Computer-Aided Diagnosis in Brain Computed Tomography Screening

Hugo Peixoto[1], Victor Alves[2]

[1]Hospital Center Tâmega e Sousa, Penafiel Portugal
[2]University of Minho, Braga, Portugal

hugo.peixoto@chts.min-saude.pt
valves@di.uminho.pt

Abstract. Currently, interpretation of medical images is almost exclusively made by specialized physicians. Although, the next decades will most certainly be of change and computer-aided diagnosis systems will play an important role in the reading process. Assisted interpretation of medical images has become one of the major research subjects in medical imaging and diagnostic radiology. From a methodological point of view, the main attraction for the resolution of this kind of problem arises from the combination of the image reading made by the radiologists, with the results obtained from using Artificial Intelligence based applications that will contribute to the reduction and eventually the elimination of perception errors. This article describes how machine learning algorithms can help distinguish normal readings in brain Computed Tomography from all its variations. The goal is to have a system that is able to detect normal appearing structures, thus identifying normal studies, making the reading by the radiologist unnecessary for a large proportion of the brain Computed Tomography scans.

Keywords: Medical imaging, Computer Aided Detection, Brain Computed Tomography, Artificial Intelligence, Machine Learning.

1 Introduction

Not only the advent of new methodologies for problem solving but also the emergence of new information and communication technologies has strongly influenced our societies, including their health care systems. Due to such a fine-tuning, Medical Informatics (MI) specialists are ever more involved in managing Information Systems (IS) in health care institutions, in particular of the hospital information ones. Indeed, the monitoring of hospital IS is a fundamental task that has to be followed when one intends to inform the physicians and the management on time, in order to provide a better care.

On the other hand an appropriate design of educational programs in MI and an increasing number of well-trained MI specialists will help to pursue the goal of transforming and improving the delivery of health care through innovative use of information and communication technology. This evolution in technology used in the

medical image practice, confront the radiology physicians with a new problem: the capacity to interpret a huge image workload. Indeed, the current workflow reading approaches are becoming inadequate for reviewing the 300 to 500 images of a routine Computed Tomography (CT) of the chest, abdomen, or pelvis, and even less for the 1500 to 2000 images of a CT angiography or functional Magnetic Resonance (MR) study. In fact, image analysis and treatment methods present enormous development and an increasing utilization in the area of medical imaging. Given to the general interest and the impressive growth in Computer Aided Detection and Diagnosis (CADD), the application of Artificial Intelligence (AI) based techniques in the interpretation and diagnosis of medical image became a rapidly growing research field [1].

For this assessment Brain CT studies were chosen due to its importance in the overall imaging market. Indeed, specialists working in Oporto region, in the North of Portugal, showed that its population, of about 1,500,000 inhabitants, present a yearly average of 450,000 CT analysis, 198,000 (44%) of which were brain CT studies[2]. However, from a methodological point of view, the main attraction for the resolution of this kind of problem arises from the combination of the image analysis taken from the radiologists, with those obtained using AI based applications, that will contribute to the selection of the urgent studies and to aid the radiologists in their readings [2].

Currently, specialized physicians almost exclusively make interpretation of medical images. The next decades will most certainly be of change, and computer-aided diagnosis systems, that have become one of the major research subjects in medical imaging and diagnostic radiology, will play an important role in the reading process. From a methodological point of view, the main attraction for the resolution of this kind of problems arises from the combination of the image reading made by the radiologists, with the results obtained from using AI based techniques that will contribute to the reduction and eventually the elimination of perception errors[3]. Undeniably, the machine learning procedures may distinguish normal studies from all its variations. Our goal is to have a system that is able to detect normal appearing structures, thus identifying normal studies, making the reading by the radiologist unnecessary for a large proportion of the brain CT scans.

Therefore, it is necessary to effect a rigorous assessment in order to establish which, are the real clinical contributions of such systems in decision support. It is here that AI appears to open the way for computer supported diagnosis in medical imaging. Since histological images are complex and their differences are quite subtle, sub-symbolic systems (e.g. Artificial Neural Networks (ANNs), Genetic Algorithms and Evolutionary Programming, Particle Swarm Organization) should be used to overcome the drawbacks of pure symbolic systems. A methodology for brain CT feature extraction and automatic diagnostic generation is proposed. Preliminary results are presented and discussed in regard to the selected features and learning algorithms.

Nowadays Computer Aided Detection (CAD) studies are mainly concerned with screening mammography and thoracic CT studies. Tao Chan et. al studied the effect of a CAD system on clinicians performance in detection of small Acute Intracranial Hemorrhage on CT concluding that there were significantly improvements in the performance of emergency physicians when they make the diagnosis with the support of CAD [4]. Timothy et. al proposed to assess the effect of CAD on the interpretation

of screening mammograms in a community breast center. Their conclusions revealed that the use of CAD in the interpretation of screening mammograms can increase the detection of early-stage malignancies without undue effect on the recall rate or positive predictive value for biopsy [5]. Several commercial CAD products are starting to proliferate and their market is also beginning to develop. Shih-Ping Wang was able to get a US Patent for his CAD system that works over the probability likelihood and predicted values for mammography examinations, aiding the physicians in achieving a correct diagnostic.

The reduction and eventually the elimination of perception errors can be achieved by using neural-network computers taught what is normal with all its variations. The computer should eventually be able to identify normal appearing structures making the reading by the radiologist unnecessary for a large proportion of images.

2 Computer Aided Detection

Automatic segmentation and machine learning are techniques that come from the computer graphics and AI field and are closely related with CAD [3]. Segmentation is used in feature extraction from the images and in many cases this results in excellent clinical information [4]-[5]. In this work, machine learning is concerned with the automatic extraction of information from data, using Knowledge Discovery and Data Mining (KDD) techniques, in our case from the study images. CAD systems should be able to return an adequate answer to the problem of automatic study reading (Fig.1). Here, we were inspired by the work of Kanazawa and Katsuragawa, and the studies follow a similar practical workflow [7]-[8].

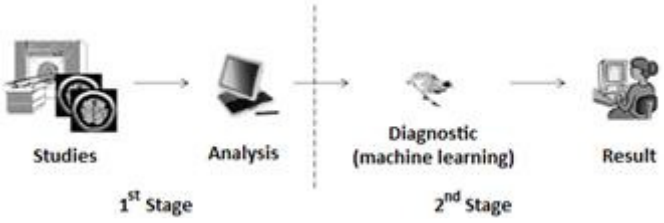

Fig. 1. Stages of CAD

Concerning the studies analyses that where performed, the most relevant features from the images had to be determined, i.e., which aspects of these images, reveals or not the existence of a pathology. From an exhaustive analysis of our case set, we determined that the image's histogram and some shape aspects were key features in determining the existence of pathology. Thus it was decided to consider the histogram and a shape function as the most appropriate features. The case set was composed by 265 studies, where the complete DICOM study, the radiologist report and some additional clinical information (e.g., age, gender) was available.

2.1 General Characteristics Vector

Using the ImageJ framework [9] a plug-in was developed to extract a General Characteristics Vector $\overrightarrow{(GCV)}$ in a csv (comma separated values) format. The global General Characteristics Vector, $\overrightarrow{GCV}_g$, is therefore, given by:

$$\overrightarrow{GCV}_g = (\mathrm{SI}, \mathrm{A}, \mathrm{G}, fc(SH), fs(S), \mathrm{xxVar}, \mathrm{yyVar}, \mathrm{xyVar}, \mathrm{D})$$

where SI, A, G, fc(SH), SH, fs(S), S, xxVar, yyVar, xyVar, and D denote, respectively, the study identification, the patient age, the gender of the patient, the content function, the Study Histogram, the shape function, the Study images, the x variance, the y variance, the xy variance, and the diagnostic. The first parameter, identification of the study, is a value of reference that will identify the study. This dimension as well as the age and gender were extracted from the DICOM images meta-data.

On the other hand, Hounsfield stated that the most relevant information of a CT belongs to the interval [-150;150], whose values are given in Hounsfield Units (HU). In a CT image each voxel corresponds to the x-ray linear attenuation coefficient measurement of the tissue in question. HU of *x* is given by:

$$HUx = 1000 \times \frac{(\mu_x - \mu_{H2O})}{(\mu_{H2O} - \mu_{air})}$$

where x denotes a given voxel, $ì_{H20}$ denotes the linear attenuation coefficients of water and $ì_{air}$ denotes the linear attenuation coefficients of air, considered at a standard temperature and pressure [10].

The standards presented in Table I were chosen as they are universally available, and suited to the key application for which computed axial tomography was developed, i.e., imaging the internal anatomy of living creatures based on organized water structures and mostly living in air, e.g. humans [11].

Table 1. Hounsfield values

Tissue	Interval
Air & sponge	(-∞;-430]
Air	[-100;-50]
Water	[-5;5]
Fluid	[10; 20]
Muscle	[10;40]
Blood	[38; 50]
Grey Matter	[38; 50]
White Matter	[38; 50]
Bone	[150; ∞)

The function *fc*(*SH*) denotes the content function for the total study; for a single image, one has the function called *fci*(*IH*), where *i* denotes a single image of the study. *fci*(*IH*) is now given by:

$$fci(IH) = \left(\frac{x_{HU1}}{t}, \frac{x_{HU2}}{t}, \frac{x_{HU3}}{t}, \ldots, \frac{x_{HUn}}{t}\right)$$

where x_{HU} denotes the total number of voxels, in image *i*, for each HU from -2000 to 2000 and t denotes the total number of voxels in the image *i*. Therefore, *fc*(*SH*) will be given by:

$$fc(SH) = fci\left(\sum_{i=1}^{l}\frac{x_{HU1}}{t}, \sum_{i=1}^{l}\frac{x_{HU2}}{t}, \sum_{i=1}^{l}\frac{x_{HU3}}{t}, \ldots, \sum_{i=1}^{l}\frac{x_{HUn}}{t}\right)$$

where *l* denotes the last image of the study. Fig. 2 illustrates the resulting $fc(SH)$ from -2000 to 2000 HU.

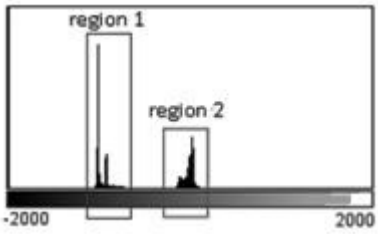

Fig. 2. Study Histogram

After looking at Table 1 and Fig. 2 a threshold from -150 to 150 was performed, since it is where the relevant information for the study lies. This way, the final *fc*(*SH*) denotes the *region 2*. *Region 1* corresponds to the ambient air and the sponge support for the patients head. On the other hand, since representing the *fc*(*SH*) for each HU leads to a huge amount of data, it was decided to sum up the number of pixels dividing the HU in intervals of five. This way we achieved 60 uniform intervals of 5 HU, being represented by $fc_I(SH)$.

From the shape function *f*(*S*) the values of *xSkew*, *ySkew* indicate the symmetry of the series, while the values of *xKurt*, *yKurt*, represent the kurtosis values in the two axes. The remaining parameters, *xxVar*, *yyVar* and *xyVar*, were added to the study and represent the variance in *x*, *y* and *xy* directions, respectively. Finally, and only in the case of the trainings set, the diagnostic value are needed. This parameter stands for a set of values that are given in the form:

- Diagnostic = Ø or
- Diagnostic = {Pathology 1, Pathology 2, …, Pathology n}

where Ø denotes the empty set. With respect to the first evaluation or assessment, it was decided to consider that if D = Ø its value is set to 0 and is judged as non pathological. If the diagnostic is a non empty set, i.e., if D ≠ Ø, the value of D is set to 1 and is taken as pathological.

3 Methodology

The first step for building a good methodology at the meta-level lead us to define a functional overall workflow of the problem solving process, where one is able to manage with all the variants (Fig. 3).

The clinical practice is at the root of the process, once it is the target as well as the basis for obtaining good knowledge. Going a step further, we obtain the image studies of brain CT's as well as the correspondent diagnostic for the training set. Afterward, ImageJ is used to extract the image features, i.e., content function and shape function.

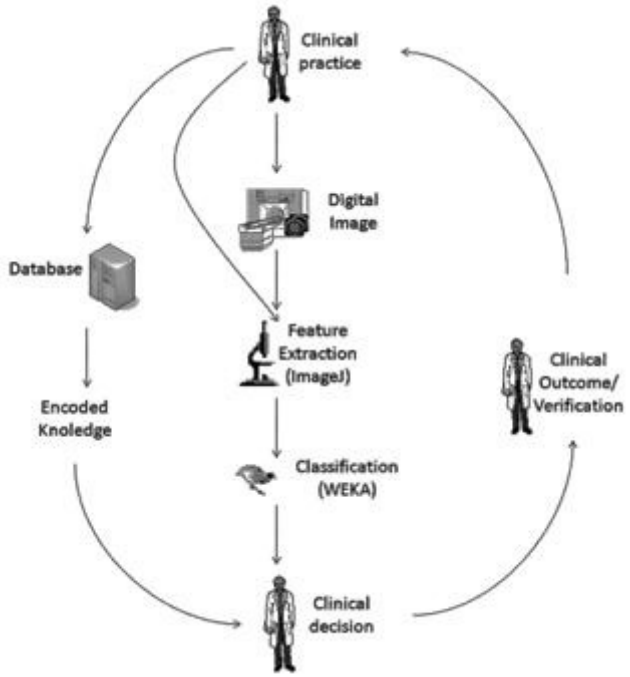

Fig. 3. Workflow Overview

The Waikato Environment for Knowledge Analysis (WEKA) [12] is then used to perform the computational analysis of the supplied data, applying KDD techniques. The $\overrightarrow{GCV}_g$ vector for WEKA is at that time imported, being therefore possible to observe the diverse attributes, distinctly enumerated.

4 Assessment

The $\overrightarrow{GCV}_g$ vector was converted for different evaluation purposes, having different dimensions with different attributes, so that the best attribute set may be obtained. Indeed, after feeding WEKA with the first output $\overrightarrow{GCV}$ several parameters were removed, since they turn out to be irrelevant to the study (e.g., patient ID). Then, $\overrightarrow{GCV}_g$ move towards $\overrightarrow{GCV}_1$, which is given by:

$$\overrightarrow{GCV}_1 = (\mathrm{A}, \mathrm{G}, f_{c1}(SH), fs(S), \mathrm{xxVar}, \mathrm{yyVar}, \mathrm{xyVar}, \mathrm{D})$$

On a second run, after importing the $\overrightarrow{GCV}_g$, other parameters were removed, in addition to the patient ID. The gender was removed and the shape function $f_s(S)$ was modified. *xKurt*, *yKurt* and *ySkew* were deleted and a new shape function, $f_{s1}(S)$ was built. This move was considered since the turnover of keeping such parameters was unknown, and its real impact had to be evaluated. Subsequently, $\overrightarrow{GCV}_g$ move towards $\overrightarrow{GCV}_2$, which is given by:

$$\overrightarrow{GCV}_2 = (\mathrm{A}, f_{c1}(SH), f_{s1}(S), \mathrm{xxVar}, \mathrm{yyVar}, \mathrm{xyVar}, \mathrm{D})$$

where $f_{s1}(S)$ is built by *xSkew*.

Two more attempts were considered. Firstly, it was removed only the gender, then, the gender was kept in and it was removed the xKurt, yKurt and ySkew. In this way we got $\overrightarrow{GCV}_3, \overrightarrow{GCV}_4$, which are given by:

$$\overrightarrow{GCV}_3 = (\mathrm{A}, f_{c1}(S\ \not\#), f_s(S), \mathrm{xxVar}, \mathrm{yyVar}, \mathrm{xyVar}, \mathrm{D})$$

$$\overrightarrow{GCV}_4 = (\mathrm{A}, \mathrm{G}, f_{c1}(SH), f_{s1}(S), \mathrm{xxVar}, \mathrm{yyVar}, \mathrm{xyVar}, \mathrm{D})$$

All the tasks referred to above were performed using only the tools provided by WEKA; however, little variations in the generation of the $\overrightarrow{GCV}_g$ were also tried to guarantee that all the possible scenarios would be covered. Firstly, the distribution of the intervals in the histogram was changed. One of the adjustments was to introduce the notion of weight to the histogram intervals. So, the intervals were multiplied by the sin function. This was made since most of the relevant information is in the middle of the histogram, and as Fig.4 shows, the sin function attributes more weight to its middle locations.

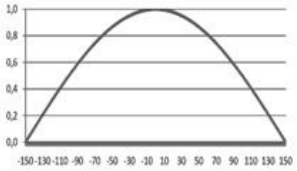

Fig. 4. Sin() function

$\overrightarrow{GCV}_5$ is therefore given by:

$$\overrightarrow{GCV}_5 = (\mathrm{A}, \mathrm{G}, \sin(f_{c1}(SH)), f_s(S), \mathrm{xxVar}, \mathrm{yyVar}, \mathrm{xyVar}, \mathrm{D})$$

Another assessment consisted in grouping the age of the patients into several groups, as it is depicted in Table 2.

Table 2. Grouping by age.

Description (Group)	Age (years)
Baby	0 – 2
Child	2 – 6
Infant	6 – 10
Starter	10 – 14
Juvenile	14 – 16
Junior	16 – 18
Young Adult	18 – 30
Junior Adult	30 – 45
Senior Adult	45 – 60
Old Adult	+ 60

The resulting $\overrightarrow{GCV}$, i.e., $\overrightarrow{GCV}_6$ is therefore given by:

$$\overrightarrow{GCV}_6 = (GROUPED\_\mathrm{A}, \mathrm{G}, f_{c1}(SH), f_s(S), \mathrm{xxVar}, \mathrm{yyVar}, \mathrm{xyVar}, \mathrm{D})$$

Finally, the last interaction was to convert the diagnostic parameter in a non 0 or 1 values, giving, instead, a probability of having pathology. This result in the last assessed $\overrightarrow{GCV}$, leads to $\overrightarrow{GCV}_7$, which takes the form:

$$\overrightarrow{GCV}_7 = (GROUPED_{\mathrm{A}}, \mathrm{G}, f_{c1}(SH), f_s(S), \mathrm{xxVar}, \mathrm{yyVar}, \mathrm{xyVar}, \mathrm{D}_1)$$

5 Results

According to the results so far obtained, it is feasible to conclude that grouping the age, giving different weights to the histogram intervals, and passing a non 0 or greater than 0 the values in the diagnostic, such changes are not so relevant as initially had been thought, i.e., the results for $\overrightarrow{GCV}_5$, $\overrightarrow{GCV}_6$ and $\overrightarrow{GCV}_7$ do not get close to the other $\overrightarrow{GCV}$. Table III presents the attained results for some algorithms using quite a few of $\overrightarrow{GCV}$s, where the best algorithms are the Simple Logistic and the LMT (Logistic Model Trees), being the most stable one the VFI (Voting Feature Intervals classification algorithm), since it presents good results not only to the best $\overrightarrow{GCV}$, but also for the remaining ones. The best $\overrightarrow{GCV}$ has the following characteristics:

- It is divided in sixty uniform intervals;
- It presents a numerical value for age;
- It presents a Boolean value for the Diagnostic; and
- It comes without gender, xKurt, yKurt and ySkew as parameters.

Table 3. Algorithm scores, false negatives and false positives.

GCV	ALGORITHM	SCORE (%)	FALSE POSITIVE (%)	FALSE NEGATIVE (%)
$\overrightarrow{\boldsymbol{GCV}}_1$	VFI	76.92	30.8	15.4
$\overrightarrow{\boldsymbol{GCV}}_1$	Random Committee	76.92	15.4	38.5
$\overrightarrow{\boldsymbol{GCV}}_2$	VFI	80,77	23,1	15,4
$\overrightarrow{\boldsymbol{GCV}}_2$	Simple Logistic	80,77	30,8	7,7
$\overrightarrow{\boldsymbol{GCV}}_2$	LMT	80,77	30,8	7,7
$\overrightarrow{\boldsymbol{GCV}}_3$	VFI	73,08	30,8	23,1
$\overrightarrow{\boldsymbol{GCV}}_4$	VFI	80,77	23,1	15,4

Once the parameters for gender, *xKurt*, *yKurt* and *ySkew* are discharged, the results improved considerably. The same happens when the gender parameter is kept in $\overrightarrow{GCV}_4$. On the other hand, removing the gender parameter and maintaining the other ones, did not score as high as in the previous dataset.

Fig. 5 gives an idea of the form as the results are provided by WEKA. It is possible to distinguish between the false positives and negatives, and the results that are marked as pathological or non-pathological.

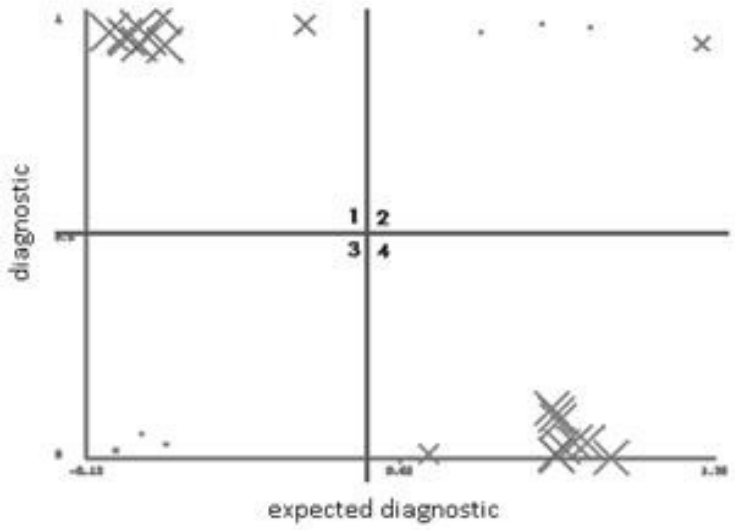

Fig. 5 - Expected Diagnostic vs Diagnostic

It is interesting to notice that since the results are presented in terms of a percentage, the system not only suggests all diagnostic possibilities but also indicates a measure of accuracy for each suggestion. The final result is a system that provides the user with diagnostic suggestions for the selected brain CT. The prototype built for real-world assessment in a medical environment is WEB based. In Fig. 6 we can see a screen shot of the web page where the study is selected for automatic generation of diagnostics. In Fig. 7 we can see the generated diagnostics where the bars stand for the probability of each diagnostic. On the left screenshot we have two big bars standing for a big probability of the diagnostic “Atrophy” and “Stroke”. On the right screenshot of Fig. 7 all bars is very small, meaning that we are in the presence of a normal study, i.e. no pathology is identified.

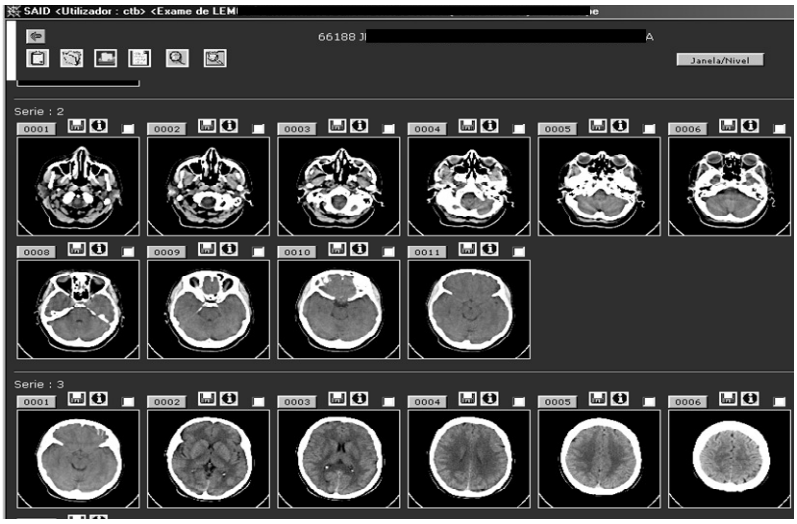

Fig. 6 – Diagnostic interface (study selection for diagnostic)

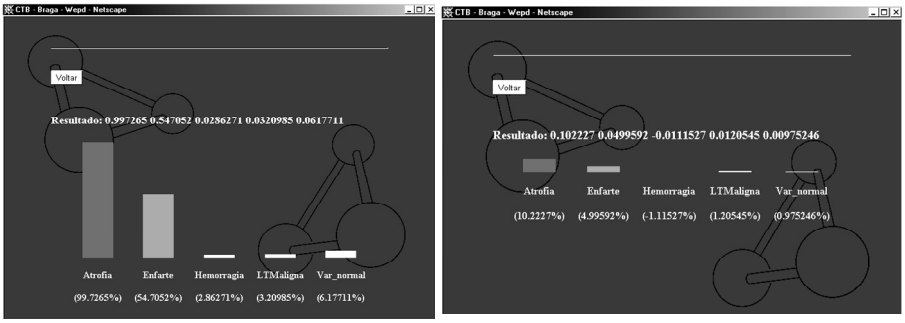

Fig. 7 – Diagnostic interface (generated diagnostics)

5 Conclusions

The primary goal of the presented work was reached. We believe that the relevant parameters for the knowledge extraction were identified. The parameters which significantly influenced the studies, as are the xKurt, yKurt, gender, ySkew were also identified.

From the analysis of the learning algorithms, we concluded that the algorithms that generated the best results, considering the presented dataset, were the Simple Logistic and the LMT fed by a vector without the parameters gender, xKurt, yKurt and ySkew, in which the rightness percentage was of 80,77 % and for false negatives was of 7,7 %. However the steadiest algorithm was the VFI that kept a more steady behavior

along all studies. It is imperative to refer that a higher number of studies could lead to better results. Though, due to some reluctance in providing medical imaging studies with its corresponding diagnostic report by the radiologists, this will always be a problem for future investigation.
This work could certainly evolve to better results, since it has a great potential of development, based on a safe foundation. The learning process, provided with a bigger number of cases will have to be performed to allow a better assessment of the learning algorithms in the knowledge extraction necessary to identify the existence or not of pathology in examinations of brain CT.

Acknowledgment

We are indebted to *CIT- Centro de Imagiologia da Trindade*, for providing the anonymized dataset and for their help in terms of experts, technicians and machine time.

References

1. Doi, K. :Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. Computerized Medical Imaging and Graphics 31: 198–211 (2007).
2. Alves, V., PhD Thesis: Distributed Problem Solving – A breakthrough into the areas of artificial Intelligence and Health (In Portuguese), Departamento de Informática, Escola de Engenharia, Universidade do Minho, Braga, Portugal (2002).
3. Castellino, R. A.: Computer aided detection (CAD): an overview. Cancer Imaging (2005).
4. Chan T., H. K. Huang: Effect of a Computer-aided Diagnosis System on Clinicians' Performance in Detection of Small Acute Intracranial Hemorrhage on Computed Tomography, Acad Radiol, 290–299 (2008).
5. Timothy W. Freer, MD Michael J. Ulissey, MD: Screening Mammography with Computer-aided Detection: Prospective Study of 12,860 Patients in a Community Breast Center, Radiology, 220:781–786 (2001).
6. Wang, Shih-Ping: Computer-aided diagnosis system and method. Patent nr: 96193822. http://www.ipexl.com/patents/others/SIPO/Wang_ShihPing/Wang_ShihPing/ 96193822.html (1996).
7. Katsuragawa, S.: Computerized Scheme for Automated Detection of Lung Nodules in Low-Dose Computed Tomography Images for Lung Cancer Screening. Academic Radiology (2004).
8. Kanazawa, K.: Computer-Aided Diagnosis for Pulmonary Nodules Based on Helical CT Images. Nuclear Science Symposium – IEEE (1997).
9. Abramoff, M.D., Magelhaes, P.J., Ram, S.J.: Image Processing with ImageJ. Biophotonics International, volume 11, issue 7, pp. 36-42 (2004).
10. Hounsfield, G.: Computerized transverse axial scanning (tomography): Part I. Description of system. Br. J. Radiology 46-1016 (1973).
11. Geise, R. A. and McCullough, E. C.: The use of CT scanners in megavoltage photon-beam therapy planning. Radiology; 133–141 (1977).
12. Ian H. Witten and Eibe Frank: Data Mining: Practical machine learning tools and techniques, 2nd Edition, Morgan Kaufmann, San Francisco (2005).