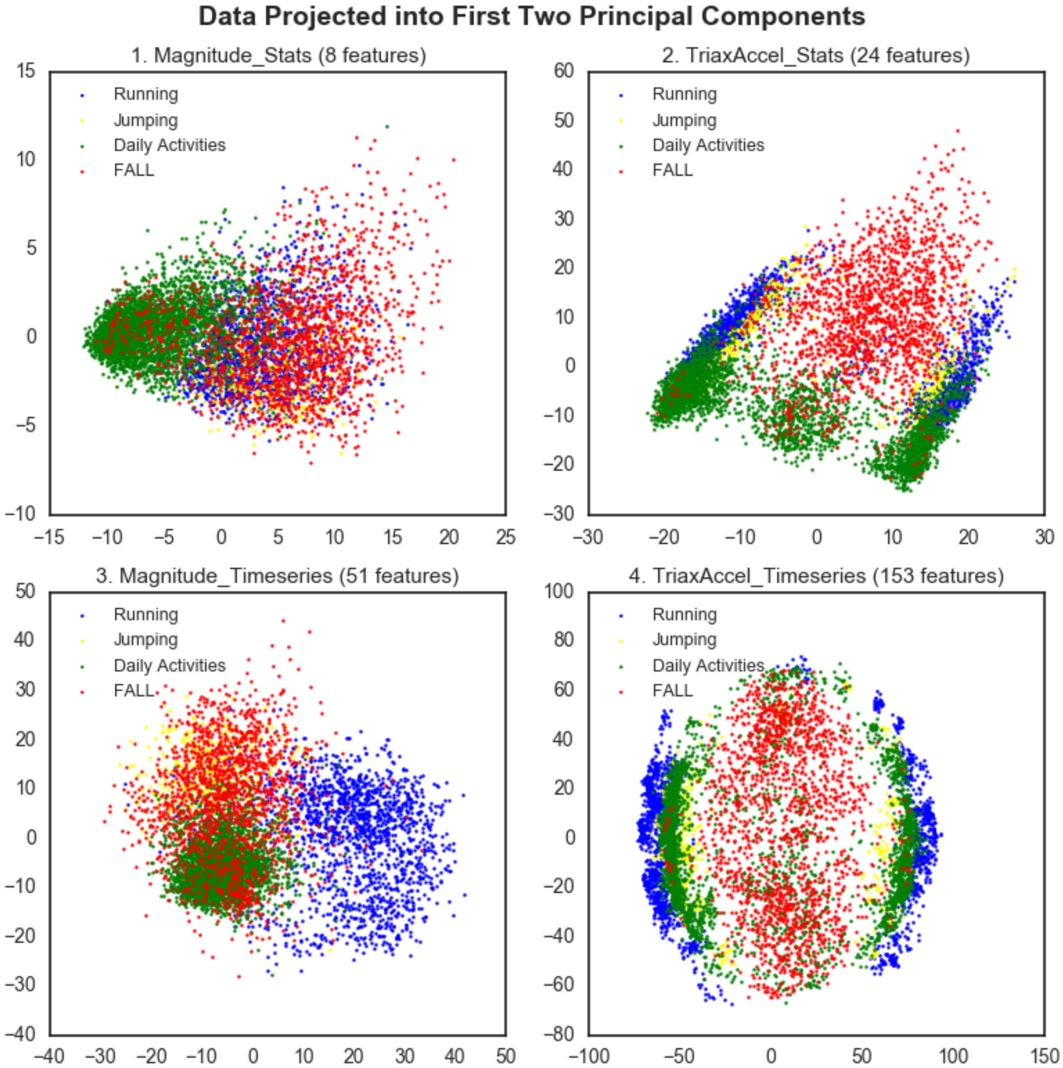


Figure 4: Separability of Classes in PCA space

In Triaxial sets (both stats and timeseries) – there is an interesting symmetrical pattern, which probably means that the direction of the projection does not really matter..

Two Class Classifiers: Daily Activities vs Falls

Linear models

We use Logistic regression as a work horse to get an understanding. This can be built on to use other classifiers: kNN, SVM, Random Forest.

We do 3 types :

a) using all activities
b) excluding jumps and running (this is for a care home)
c) rebalance the data by oversampling falls (duplicate fall entries)

Figure 5 gives ROC curves for each feature set based on 3 data types.

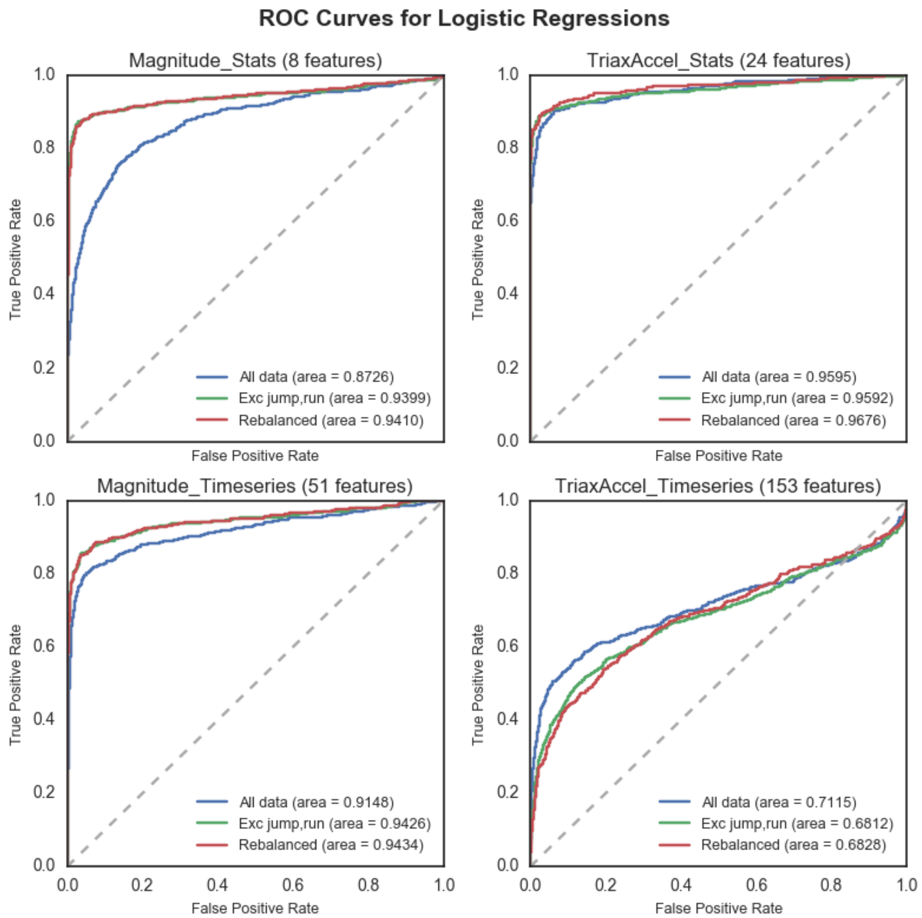


Figure 5: ROC Curves Using Different Feature Sets

The best is Triaxial Stats. Triaxial Timeseries (raw data) is very poor by some reason…

Model perforamce is in the Table 2 below

Table 2: ***Model Performance for Logistic Regression Classifiers***

	Feature Set	Data Type	Accuracy	PoD	PFA	AUC
1	**Magnitude_Stats**	a) All activities	85.9%	**60.5%**	5.9%	0.8726
	(8 Features)	b) Exc jump, run	93.5%	**87.5%**	3.4%	0.9399
		c) Exc jump, run + Rebalanced	91.4%	**89.4%**	7.6%	0.9410
2	**TriaxAccel_Stats**	a) All activities	94.3%	**84.0%**	2.4%	0.9595
	(24 Features)	b) Exc jump, run	93.9%	**89.3%**	3.8%	0.9592
		c) Exc jump, run + Rebalanced	93.2%	**90.8%**	5.6%	0.9676
3	**Magnitude_Timeseries**	a) All activities	91.9%	**76.6%**	3.1%	0.9148
	(51 Features)	b) Exc jump, run	92.3%	**83.8%**	3.4%	0.9426
		c) Exc jump, run + Rebalanced	91.1%	**87.2%**	6.9%	0.9434
4	**TriaxAccel_Timeseries**	a) All activities	82.6%	**34.2%**	1.6%	0.7115
	(153 Features)	b) Exc jump, run	75.1%	**43.5%**	8.7%	0.6812
		c) Exc jump, run + Rebalanced	66.9%	**62.2%**	30.8%	0.6828

PoD = Probability of Detection; PFA = Probability of False Alarm

Interestingly, rebalancing by oversampling the falls increases the probability of detection, but makes the False Alarm rate to go up..

And these are best classifiers

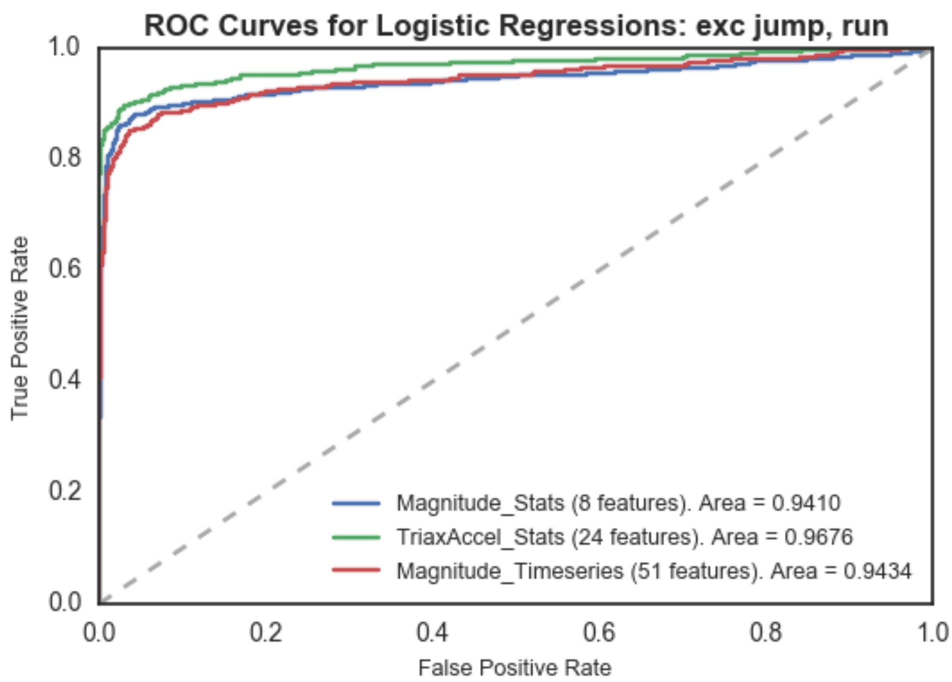


Figure 6: ROC Curves, data excl jumping/running and oversampling falls