Application of Machine learning algorithm in predicting force from EMG and Classification of different type of Grip Movement

Hasan E. Issa
University of applied science upper Austria
Garnisonstrasse 21 Linz Upper Austria /Austria
Hasan.issa@students.fh-linz.at

ABSTRACT

Relationship between daily routine Repetitive movement and repetitive strain injury (RSI) has been considered a major topic in ergonomic.
worker who may have exposure to undue physical stress strain and overexertion including vibration, awkward posture, repetitive movement could develop RSI dramatically.
Machine learning Algorithms of EMG data is a significant method to extract the important features of the EMG signal to build a statistical model can predict the exerted force by a muscle and to be able to classify different type of grip movement.
In this study we tried to build a uniform model can predict the force from collected EMG data only, and to classify three different types of grip movement. Reaching an accurate model which can implemented in an embedded EMG modality could accomplish many applications in ergonomics and Machine human interface.

KEY WORDS
Machine learning, sEMG, linear regression

1 Introduction

According to the U.S. Department of Labor, Occupational Safety and Health Administration (OSHA), repetitive strain injuries (RSI) are the nation's most common and costly occupational health problem, affecting hundreds of thousands of American workers, and costing more than $20 billion a year in workers' compensation. According to the U.S. Bureau of Labor Statistics, nearly two-thirds of all occupational illnesses reported, were caused by exposure to repeated trauma to workers' upper body (the wrist, elbow or shoulder). One common example of such an injury is carpal tunnel syndrome. The main problem with RSI that it can appear with occupations involving heavy labors and with computer users and typist. A simple calculation can show that if you type 40 words a minute you press 12000 keys per hour or 96,000 keys per eight-hour day, almost of 16 tons of force will be exercise by your finger. Presently, 25% of all computer operators have Carpal Tunnel Syndrome, with estimates that by the year 2000, 50% of the entire workforce may be affected. Only 23% of all Carpal Tunnel Syndrome patients were able to return to their previous professions following surgery. Up to 36% of all Carpal Tunnel Syndrome patients require unlimited medical treatment [1].one of the possible way to minimize the effect of RSI is to know the minimum force required to perform a certain task, such approximation can be done using a force sensor but there is a lot of limitation with using force sensor directly since such sensors need to be connected directly to the palm of the hand or need a special glove and this come with loss of real measurement scenario, another approach is to use the EMG data to measure the force by depending by the relationship between force and EMG [2].

2 Material and Method

Data acquisition and measurements are done by Richard Schmidt as part of his master thesis, this study focus in further data processing and algorithm development.

2.1 Subject

Five right-handed healthy males between 24 and 28 years with a mean of 26 years were recorded. All subjects agreed voluntarily to the participation and the publication of the recorded data.

2.2 Grip Type

Three basic everyday life grips were selected for the study: precision, force and pinch grip. Figure 1 shows the three-different grip movement.

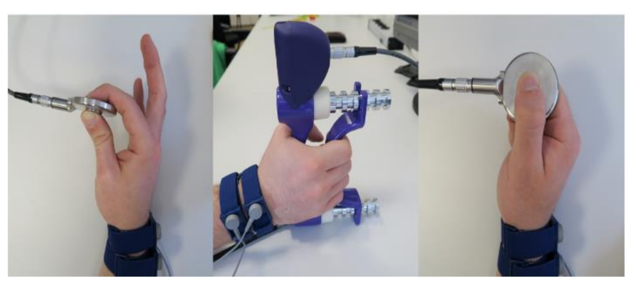

Figure 1 Grip type from left to right: Precision, force, Pinch

2.3 Paradigm

The records consist of three identical blocks with same structure and sequence of the measurements. The only difference between each block is the grip type (pinch, force, precession), first block is the pinch grip, second is the force and third block is the precession grip, instruction for subject took around half hour before the measurement, each block took almost one hour and between each block there was a break for half hour which make the whole measurement for one subject take four hours and half.
Each block consists of 31 tasks. At the beginning and at the end of each block four calibration tasks take place. The calibration tasks were maximum force tasks for every grip type and in addition one task with resisted maneuvers of the forearm in neutral position. Table 1 shows an overview of force and frequency tasks performed by each subject. Figure 2 shows the blocks diagram the arrows at the beginning and at the end of the block indicate the calibration tasks.

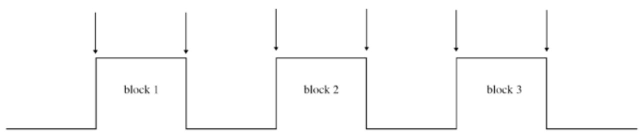

Figure 2 Paradigm Blocks

MVC	Repetition Frequency in Hz				
5 %	0.1	0.25	0.5	1	2
10%	0.1	0.25	0.5	1	2
20%	0.1	0.25	0.5	1	
35%	0.1	0.25	0.5	1	
50%	0.1	0.25	0.5		
66%	0.1	0.25			

Table 1 paradigm measurement

2.4 EMG Channels Configuration

To record all EMG signals the whole measurement is repeated with two different EMG electrodes configurations. The first one is six radial distributed EMG electrodes on the forearm. The second configuration are 4 electrodes in special position on the forearm. Table 2 shows the channel configuration and placement of the electrodes.

Configuration	Channel	Sensor type	Position
First Configuration	Ch1	Force sensor (G200, P200	Hand (grip type dependant
	Ch2 to Ch7	EMG electrode (SX230)	Clockwise uniformly distributed
	Ch8	-	-
	D1	EMG ground (R 506)	wrist
	D2	Synchronization cable	
Second Configuration	Ch1	Force sensor (G200, P200	Hand (grip type dependant distributed
	Ch2	EMG electrode (SX230)	extensor capriulnars
	Ch3	EMG electrode (SX230)	Extensor digitorum
	Ch4	EMG electrode (SX230)	Flexor digitorum profundus
	Ch5	EMG electrode (SX230)	Flexor pollicis-longus

Table 2 Channel Configurations and placement of electrodes

2.5 Data Processing

The recorded raw data are in (‘txt’) format, each subject has 93 file which make the total numbers 465 files for fives subjects. To facilitate the access of the raw data a strict naming method was performed. Each file name consists of six sections which are separated by an underscore (‘_’) as follow:
Subject_Block_Grip_MVC_frequency for example the file (02_03_pinch_20_025) can be interpreted as:
Second Subject _Third Block_ Pinch Movement _20% MVC _0.25Hz.
the first thirteen lines in each text files are header for the file which contain the information about the EMG channel and measuring units.
The data values start at line 14 and the data are separated by a comma in 11 columns, first columns for the force sensor values and the remaining 10 are for EMG channels values (6 for the first configuration and 4 for the second configuration).
the first step is to automate the selection of data files and since further processing could be done to any group of files (subject, MVC level, …etc.) or any combination of the previous sections, function was implemented to group the data files as per the user input,

Pandas which is a module in python 3.0 facilitate this process by creating a data Frame for the file name where each column of the data frame represent a unique scenario, for example below are the data frame columns

{'MVC_20', 'Freq_100', 'Fifth_Subject', 'Freq_25', 'MVC_10', 'MVC_50', 'Forth_Subject', 'First_Subject', 'Freq_200', 'Freq_10', 'Precesion_Movment', 'Third_Subject', 'MVC_66', 'MVC_Max', 'MVC_35', 'Freq_50', 'Force_Movment', 'MVC_5', 'Pinch_Movment', 'Second_Subject'}

The function will match the user input from the keyboard to the names of the columns in the data frame for example the user will enter MVC_20 and Second_Subject and Freq_10 and Pinch_Movment, the function will return all data file of the Second Subject Performing Pinch Movement with 20% MVC and with repetition frequency of 0.1 Hz.

2.6 Preprocess

Moving Root mean square (RMS) widow with size of 197 was used to filter the raw data, applying the RMS has been used to process EMG in much the same way as applying a linear envelope. By squaring the raw signal, you effectively transform it into a zero-mean signal. Then the mean (and square root) are applied using a moving time window, the length of which the user can define. The selected length of the window (i.e., the time constant) will impact how "smoothed" the signal becomes, Figure 3 shows the raw force and EMG data while Figure 4 shows the force and EMG data after applying the RMS filter.

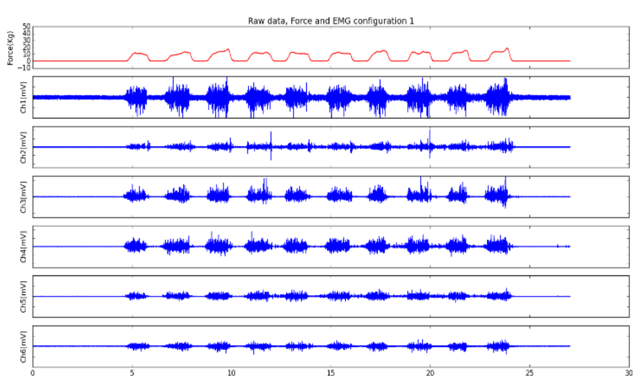

Figure 3 Raw EMG data

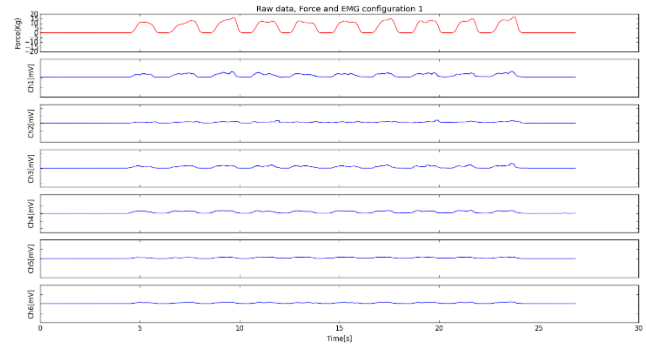

Figure 4 Force and EMG data after applying RMS filter

2.7 Linear Regression for a single channel

Linear Regression is a very simple approach for supervised learning. It is a useful tool for predicting a quantitative response [3], our first model was to build the linear regression model where the data from the force sensor is the response and the data from one electrode is the predictor. Equation [1] represent the single linear regression model

$$Force = \beta 0 + \beta 1 * EMGs \quad (1)$$

In such model the main concern will be which EMG electrode is involved more in the grip movement, since each grip movement involve the activation of one or more muscle higher than others, the best way to determine which EMGs electrode data contribute more to the movement, is by calculating the correlation coefficient for each sensor data with force sensor data. Table 3 shows the correlation coefficient for Grip Force Movement with 10% of MVC and repetition frequency of 0.5 Hz. The Correlation coefficient between two variables answers the question are the two variables related? Or if one variable changes, does the other also changes? [4] Good correlation coefficient is usually near 1.

Selecting the EMGs Electrode now is easy by choosing the highest correlation to the force data. In this case, we can see the EMG radial 6 and EMG radial 1 is relatively higher than other so we can build our model using its data as a predictor.

Sensor	Correlation coefficient
Force sensor	1.000000
EMG_radial_1	0.905168
EMG_radial_2	0.890762
EMG_radial_3	0.860325
EMG_radial_4	0.839110
EMG_radial_5	0.877085
EMG_radial_6	0.926253
EMG_special_1	0.912436
EMG_special_2	0.902278

EMG_special_3	0.865208
EMG_special_4	0.864327

Table 4 Correlation Coefficient for Grip Force

Figure 5 shows the best -fit linear regression line to a single channel.

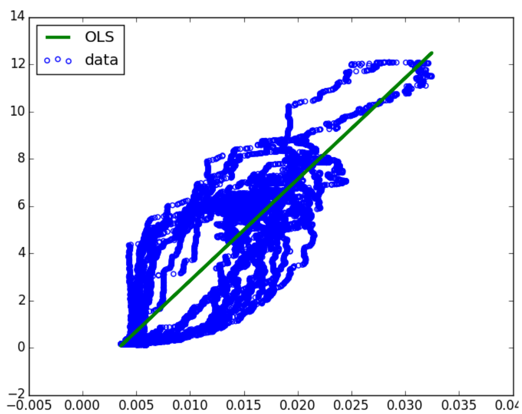

Figure 5 best -fit linear regression for single EMG channel

The model fit is done by the statsmodels module in python3 which gives a wealth of information about the model, our model summary as below

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           Force_sensor   R-squared:                       0.819
Model:                            OLS   Adj. R-squared:                  0.819
Method:                 Least Squares   F-statistic:                 1.209e+05
Date:                Tue, 14 Mar 2017   Prob (F-statistic):               0.00
Time:                        09:27:14   Log-Likelihood:                -44830.
No. Observations:               26671   AIC:                         8.966e+04
Df Residuals:                   26669   BIC:                         8.968e+04
Df Model:                           1
Covariance Type:            nonrobust
================================================================================
                 coef    std err          t      P>|t|      [95.0% Conf. Int.]
--------------------------------------------------------------------------------
const         -1.4533      0.015    -98.892      0.000        -1.482    -1.425
EMG_radial_1  429.0483     1.234    347.767      0.000       426.630   431.466
==============================================================================
Omnibus:                      995.001   Durbin-Watson:                   0.001
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1669.214
Skew:                           0.327   Prob(JB):                         0.00
Kurtosis:                       4.037   Cond. No.                         155.
==============================================================================
```

With our model summary, we can check the Null Hypothesis which is β0 = β1 = 0 in equation (), that can be interpreted as there is no relationship between Force and EMG channel, we can look to the P value corresponding to the F statistics, if the P values is low enough we can consider there is a relationship between the Force and the EMG Channels. And here we can reject our null hypothesis. In our model, we can see that P value is zero so definitely there is a relationship between Force data and EMG data.

After confirming the relationship between the variables now we can discuss how strong is the relationship between the Force and the EMG data?

It's a question of the Model accuracy and it can be done by R^2 statistics which records the percentage of variability in the response that is explained by the predictor, R^2 values which is close to 1 indicate that big portion of the data was explained by the linear regression model [3]. In our model the R^2 value is 0.819. another important parameter is to plot the residuals, studentized residual and the leverage. Figure 6 shows the three plots.

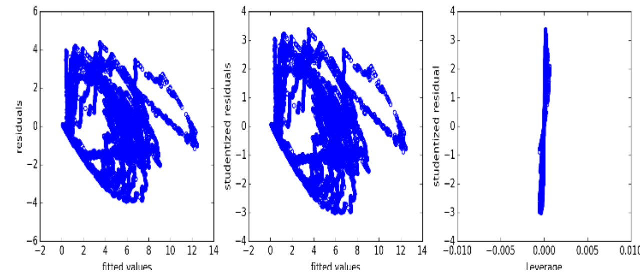

Figure 6 residual, studentized residuals and leverage

A good model will show no pattern of residuals. usually a U-shaped residual can indicate to non-linearity problem, while the studentized residuals, computed by dividing each residual by its estimated standard error. Observations whose studentized residuals are greater than 3 in absolute value are possible outliers. Hence we can see that most of our data lies between -3 and +3 which indicate low number of outliers, observations with high leverage usually have an unusual value for x values, like studentized residuals good model will show leverage between -3 and +3.

2.8 Multiple Linear Regression

In multiple linear regression, we use all EMGs channels data as predictors and the force as response, equation [2] represent the multiple linear regression model

$$\text{Force} = \beta 0 + \sum_{i=0}^{n} \beta \text{i} * \text{EMGi} \ (2)$$

2.9 Force Regression

The main problem with the previous linear regression models that we are trying to predict the Force values from the EMG data in each point of time, with long measurement like our case, the model will not fit the data perfectly because of the existence of outliers and high leverage points, since our measurement include resting time between activation the proposed model is to perform the regression only during contraction which will eliminate the effect of the outliers and will reduce the huge amount of data to a acceptable range which can be computed fast enough, below are the steps to build the force regression model :

- Apply root mean square filter to the raw data

Separate the contractions from the force signal by choosing a certain threshold and delete all values point below this threshold. We chose the threshold 30 % of maximum contraction, Figure 7 shows the force data after removing all data points below the threshold

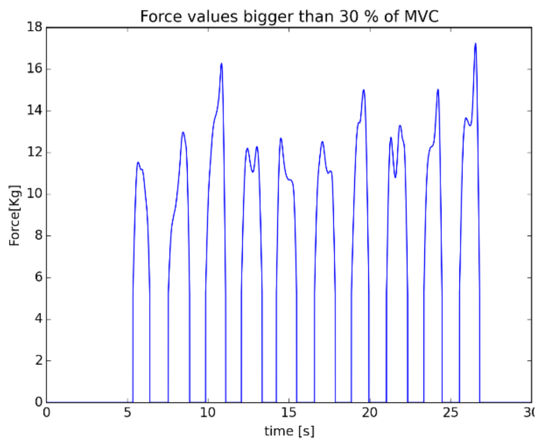

Figure 7 force data which are bigger than 30% MVC

- Take the median value for each contraction
- Group the median values of the contractions for each subject for a certain MVC level, in this case we will have five data sets for each grip type for example Figure 8 shows the plot of the median values for force grip for all five subjects, we can notice that relationship take almost linear form.

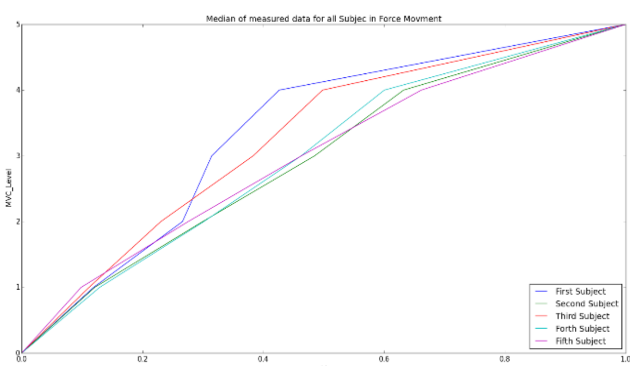

Figure 8 the median values for force grip for all Subject

Fit the data to a linear function and calculate the real intercept and slop Table 5 shows the calculated coefficients from the model for each grip type, all data are normalized to the maximum contraction Figure 9 and Figure 10 and Figure 11 show three model fits for three types of grip.

Grip type	Coefficient ax+b	Tolerance
Pinch	a = 0.011 b = -0.124	+/- 0.0003 +/- 0.023
Precision	a = 0.011 b = -0.1182	+/- 0.0003 +/- 0.023
Force	a = 0.39203 b = 0.001	+/- 0.024 +/- 1.421

Table 5 calculated coefficients from the model for each grip type

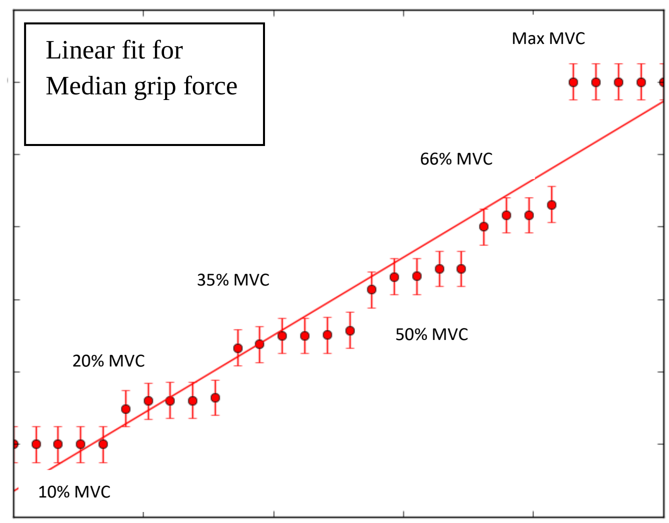

Figure 9 Linear fit for Median Grip Force

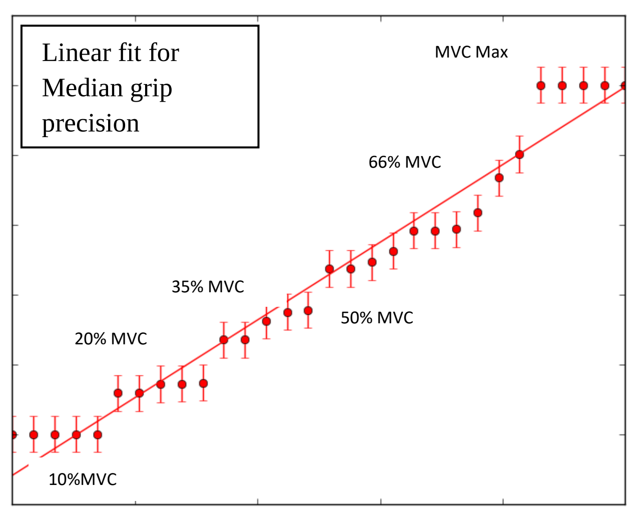

Figure 10 Linear fit for Median Grip precision

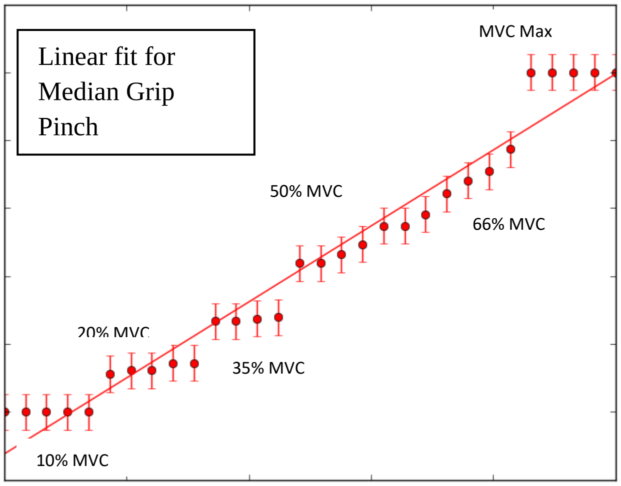

Figure 11 Linear fit for Median Grip Pinch

- Estimate the model

We can estimate the model by plotting the linear fit with error of one standard deviation and three standard deviation, we can notice that all data points lies between +/- 3 standard deviations. Figure 12, Figure 13 and Figure 14 represent the model fits with fit errors.

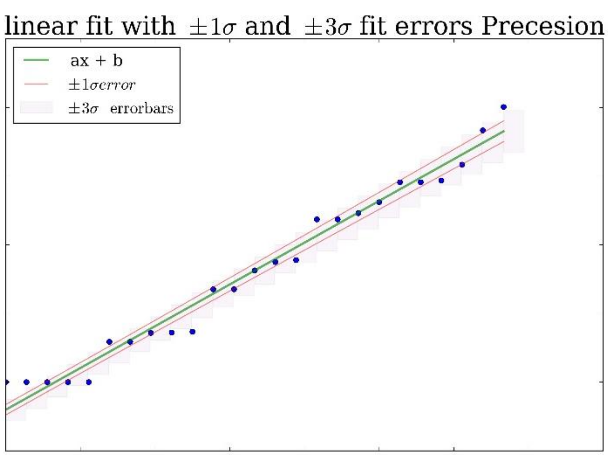

Figure 12 Linear fit with error for Precision

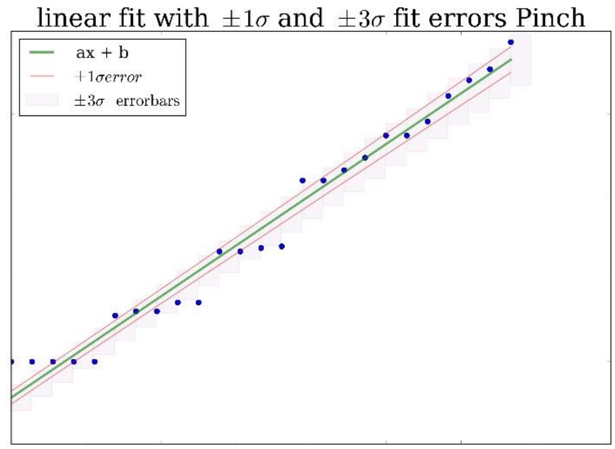

Figure 13 Linear fit with error for Pinch

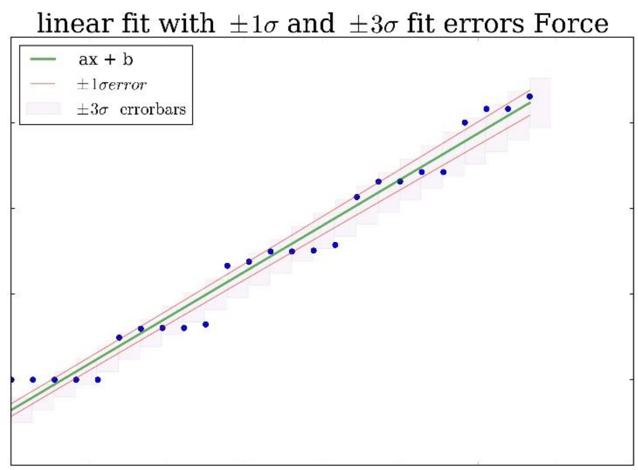

Figure 14 Linear fit with error for Force

3 Conclusion

We can see that our force regression model could fit the data perfectly with acceptable error, based on this model we can predict that any contraction on certain MVC level should fit the line which describe this certain movement, the next step will be the feature extraction of the EMG signal and apply the Principal component analysis to reduce the dimensionality on multivariate data, then applying Linear discriminant analysis to reach to acceptable accuracy in grip type classification.

One of the difficulties we could find during the study is the huge amount of the data which can consume a lot of time during application of any algorithm to a certain sub data files, one way to solve the issues is to work with guidata which is a package provide with python 3 which can create interactive user-interface to deal with directories and drawings.

4 References

[1] NISOH, "Hand/Wrist Musculoskeletal Disorders ,Carpal Tunnel Syndrome, Hand/Wrist Tendinitis, and Hand-Arm Vibration," NISOH publication NO 97, pp. 5-7.

[2] F. M. d. A. L. S. O. T. E. M. M. Heloyse Uliam Kuriki, "The Relationship Between Electromyography and Muscle Force," in EMG Methods for Evaluating Muscle and Nerve Function, 2012.

[3] D. W. H. T. Gareth James, An Introduction to statistical learning with application in R, Springer New York Heidelberg Dordrecht London, 2013.

[4] T. Haslwanter, An Introduction to Statistics with Python, Springer , 2016.