Chapter 13
Applying Soft Computing to Clinical Decision Support

José Machado
Universidade do Minho, Portugal

Luís Barreiro
Universidade do Minho, Portugal

Lucas Oliveira
Universidade do Minho, Portugal

Serafim Pinto
Universidade do Minho, Portugal

Ana Coimbra
Universidade do Minho, Portugal

ABSTRACT

This article aims to explain the construction process of the learing systems based on Artificial Neural Networks and Genetic Algorithms. These systems were implemented using R and Python programming languages, in order to compare results and achieve the best solution and it was used Diabetes and Parkinson datasets with the purpose of identifying the carriers of these diseases.

INTRODUCTION

Clinical Decision Support is a process whose main goal is to improve heath and healthcare delivery by supporting health-related decisions and actions. (Healthcare Information and Management Systems Society, n.d.) (HealthIT.gov, n.d.)

Nowadays, there is a huge supply of data in Internet through abundant data sources in several areas of knowledge, such as commerce, science and society. Often, is difficult to analysis, obtain conclusions and/or predict future results in this

DOI: 10.4018/978-1-4666-9882-6.ch013

data. Through the use of some tools and techniques it is possible realize a more specify and objective analysis of data, with this analysis it is possible to find results, determine which are the most relevant results of the analysis, and interprets them in order to decide on how they may be useful in the context of the data analyzed.

This way, through Neural Networks and Genetic Algorithms, in this work were analyzed two sets of different data. It was used one dataset about the Diabetes disease and another about Parkinson, which will be explained later. The main goal was to create many different scenarios, make several tests, and then find the best possible solution for these learning systems. Always with the goal of optimize the problem and be able to make a prediction of results with the lowest possible error.

DATASETS

For this work it was select two datasets of public domain, available for download in https://archive.ics.uci.edu/ml/datasets.html. The first is about the Diabetes disease, more particularly in female individuals with at least 21 years old and Pima Indian descent. It is composed by 9 attributes and 768 instances, to know:

1. Number of times pregnant
2. Plasma glucose concentration a 2 hours in an oral glucose tolerance test
3. Diastolic blood pressure (mm Hg)
4. Triceps skin fold thickness (mm)
5. 2-Hour serum insulin (mu U/ml)
6. Body mass index (weight in kg/(height in m))^2)
7. Diabetes pedigree function
8. Age (years)
9. Class variable (0 or 1)

The second dataset analyzes various biomedical measures corresponding to Parkinson disease or not. The main goal of the information is discriminate the people that are Parkinson patients. The information is divided in 23 attributes and 197 instances, to know:

1. Name - ASCII subject name and recording number
2. MDVP:Fo(Hz) - Average vocal fundamental frequency
3. MDVP:Fhi(Hz) - Maximum vocal fundamental frequency
4. MDVP:Flo(Hz) - Minimum vocal fundamental frequency
5. MDVP:Jitter(%),MDVP:Jitter(Abs),MDVP:RAP,MDVP:PPQ,Jitter:DDP - Several measures of variation in fundamental frequency