|  |
| --- |
| **Intelligent Systems** |
| Project |
|  |
| Developing of a neural system, a Mamdany-type fuzzy inference system and a Sugeno-type fuzzy inference system for the identification of a volunteer position/activity |
|  |
| **Francesco Paolo Culcasi** |
| **30/05/2016** |
|  |

This page intentionally left blank.

Summary

[Summary 3](#_Toc508459083)

[1 - Project specifications 5](#_Toc508459084)

[1.1 - Data description 5](#_Toc508459085)

[1.2 - Objective 6](#_Toc508459086)

[2 - Data analysis 7](#_Toc508459087)

[2.1 – Data initialization 7](#_Toc508459088)

[2.2 – Data cleaning 7](#_Toc508459089)

[2.2.1 – Signal smoothing 7](#_Toc508459090)

[2.2.2 – Normalization 8](#_Toc508459091)

[2.3 - Feature extraction 8](#_Toc508459092)

[2.4 – Feature selection 9](#_Toc508459093)

[2.4.1 – Principal component analysis 9](#_Toc508459094)

[3 - Neural Network 11](#_Toc508459095)

[Appendix A - MATLAB Scripts 13](#_Toc508459096)

[A.1 init.m 13](#_Toc508459097)

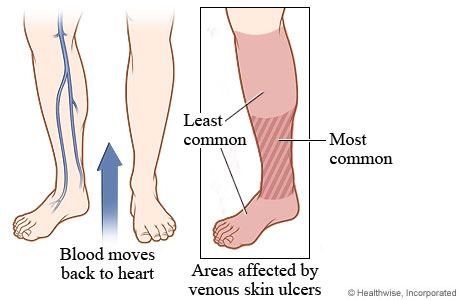
[A.2 split.m 14](#_Toc508459098)

[A.3 cleaning.m 15](#_Toc508459099)

[A.4 featureEx.m 16](#_Toc508459100)

# 1 - Project specifications

The project requires the analysis of a set of medical data. The application context is dermatology and, in particular, the compression therapy by means of bandages in the treatment of venous ulcers of the leg.

In dermatology, venous ulcers are very frequent lesion of the skin of the legs, due to a compromised functioning of the blood circulation. To repair these ulcers, the blood circulation must be enhanced by exploiting the calf muscle pump which allows the blood to be moved back to the heart. The basic way to re-establish the blood circulation is the compression therapy.

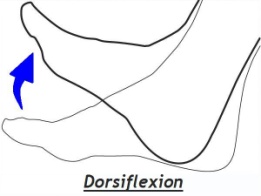
**Figure 1** - Compression therapy

In the compression therapy, the doctor applies a compression bandage on the leg of the patient providing a given pressure of the ulcer area. The pressure applied by the bandage allows to repair the ulcer, typically, in a few months.

The pressure applied by the bandage, and thus the efficiency of the therapy, depends on several factors:

* The type of bandage (depending, e.g., on the elasticity);
* The correct application of the bandage: an appropriate pressure should be applied in several parts of the leg according to the position of the ulcer, and the pressure should remain constant as much as possible between bandage applications;
* The activities performed by the patient during the therapy (e.g., walking, standing, etc.).

## 1.1 - Data description

The data refer to 10 volunteers. A compression bandage was applied to their calf. Three sensors were applied in three different position to measure sub-bandage pressure. Each volunteer wears the bandage for 12 minutes. During this time, the volunteer perform different activities or maintains their positions. Each activity/position is performed/maintained for about 3 minutes. The sensors measure the pressure with the sampling time of about 82 ms. The positions/activities taken into account are the following:

1. Supine position
2. Dorsiflexion standing

**Figure 2** - Dorsiflexion activity

1. Walking
2. Stair climbing

The data are organized as follows. One folder for each volunteer is provided. In each folder there are 4 files, one for each position/activity performed. Each file contains the measurements of the three sensors (in the first three columns), and the corresponding sampling time (in the last column). Please, note that the sensor measurements are electrical resistance and are measured in Ohms.

## 1.2 - Objective

The objective of the project is twofold: on the one hand, we want to identify the position/activity of the volunteer by analyzing the pressure of the bandage, i.e., to distinguish among supine position, dorsiflexion standing, walking and stair climbing; on the other hand, we want to find out the least temporal interval that is necessary and sufficient to recognize the position/activity.

# 2 - Data analysis

Given the data organization decrypted in the previous chapter, each signal coming from a sensor is a set of approximately 2000 samples. Rather than give all these data as input of our system, we prefer to appropriately represent the signal in terms of a reduced number of features that can summarize the information deriving from the original.

For this reason on the following paragraphs we are going to evaluate the signal after a preliminar feature extraction phase.

## 2.1 – Data initialization

At the starting point we have to deal with raw signals that are quite difficult to analyze. This implies a pre-processing phase with the purpose to get some structures in order to ease the computations on data ([*Appendix A.1*](#_A.1_init.m)). In the pressure measurement of the activity we observe that the three different sensors may show different shapes and shifts due to their position on the calf sub-bandage. For this reason we also introduced a fourth “virtual” sensor, obtained as the sum of the former three, to summarize the overall measurement.

Moreover our data should be sliced to obtain several pieces of the signal with smaller duration than the 3 minute signal for each activity. This has twofold profit: we split every signal to find out the least temporal interval that is necessary and sufficient to recognize the position/activity and even more we have much more signals that we can use to train/test/valuate our neural network ([*Appendix A.2*](#_A.2_split.m)).

We notice from the signals that they have been sampled with a sampling period of ~82ms (or 12.2Hz sampling frequency). Therefore we can compute the number N of samples needed to form a M-seconds-long piece of the signal with the following formula:

A vector with several of these values has been defined to faster switch between different temporal intervals in our project. The values are computed for the following intervals: 3sec, 5sec, 10sec, 12sec, 15sec, 20sec, 30sec, 1min, 2min.

## 2.2 – Data cleaning

The signals we are going to processing are obtained from pressure sensors. As like as every real world signal they are affected from noise that deteriorates the signal.

For this reason it is a good rule to apply some cleaning techniques before any feature extraction.

### 2.2.1 – Signal smoothing

One of the most common smoothing technique is the use of Savitzky-Golay filter. Savitzky–Golay filter is a digital filter that can be applied to a set of digital data points for the purpose of smoothing the data, that is, to increase the signal-to-noise ratio without greatly distorting the signal. This is achieved, in a process known as convolution, by fitting successive sub-sets of adjacent data points with a low-degree polynomial by the method of linear least squares.

For our project we used a third-order polynomial and a frame length of 7 ([*Appendix A.3*](#_A.3_cleaning.m)).

(**IMMAGINE PRIMA E DOPO**)

### 2.2.2 – Normalization

Another important point to deal with is the following: signals may have different scale and shift. This is a huge problem since we would like to compare signals. To solve the problem we apply a normalization phase.

Since our dataset may contain some outliers we use Z-score normalization of the signal X, that is not affected from outliers. It can be described with the formula belove:

where we define with μ and σ respectively the signal mean and standard deviation.

## 2.3 - Feature extraction

The goal is to reduce the number of inputs sufficient to distinguish among positions/activities. In order to do that we start analyzing signals shape.

(**IMMAGINE SEGNALI A CONFRONTO**)

From the above figures, representing the 4 activities of the volunteer 1, it become immediately evident that signals belonging in different activities can be discriminated by the following temporal characteristics:

* **Max/Min**: the maximum and minimum values of each signals may be significatively important;
* **Standardized moments**:
  + **1st std. moment (mean)**: after the normalization phase the mean value has been nullified;
  + **2nd std. moment (variance)**: the same as for mean value.
  + **3rd std. moment (skewness)**: is a measure of the asymmetry, computed as
  + **4th std. moment (kurtosis)**: is a measure of the "tailedness", measured as:
* **Autocorrelation**: is the correlation of a signal with a delayed copy of itself as a function of delay, and can be obtained with

For this project we take in consideration the first 20 lags (Δ=1,2,…,20).

In addition, we can introduce some frequency features that may look appropiated:

* **Fundamental frequency** (from now on f0): achieved by detecting the maximum peak abscissa of the signals’ spectrum, obtained by Fast Fourier Transform (FFT) formula,

To faster compute f0 we can approximate it with the maximum value of the FFT;

* **Amplitude of the** f0 **peak** (amp);
* **Power Spectral Density** (PSD): obtained from the spectrum through the formula,

The feature extraction phase is showed in the [*Appendix A.4*](#_A.4_featureEx.m).

## 2.4 – Feature selection

So far we have obtained 25 features: max, min, skewness, kurtosis, autocorrelation (20), f0, amp and PSD. But we have to compute these for each of the 4 sensors (3 + 1 virtual). Then we have 100 features.

There are still a lot of inputs for system, so we have to select only the more significant for the activity classification.

### 2.4.1 – Principal component analysis

# 3 - Neural Network

# Appendix A - MATLAB Scripts

## A.1 init.m



## A.2 split.m



## A.3 cleaning.m



## A.4 featureEx.m

