UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

Visión Electrónica
Más que un estado sólido
https://doi.org/10.14483/issn.2248-4728

VISIÓN ELECTRONICA
A RESEARCH VISION

Supervised classifiers of prostate cancer
A geometric study on magnetic resonance images T2 weighted (T2W), by diffusion (DWI-ADC)

Clasificadores supervisados del cáncer de próstata
Un estudio geométrico sobre imágenes de resonancia magnética ponderadas T2 (T2W) y por difusión (DWI-ADC)

Natalia Andrea Ramírez-Pérez[1], *Lilia Edith Aparicio-Pico*[2], *Ernesto Gómez-Vargas*[3]

INFORMACIÓN DEL ARTICULO

Historia del articulo
Enviado: 23/09/2018
Recibido:24/09/2018
Aceptado: 12/11/2018

Keywords
cancer,
diagnosis,
geometry,
magnetic resonance,
prostate.

Palabras clave:
cáncer,
diagnóstico,
geometría,
resonancia magnética,
próstata.

ABSTRACT:

Prostate cancer is a common type of cancer in men, it is slow and silent, and they respond to timely treatment.
There is great importance in the early diagnosis when it has not yet invaded the prostate gland, currently there is a need to deepen the objective prediction tools. This article presents an analysis and development of an application for early diagnosis of this cancer from the multiparameter magnetic resonance of patients, in this case patient images are used to detect lesions in the prostate and the Data and Reports System Prostate images (PI-RADS).
The objective of the method is to provide a contribution to medicine in the hands of all clinical staff responsible for the diagnosis of prostate cancer, a method of prediction, as a support for diagnosis, seeking stability and tranquility of the affected patients.

RESUMEN

El cáncer de próstata es frecuente en los hombres, es lento y silencioso, y responde a un tratamiento oportuno. Existe una gran importancia del diagnóstico temprano cuando aún no ha invadido la glándula prostática; actualmente, se ha visto la necesidad de profundizar en las herramientas de predicción objetiva. En este artículo se presenta un análisis y desarrollo de una aplicación para diagnóstico temprano de este cáncer a partir de la resonancia magnética multiparamétrica de pacientes, usándose imágenes de pacientes para detectar lesiones en la próstata y el Sistema de Datos e Informes de Imágenes de próstata (PI-RADS). Se obtiene un método que proporciona al personal clínico encargado del diagnóstico de cáncer de próstata un método de predicción aceptable como soporte al diagnóstico, estable, y que no intranquiliza pacientes afectados.

1 Mathematician, Universidad Distrital Francisco José de Caldas, Colombia. MSc. (c) in Information and Communications Sciences, Universidad Distrital Francisco José de Caldas, Colombia. Current position: Support Teacher at Universidad Distrital Francisco José de Caldas, Colombia. E-mail: naaramirezp@correo.udistrital.edu.co ORCID https://orcid.org/0000-0003-4389-7295
2 BSc. in Special Education, MSc. in Teleinformatics, Specialist in Management of Educational Projects, Universidad Distrital Francisco José de Caldas, Colombia. PhD. in Technical Sciences, Universidad Central Marta Abreu De Las Villas, Cuba. Current position: Professor at Universidad Distrital Francisco José de Caldas, Colombia. E-mail: leap0763@gmail.com medicina@udistrital.edu.co ORCID https://orcid.org/0000-0003-1841-4423
3 Ph.D. in Engineering, Pontificia Universidad Javeriana, Colombia. Current position: Professor at Universidad Distrital Francisco José de Caldas, Colombia. E-mail: egomez@udistrital.edu.co ORCID: https://orcid.org/0000-0003-4957-7313
Cite this article as: N. A. Ramírez-Pérez, L. E. Aparicio-Pico and E. Gómez-Vargas, "Supervised classifiers of prostate cancer", *Visión Electrónica*, vol. 2, no. 1, Special edition, January 2019. https://doi.org/10.14483/issn.2248-4728

1. Introduction

The Magnetic resonance imaging is increasingly used to diagnose prostate cancer, as it has improved the sensitivity and specificity of PSA [1], being such the reason that the mathematical-computational support of prostate cancer in magnetic resonance imaging is an active research area [2].

Prostate cancer, due to the accumulation of prostate fluid in the male gland called the prostate, is presented with a blockage of the urethra that interrupts urine output, it is the most common type of cancer worldwide in men with 1.6 million cases, it is a neoplasm that can become fatal if not diagnosed in time and apply appropriate treatment, thanks to the new tools used for diagnosis, prostate cancer has had a tendency to decrease annually in the world , but not in Latin America, where the mortality rate has increased, becoming the leading cause of cancer death in Latin America [3]. The Bahamas is the country with the highest rate of death from prostate cancer with 78.86 deaths per 100,000 population, followed by Barbados with 76.8 and Trinidad and Tobago with 70.67. Then there is Mexico with a predominance in Colima, Nayarit, Jalisco second of Tlaxcala, Quintana Roo and Yucatán.

In the first part of this study, there is a synthesis on the use of classification models for the diagnosis of prostate cancer, highlighting the importance of the implementation of computational models to support the diagnosis of the disease [4]. Currently, Magnetic Resonance is used as a prostate cancer diagnosis system; Because there are inaccurate diagnoses due to lack of symptoms, a contribution is shown here with the help of an expert system that makes it possible to diagnose more accurately cancer lesions that facilitate identification and thus adequately prevent evolution [5].

The second part shows a list of the equations used to make the geometric study, on magnetic resonance imaging with weighting in sequence T2 (T2W), by diffusion (DWI-ADC), together with the analysis of the structures involved, through the extraction of geometric characteristics such as the measurement of the marking of the lesion for the three images (T2, DWI and ADC), eccentricity, area, arc length, extreme points, centroid, convexity, radius, major axis length, length of minor axis, orientation, solidity, extension, equivalent diameter and fractal dimension per case count, together with the data of the coordinates of the position of the lesion, the matrix that describes the orientation and scale of the image, the values of i, j and k being the column, row and coordinate of the cut to be found respectively, the vector with spacing scalars x, y, z and the position of the lesion TZ: Transitional zone, PZ: Peripheral Zone and A S: Fibromuscular stroma, taking into account the PI-RADS evaluation categories for the probability of having clinically significant prostate cancer.

For the use of the prostate cancer diagnosis support system, Magnetic Resonance imaging was used taking into account the general evaluation of PI-RADS, which is different due to the location of the lesion, because in the transition zone The evaluation of PI-RADS is determined by the T2W score and is sometimes modified by the DWI score, but in the peripheral area the evaluation of PI-RADS is determined by the DWI and ADC scores.

The methods currently used to evaluate prostate cancer are based on Magnetic Resonance imaging that detect different prostate cancer lesions and the PI-RADS System or 5-point classification system responsible for classifying the lesions found in Magnetic Resonance , this is a contribution that combines Machine Learning with images, a branch of medicine that uses algorithms to extract characteristics of medical images, with the use of data sets and their validation and have allowed researchers to classify prostate cancer in patients and therefore a diagnosis with a better predictive value [6].

On the other hand, it continues with the analysis and results sample using the data of 63 patients where logistic regression models, Neural network, decision tree, random forests are used, where categories 4 and 5 of PI-RADS are used. The benefits that can be accessed with the classification models, through the extraction of geometric characteristics, will be a system that will be available to the expert doctor for the diagnosis of cancer, to increase the reliability and tranquility of the patient by improving the appropriate intake of decisions that will direct and define the course of the disease where the goal will be to use the most appropriate treatment that can end the problem and provide a better quality of life for patients, also aimed at the entire branch of oncology, hospital centers and other interested parties, which will be a great ally as a contribution to speed up the diagnosis, as a complement to help using an automatic classifier and geometric variables for improvement and speed.

2. Materials and methods

The research, together with the innovation of science and technology, has advanced until finding important contributions in the field of the study of diseases where oncology occupies a primary place for diagnosis, with the new Automatic Learning System it is possible to use a set of data to recognize patterns associated with cancer which until now have been difficult to see

conventionally by expert doctors.

It was taken into account that the Prostate Specific Antigen (PSA) a protein produced by the cells of the prostate gland is obtained with a blood test that measures PSA levels, since PSA does not exist for a certain age and for the presence of Prostate cancer PSA levels rise, but they are not a determinant for the diagnosis as they are also elevated by a benign infection called, Benign Prostatic Hyperplasia or benign enlargement of the prostate [7].

In addition, histology elements were considered, branch of the anatomy that analyzes the tissues under the microscope, determines an adenocarcinoma for prostate cancer, there being less common histological types such as neuroendocrine prostate cancer and small cell prostate cancer than by Being aggressive variants produce less PSA spreading outside the prostate [8]. These rare variants tend to be more aggressive, produce much less PSA and spread outside the prostate before [10].

The prostate specific antigen (PSA) is a protein produced by the cells of the prostate gland and released into the bloodstream. PSA levels are measured by a blood test. Although there is no "normal PSA" for any man at a certain age, a higher than normal PSA level can be found in men with prostate cancer. Other non-cancerous prostate conditions, such as prostatitis, can also lead to an elevated PSA level. Prostatitis is inflammation or infection of the prostate. In addition, some activities such as ejaculation may temporarily increase PSA levels. This should be avoided before a PSA test to avoid falsely elevated tests [9].

2.1. Clinical presentation

Next, in table 1, the clinical presentation of prostate cancer is shown:

Prostate cancer is usually detected by:	Clinically patients may present:	Clinical scores that may be useful for assessing the severity of symptoms include:
A high prostate specific antigen (PSA) (greater than 4 ng / dL); normal is 1-4 ng / dL Exam abnormal digital rectal • The prostate gland is divided into several anatomical areas and prostate cancer usually occurs within the peripheral area near the rectum, so a digital rectal exam (DRE) is a useful screening test.	• Urinary symptoms, p.eg nocturia, hesitation, urgency, drip terminal • Back pain	• IPSS: international prostate symptom score, 35 points score mainly when assessing urinary symptoms • IIEF: international index of erectile function.

Table 1. Clinical presentation Prostate cancer [11].

2.2. Anatomy

Zonal anatomy

Scroll through the images to see the areas in which the prostate is divided:

- The transition zone surrounds the prostate urethra. This area is enlarged in elderly men, resulting in benign prostatic hyperplasia.
- The central area is located at the base of the prostate, behind the transition zone and surrounds the left and right ejaculatory duct.
- A posterior fibromuscular stroma is a small area of tissue that is located on the anterior side of the prostate.
- The peripheral area is located on the back and side of the prostate.

70-75% of all prostate cancers originate in the peripheral area (PZ). The posterior aspect of this area can be examined with a digital rectal exam. 25% of prostate cancers originate in the transition zone (TZ). Very few prostate cancers manifest in the central zone or in the anterior fibromuscular stroma [12]

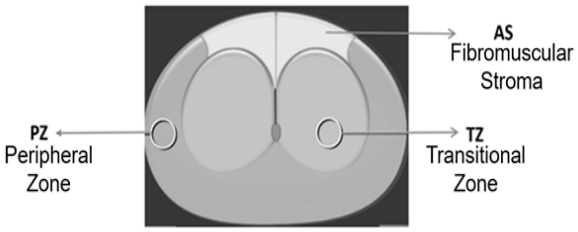

Figure 1. Parts of the prostate [13].

2.3. Prostate image and data system

PI-RADS (Prostate Imaging Reporting and Data System) refers to a structured report scheme to evaluate the prostate for prostate cancer. It is designed to be used in a patient prior to therapy.

The original PI-RADS score was annotated, reviewed and published as the second version, PI-RADS v2, by a steering committee composed of the joint efforts of the American College of Radiology (ACR), European Society of Urogenital Radiology (ESUR) and the AdMeTech Foundation.

The score is calculated from the evaluation of a prostate-specific MRI. Images are obtained using a multiparameter technique that includes T2-weighted images, a dynamic contrast study (DCE), and DWI. If DCE or DWI are insufficient for interpretation, the most recent guidelines recommend omitting them in the score [14].

A score is given according to each variable. The scale is based on a score of 1 to 5 (which is administered for each lesion), 1 being probably benign and 5 being highly suspected of malignancy:.

- **PI-RADS 1:** very low (clinically significant cancer is

is very unlikely)
- **PI-RADS 2:** low (clinically significant cancer is unlikely)
- **PI-RADS 3:** intermediate (the presence of clinically significant cancer is equivocal)
- **PI-RADS 4:** high (a clinically significant cancer is likely)
- **PI-RADS 5**: very high (a clinically significant cancer is very likely) [15].
-

3. Development of the topic

3.1. Considerations on the current status of the prostate cancer study.

It is important to note that, one of the reasons why magnetic resonance imaging has not yet progressed to a first-line modality for the diagnosis of prostate cancer is that it requires a substantial experience of the radiologist to read the magnetic resonance of the prostate and such experience is not widely available [16].

In the article "Computer-Aided Detection of Prostate Cancer in MRI" the authors expose the importance of having diagnostic support from computer assistance where they use prostate segmentation based on multiple atlases, extraction of voxel characteristics and classification and detection of local maxima, where thanks to the laser classification they find the probability of cancer for each patient [17], in the research "Advances in Intelligent Systems and Computing" the author recommends the use of machine learning techniques to identify areas of Prostate cancer in complex histological images [18], in which it is taken into account that the separation of benign glands and stromal cancer areas is one of the vital steps towards the automatic classification of prostate cancer in digital images, presents a novel tool that uses a supervised classification of hist Component ograms in hematoxylin and eosin imaging to delimit areas of benign glands and cancer. Using high-resolution images of complete slide prostatectomies, they compared several image classification schemes that included intensity histograms, oriented gradient histograms and concatenations with manual tissue annotations by a pathologist [19].

Classification models for prostate cancer are applied, such as logistic regression, decision tree, Random forests, Naive-Bayes (Gaussian), K-neighbors, multilayer perceptron, Naive-Bayes (Bernoulli), support vector classifier , stochastic descending gradient, support vector classifier (Linear), support vector classifier (Nu-support), vector support machines, emphasizing theoretical concepts, knowledge and research that contribute to the support of medical diagnosis of this type of cancer, taking into account the experience and background of many studies, with data, for the experimentation and classification that supports the medical diagnosis of prostate cancer, striving to have an accurate and early diagnosis through the extraction and analysis of geometric characteristics, as a valuable contribution of science and technology that generates hope for patients and doctors that cool They have such a careful disease.
To avoid subjectivity in reports, a standardized reporting system, known as PIRADS (Prostate Imaging and Reporting Archiving Data System scoring system) is being used. It is essential to have radiologists specialized in this test for a precise interpretation [20]. The process in general that is applied to each of the images per patient, consists of a filtering of the image to enhance and detect contours, from which the different characteristics are determined, which will be the attributes of the data set constructed, said measures eccentricity, area, arc length, extreme points, centroid, convexity, center and radius, major axis length, minor axis length, orientation, solidity, extension, equivalent diameter and fractal dimension per box count, together with the data of the coordinates of the position of the lesion, the matrix that describes the orientation and scale of the image, the values of i, j and k being the column, row and coordinate of the cut to be found respectively, the vector with spacing scalars x, y, z, and the position of the TZ lesion: Transitional zone, PZ: Peripheral Zone and AS: Fibromuscular stroma. All the mentioned measures are in pixels and are carried out considering the data system and report in prostate imaging (PI-RADS), which was created in order to help in the detection, location of prostate cancer with magnetic resonance, with which is intended to reduce variability in the interpretation of prostatic magnetic resonance studies.

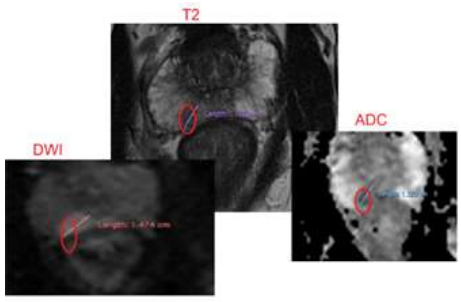

Figure 2. Sequences with the marking of the lesion corresponding to PI-RADS 4. Source: own.

Most cancers in their initial stage do not cause symptoms, which happens with prostate cancer, in an advanced stage it presents symptoms such as; problems with urination, bleeding, erectile dysfunction, back and hip pain, numbness in the feet among others.

It is the experience of a Radiologist who reads and interprets the Magnetic Resonance images of the prostate who are responsible for issuing the final diagnosis of prostate cancer, sometimes resulting in a doubtful diagnosis due to the limited availability of a substantial effective experience. The article on Artificial Intelligence in prostate cancer by María Peñas Leonor, Laura Vizoso Tercero, on the diagnosis of prostate cancer uses automatic segmentation of the prostate using Marginal Space Learning (MSP) that aims at automatic segmentation of the prostate, the which consists of detecting approximately the shape of the prostate by normalizing the data, and then refining the edges and shape, with results obtained on magnetic resonance imaging enhanced in T2 of the prostate demonstrating the effectiveness of learning-based methods such as Efficiency of the deformations that apply to the limits of the segmentation of the lesion during the process.

Automatic segmentation of the prostate remains a challenge due to its resemblance to the seminal vesicles, the geometric variability caused by the disease and the presence of adjacent structures such as the rectum and bladder, manual delineation of the prostate is still necessary as a method starting to perform automatic segmentation [21].

3.2. T2-DWI image processing

The method for processing magnetic resonance images in T2 sequences and DWI - ADC diffusion consists in marking the lesion, the detection process, which consists of converting the image to grayscale and then to a binarized image. In the latter, the pixels consist of values of 0 and 1 only, resulting in a black and white image. Therefore, noticing a change in pixel value identifies those with this property and then draw it. Subsequently, the described geometric variables are extracted.

The classification report confirms the accuracy of the models for the predictions, which can be seen in the Table 2. The accuracy is verified taking into account that it is calculated thanks to the sum of the positive predictions classified as positive and the negative ones classified as negative divided over the total data.

Classification Model	Preciseness
Logistic regression	0.80
Decision tree	0.94
Random Forests	0.96
Naive-Bayes (Gaussian)	0.79
k-neighbors	0.74
Multilayer Perceptron	0.74
Naive-Bayes (Bernoulli)	0.52
Support Vector Sorter	0.50
Stochastic Descending Gradient	0.70
Support vector classifier (Linear)	0.75
Support vector classifier (Nu-support)	0.50
Vector Support Machines	0.90

Table 2. Results of implementation of the classification models. Source: own.

4. Equations

A contour line is a closed curve that represents the edge of a figure taking into account the same colors or same intensity in pixels. The detection process consists of converting the image to grayscale and then to a binarized image. In the latter, the pixels consist of values of 0 and 1 only, resulting in a black and white image. Therefore, noticing a change in pixel value identifies those with this property and then draw it. Next, the geometric image contour characteristics used in this work and which were developed from Python programming version 3.7.1 will be described.

4.1. Area

It refers to the area enclosed within an outline, and also to the number of non-zero pixels in the image through the Green formula.
Let C be a simple closed curve oriented counterclockwise, smooth to pieces on the plane, if D is the internal region enclosed by C, L and M are functions of $\overline{D}$ or $\mathbb{R}$ with continuous partial derivatives then

$$\oint_C (Ldx + Mdy) = \int \int_D \left(\frac{\partial M}{\partial x} - \frac{\partial L}{\partial y}\right) dxdy$$

And so for the area of region D which is $\int \int_D dA$ For a binarized image the area is given by:

$$Área(D) = \sum_i \sum_j I(i,j)$$

Also called discrete Green formula. $I(i,j)$ is the value of the pixel in the i-th row and j-th column of the image. In the case of being a binarized image, $I(i,j)$ takes values of 0 or 1, for which, the area enclosed within the lesion contour will have pixels of value one and the area results in a pixel count of value 1.

4.2. Length

Also called perimeter or arc length, calculates the length of the contour curve. Normally the calculation of this magnitude is

$$L(C) = \int_a^b \sqrt{1 + [f'(x)]^2}$$

if C is defined by a differentiable function f, or if C is defined parametrically by $(x(t), y(t))$ so so

$$L(C) = \int_a^b \sqrt{[x'(t)]^2 + [y'(t)]^2}$$

However, since the image results as a discrete subset of the plane and the points that form the 'curve' are finite, say N, then the length is determined by

$$L(C) = \sum_{i=1}^{N} \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2}$$

which is the sum of the Euclidean distances between consecutive points (x_{i-1}, y_{i-1}) and (x_i, y_i).

4.3. Moment

The moment in general is a quantitative measure of the shape of a function, which represent various physical characteristics, such as density, mass, center of mass, etc. In particular, the moment on an image is the different amounts of particular weights given to the intensity of pixels on the image.

Again by Green's theorem the different quantities at which the moment is treated are calculated from:

$$m_{ij} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^i y^j f(x,y) dxdy$$

That being located on an image that is a discrete set of the plane:

$$m_{ij} = \sum_x \sum_y x^i y^j I(x,y)$$

It should be noted that $m_0 0$ it is the area of the contour.

4.4. Centroid

O geometric center represents the average position with respect to the shape of the contour. It can be calculated as follows:

$$C_x = \frac{\sum_x \sum_y xI(x,y)}{\sum_x \sum_y I(x,y)} = \frac{m_{10}}{m_{00}} \quad C_y = \frac{\sum_x \sum_y yI(x,y)}{\sum_x \sum_y I(x,y)} = \frac{m_{01}}{m_{00}}$$

For (c_x, c_y) centroid coordinates.

4.5. Hu Moments

The central moments are defined as

$$mu_{ji} = \sum_x \sum_y (x - C_x)^j (y - C_y)^i I(x,y)$$

and the respective standardized as:

$$nu_{ji} = \frac{mu_{ji}}{m_{00}^{((i+j)/2)+1}}$$

These are invariant under translations if there are similar figures in various places in space, their central moment will be the same. Now for this invariance to be maintained for translations, rotations and escalations, it is that Hu's moments are needed. Which are defined by

$$h_0 = nu_{20} + nu_{02}$$

$$h_1 = (nu_{20} - n_{02})^2 + 4nu_{11}^2$$

$$h_2 = (nu_{30} - 3nu_{12})^2 + (3nu_{21} - nu_{03})^2$$

$$h_3 = (nu_{30} + nu_{12})^2 + (nu_{21} + nu_{03})^2$$

$$h_4 = (nu_{30} - 3nu_{12})(nu_{30} + nu_{12})[(nu_{30} + nu_{12})^2 - 3(nu_{21} + nu_{03})^2] + (3nu_{21} - nu_{03})[3(nu_{30} + nu_{12})^2 - (nu_{21} + nu_{03})^2]$$

$$h_5 = (nu_{20} - nu_{02})[(nu_{30} + nu_{12})^2 - (nu_{21} + nu_{03})^2 + 4nu_{11}(nu_{30} + nu_{12})(nu_{21} + nu_{03})]$$

$$h_6 = (3nu_{21} - nu_{03})(nu_{21} + nu_{03})[3(nu_{30} + nu_{12})^2 - 3()^2] + ()()[3()^2 - ()^2]$$

4.6. Extremal Points

They are defined as the extreme points of the curve, that is, the coordinates of the points that are higher, lower, more to the right and more to the left. More formally means if $\{x_i\}$ and $\{y_i\}$ the sets of first and second coordinates of points that are on the curve.

Therefore if

$$x_j = \max_i x_i, \quad x_k = \min_i x_i, \quad y_l = \max_i y_i, \quad y_m = \min_i y_i$$

So the extremal points are:

$$(x_j, y_j), \quad (x_k, y_k), \quad (x_l, y_l), \quad (x_m, y_m)$$

4.7. Convexity

The convexity of a curve is about the convexity of the region that encloses that curve. A set It is A convex if for everything $x, y \in A$ for all $t \in [0,1]$

$$(1 - t)a + tb \in A$$

In other words, the line segment between two points within the region is contained in the region.

4.8. Center and radio

If the points that generate the curve are considered, it is necessary to generate the smallest circumference that encloses such points. The center and radius of said circumference is what these measures refer to.

Now as the set taken A is finite it exists if the diameter of this set is defined as

$$diam(A) = \sup\{|x - y| : x, y \in A\}$$

Jung's theorem states that the radius of the smallest circle that encloses A cannot exceed diam $(A)/\sqrt{3}$.

Additionally, geometrically, it can be observed that this circumference is only determined and can be found by finding a maximum of three "limit" points that must belong to the cycference.

4.9. Ellipse

If the contour is made analogous to the previous problem, a rectangle that encloses the contour with the smallest possible area. And if another condition is added to it like the rectangle can be rotated. From this rectangle, the ellipse inscribed in it is now generated.

4.10. Major and minor axis length

The major and minor axis length is the diameter of the circumscribed and inscribed circumference respectively.

4.11. Orientation

It is the angle between the x axis and the main or focal axis that, as the name implies, is the axis of the ellipse where the foci are located

4.12. Eccentricity

It is a measure that relates the measure of the foci with the length of the main axis and can be calculated by

$$Excentricidad = \sqrt{1 + \frac{Longituddeejemenor^2}{Longituddeejemayor^2}}$$

4.13. Solidity

If you take the collection of points A on the contour, you can generate the smallest convex set that contains A . Especially if it is done on convex polygons, it is known as a connected envelope or convex hull of A More formally it is the set of all the convex combinations of A that is to say,

$$Conv(A) = \{\sum_i \alpha_i x_i : \alpha_i \geq 0, \sum_i \alpha_i = 1\}$$

If, for example, the points were on the same line Conv (A) it is the segment that joins the extreme points. Then the solidity is defined as the ratio of the area of its contour to its convex area of the hull:

$$solidity = \frac{Contour\ Area}{Convex\ hull\ area}$$

4.14. Tumor Extension

It is defined as the proportion of the area of the contour with respect to the area of the bounding rectangle:

$$Extension = \frac{InjuryArea}{BoundaryRectangleArea}$$

4.15. Diameter

It is the diameter of the circle whose area is the same as the area of the contour, and can be found by the following expression

$$Equivalent\ diameter = \sqrt{\frac{4(ContourArea)}{\pi}}$$

4.16. Fractal dimension

A fractal is an irregular geometric structure with the property of self similarity, meaning that it is repeated at different scales. There is a way to provide this type of object with a dimension, this is known as the box count dimension, the Minkowski or Hausdorff dimension. In it, the complexity with which the figure changes through the scales is measured.

Let $X \subset \mathbb{R}^2$ be the fractal and $\{R_i(\epsilon)\}_{i\in\mathbb{N}}$ the collection of open boxes or squares whose sides measure ϵ, that partition the space (image). If $f_X: \mathbb{R} \to \mathbb{N}$ such that

$$f_X(\epsilon) = \sum_{i\in\mathbb{N}} \chi_{R_i(\epsilon)}(X)$$

where $\chi_{R_i(\epsilon)}(X)$ is 1 if there is an $x \in X$ for which $x \in R_i(\epsilon)$ or otherwise 0. The upper Minkowski dimension of X as

$$\overline{dim}(X) = \limsup_{\epsilon\to 0} \frac{\log(f_X(\epsilon))}{log(1/\epsilon)}$$

and in counterpart, the lower Minkowski dimension by

$$\underline{dim}(X) = \liminf_{\epsilon\to 0} \frac{\log(f_X(\epsilon))}{log(1/\epsilon)}$$

And when both coincide:

$$\overline{dim}(X) = \underline{dim}(X)$$

It is the Minkowski box count dimension of X [22].

5. Analysis of results

Note that in addition to the variables mentioned in 3 and convexity, other variables such as 24 moments, including centralized and normalized moments, were considered. These were excluded because the accuracy of the classification models decreased; in particular, the different moments are contained in the moments of Hu being these invariants under translations, rotations, and reflections. Convexity on the other hand, generated only one value and therefore did not contribute to the distinction between the PIRADS contemplated, thanks to the fact that these variables were excised, greater precision is obtained by the random forest classification model set out in Table 2 of the 96%

6. Conclusions

The variables obtained from the data set are included in each classification model to know their performance in terms of accuracy, and with this the results obtained are compared; it was found that supervised classifiers such as random forests and the decision tree have a better performance than logistic regression, Naive-Bayes, k-neighbors, multilayer perceptron, support vector classifier, stochastic descending gradient and support machines Vector, for the classification of the PI-RADS evaluation categories, on magnetic resonance imaging with T2 weighting and DWI-ADC diffusion, with an accuracy of 0.96 for random forests and 0.94 for the decision tree.

In the 21st century, science and technology come together to provide great solutions and in the area of medicine they impact on the development of a Machine Learning Machine Learning (ML) system, to support medical diagnosis, which using Magnetic Resonance images of humans and with the different supervised classifiers evaluated, taking into account the Prostate Imaging Data and Reporting System (PI-RADS), we have a system aimed at supporting the diagnosis of prostate cancer. The diagnosis of prostate cancer is a long process that is currently carried out in several steps ranging from rectal examination, PSA tests and magnetic resonance imaging, which have a subjective classification where the final decision depends on the experience of the Professional, expert doctor in charge of interpreting the images, therefore with the machine learning method and the PIRADS classification, the images will be classified at a more exact level for diagnostic support.

With the extraction of geometric characteristics, the numerical characterization in pixels of the lesion is performed, with the PIRADS evaluation categories that will help the expert doctor in charge of issuing the diagnosis of prostate cancer who will have a tool that is implemented to the images obtained by Resonance Magnetic as a numerical value of easy representation on the categories of evaluation of the patient of the different variables extracted as a reference for the classifiers.

To quantify the utility of all variables throughout the

random forest, the relative importance of the variables is observed, which represent how much, including a particular variable, the prediction improves. These variables are the measure in cm of the lesion, orientation, extreme points, solidity, extension, diameter, eccentricity, Hu_momento_h2, Hu_momento_h3 and fractal dimension, standing out by 40% over the other variables.

Mathematics drives medicine with its precision, thanks to the quantification provided by methods for doctors to establish balance through applications and calculations when issuing diagnoses, executing treatments and surgeries, geometry is used in specialties such as analysis and Image processing, where the application of angles, planes and measures, are decisive to grant an accurate diagnosis, mathematics is the main basis of the scientific development of health, by the applications of the science of numbers in the medical field. The geometric characteristics extraction module, obtains the information of each image and the classification module, allows, from the characteristics of the magnetic resonance images (T2 - DWI - ADC) obtained in the previous module, to place the image inside of a class, providing you with an encoding. For this implementation, a study of different characteristics that can be obtained from the images and a learning algorithm that allows classifying them after obtaining their content is carried out.

Acknowledgments

Telemedicine Research Group (GITEM ++), Universidad Distrital Francisco José de Caldas. Python 3.7.1

References

[1] K. Kitajima, Y. Kaji, Y. Fukabori, K. Yoshida, N. Suganuma and K. Sugimura, "Prostate cancer detection with 3 T MRI: Comparison of diffusion-weighted imaging and dynamic contrast-enhanced MRI in combination with T2-weighted imaging", *Journal of Magnetic Resonance Imaging*, vol. 31, no. 3, pp. 625-631, 2010. https://doi.org/10.1002/jmri.22075

[2] S. Liu, H. Zheng, Y. Feng and W. Li, "Prostate cancer diagnosis using deep learning with 3D multiparametric MRI", Spie Medical Imaging 2017: Computer-Aided Diagnosis, 2017. https://doi.org/10.1117/12.2277121

[3] H. Lamadrid Figueroa, "Cáncer de próstata: Resultados del estudio de la Carga Global de la Enfermedad", p. 35, 2015.

[4] G. Litjens, O. Debats, J. Barentsz, N. Karssemeijer and H. Huisman, "Computer-Aided Detection of Prostate Cancer in MRI", *IEEE Transactions on Medical Imaging*, vol. 33, no. 5, pp. 1083-1092, 2014. https://doi.org/10.1109/TMI.2014.2303821

[5] F. Gómez and R. Gaston, "Cáncer de Próstata", 2015. [Online]. Available at: https://www.icua.es/urologia-avanzada/cancer-de-prostata/

[6] C. Hoeks et al., "Prostate Cancer: Multiparametric MR Imaging for Detection, Localization, and Staging", *Radiology*, vol. 261, no. 1, pp. 46-66, 2011. https://doi.org/10.1148/radiol.11091822

[7] American Society Clinical Oncology, "Cáncer de próstata: diagnóstico", 2018. [Online]. Available at: https://www.cancer.net/es/tipos-de-c%C3%A1ncer/c%C3%A1ncer-de-pr%C3%B3stata/diagn%C3%B3stico

[8] D. Mitchell and M. Cohen, "MRI principles". Philadelphia, Pa.: Saunders, 2004.

[9] T. Hambrock et al., "Prospective Assessment of Prostate Cancer Aggressiveness Using 3-T Diffusion-Weighted Magnetic Resonance Imaging–Guided Biopsies Versus a Systematic 10-Core Transrectal Ultrasound Prostate Biopsy Cohort", *European Urology*, vol. 61, no. 1, pp. 177-184, 2012. https://doi.org/10.1016/j.eururo.2011.08.042

[10] K. Nguyen, B. Sabata and A. Jain, "Prostate cancer grading: Gland segmentation and structural features", *Pattern Recognition Letters*, vol. 33, no. 7, pp. 951-961, 2012. https://doi.org/10.1016/j.patrec.2011.10.001

[11] R. Christian, F. Juan and M. Alejandro, "Detección precoz de cáncer de próstata: Controversias y recomendaciones actuales", *Revista Médica Clínica Las Condes*, vol. 29, no. 2, pp. 128-135, 2018. https://doi.org/10.1016/j.rmclc.2018.02.013

[12] K. Clark et al., "The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository", *Journal of Digital Imaging*, vol. 26, no. 6, pp. 1045-1057, 2013. https://doi.org/10.1007/s10278-013-9622-7

[13] I. Chan et al., "Detection of prostate cancer by integration of line-scan diffusion, T2-mapping and T2-weighted magnetic resonance imaging; a multichannel statistical classifier", *Medical Physics*, vol. 30, no. 9, pp. 2390-2398, 2003.

[14] N. Karssemeijer, J. Otten, H. Rijken and R. Holland, "Computer aided detection of masses in mammograms as decision support", *The British Journal of Radiology*, vol. 79, no. 2, pp. 123-126, 2006. https://doi.org/10.1259/bjr/37622515

[15] A. Firjani et al., "A diffusion-weighted imaging based diagnostic system for early detection of prostate cancer", *Journal of Biomedical Science and Engineering*, vol. 6, no. 3, pp. 346-356, 2013. https://doi.org/10.4236/jbise.2013.63a044

[16] M. Stumpe, "Improved Grading of Prostate Cancer Using Deep Learning", 2018. [Online]. Available at: https://ai.googleblog.com/2018/11/improved-grading-of-prostate-cancer.html

[17] N. A. Ramírez Pérez, E. G. Vargas and O. M. Forero Cuellar, "Supervised Classi ers of Prostate Cancer from Magnetic Resonance Images in T2 Sequences", 14th Iberian Conference on Information Systems and Technologies (CISTI), Coimbra, Portugal, 2019. https://doi.org/10.23919/CISTI.2019.8760647

[18] E. Pietka, "Advances in Intelligent Systems and Computing", pp. 295–296, 2012.

[19] D. K. Jones, "Diffusion MRI: theory, methods, and applications", 2010, Oxford University Press. https://doi.org/10.1093/med/9780195369779.001.0001

[20] P. Vos, T. Hambrock, J. Barenstz and H. Huisman, "Computer-assisted analysis of peripheral zone prostate lesions using T2-weighted and dynamic contrast enhanced T1-weighted MRI", *Physics in Medicine and Biology*, vol. 55, no. 6, pp. 1719-1734, 2010. https://doi.org/10.1088/0031-9155/55/6/012

[21] M. P. Leonor and L. V. Tercero, "Inteligencia artificial en el cáncer de próstata", 2018. [Online]. Available at: http://www.die.upm.es/sites/default/files/files/Congreso_IMA2_2018-19/Poster_Penas_Vizoso.pdf

[22] Y. Ba tanlar and M. Özuysal, "Introduction to Machine Learning", *miRNomics: MicroRNA Biology and Computational Analysis*, vol. 1107, pp. 105-128, 2013. https://doi.org/10.1007/978-1-62703-748-8_7