International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:13, No:3, 2019

Performance Analysis of Search Medical Imaging Service on Cloud Storage Using Decision Trees

González A. Julio, Ramírez L. Leonardo, Puerta A. Gabriel

Abstract**—**Telemedicine services use a large amount of data, most of which are diagnostic images in Digital Imaging and Communications in Medicine (DICOM) and Health Level Seven (HL7) formats. Metadata is generated from each related image to support their identification. This study presents the use of decision trees for the optimization of information search processes for diagnostic images, hosted on the cloud server. To analyze the performance in the server, the following quality of service (QoS) metrics are evaluated: delay, bandwidth, jitter, latency and throughput in five test scenarios for a total of 26 experiments during the loading and downloading of DICOM images, hosted by the telemedicine group server of the Universidad Militar Nueva Granada, Bogotá, Colombia. By applying decision trees as a data mining technique and comparing it with the sequential search, it was possible to evaluate the search times of diagnostic images in the server. The results show that by using the metadata in decision trees, the search times are substantially improved, the computational resources are optimized and the request management of the telemedicine image service is improved. Based on the experiments carried out, search efficiency increased by 45% in relation to the sequential search, given that, when downloading a diagnostic image, false positives are avoided in management and acquisition processes of said information. It is concluded that, for the diagnostic images services in telemedicine, the technique of decision trees guarantees the accessibility and robustness in the acquisition and manipulation of medical images, in improvement of the diagnoses and medical procedures in patients.

Keywords**—**Cloud storage, decision trees, diagnostic image, search, telemedicine.

I. Introduction

CLOUD computing is a concept that has created a great number of possibilities for different organizations focused on the supply and offer of health care, due to its flexibility as a rising IT service [1]. Besides this, Cloud can be considered as an essential mechanism for the health sector due to its scalability, adaptability, connectivity and accessibility, among other high-performance characteristics that improve the efficiency and quality of health care [2]-[4].

The volume of medical information that is generated every day has occasioned a problem regarding its management, handling and storage. This is caused by the high flow of data, the type of data handled and the attention that must be applied to this type of information [5].

The proper approach to large volumes of data has turned into a challenging task that requires both a conceptual and a computational infrastructure, such as the implementation of different data analysis mechanisms that guarantee the quality and efficiency in the treatment of data and information [6], [7].

The data generated in the health sector must be managed and handled with the greatest possible priority, since this is not information that only contains numerical, financial or statistical data, but is also a type of information that, for example, can help a doctor make choices during a procedure a patient will be subjected to (considering that the health and wellness of the patient are involved). For this purpose, the storage systems in the cloud must provide the tools that will allow the proper handling and management of data in the health sector, guaranteeing the availability, accessibility and strength of the service, as well as the QoS metrics for the Cloud design, development, deployment and operation [8], [9]. On the other hand, the integration of data mining techniques in daily activities has been involved in the advancement and development of information management [10]. The provision of health care is improved when the extraction of medical data has an answer and specific purpose in said field. In general, health care processes generate large amounts of information and specific patterns that may be discovered through different types of algorithms and data mining techniques, that at the same time, can improve the quality of the processes focused on health care [11].

This article has the purpose of analyzing the offer level of a service from the Cloud server belonging to the TIGUM telemedicine group from the Universidad Militar Nueva Granada, based on the QoS metrics, guaranteeing availability, accessibility and strength parameters for the service, to subsequently apply data analysis mechanisms that will help in the search and timely delivery, and control of the information stored in the repository server (Cloud) [12]. Basic concepts such as quality of service (QoS), quality of experience (QoE) and service level agreements (SLA) are discussed in the management models from Cloud services, applying them to the management and handling of information offered to the user. Accomplishing efficiency in the management of Cloud architectures eases the handling and acquisition of software, hardware, growths and new developments.

Considering what we described before, the research problem is focused on: How the accurate selection of a cloud

Gonzalez. A. Julio is a research assistant of TIGUM research Group of Universidad Militar Nueva Granada, Bogotá, Carrera 11 No.101-80, Colombia (phone: +51 3188808889; e-mail: u1401109@unimilitar.edu.co).

Ramírez L. Leonardo is a researcher and director of TIGUM research group of Universidad Militar Nueva Granada, Bogotá, Carrera 11 No.101-80, Colombia (phone: +51 0316500000; e-mail: tigum@unimilitar.edu.co).

Puerta A. Gabriel is PhD student and researcher of TIGUM research group of Universidad Militar Nueva Granada, Bogotá, Carrera 11 No.101-80, Colombia (phone: +51 3115987518; e-mail: u7700110@unimilitar.edu.co).

management model and a data analysis mechanism can contribute to the solution of accessibility, strength, availability, ubiquity and reliability issues in the collection and management of data in a cloud service. For this purpose, we implemented a cloud service management model for the analysis of metadata from the health sector, through the performance evaluation of the deployment and operation of said service applied the proper data mining technique for the analysis of metadata and innovate with the use of IoT [13], [14].

II. Method

A. Methodology

The methodology implemented in the development of this paper was a descriptive and analytical methodology. Descriptive methodology is understood as the process of searching for information and collecting data related to previous works to determine the factors and parameters that should be considered to conduct the project, as well as the definition of the QoS metrics for the analysis of the performance, the offer level of the cloud system and the selection of the decision trees data mining technique.

Regarding the analytical methodology, we made a comparative study using statistical results between two information gathering methods, one applied the decision trees data mining technique to the Cloud server and the other did not, in other words, it was done using a sequential search. This was done to determine which search method is more efficient and helps solve the optimization issues related to the search times, optimize the resources and avoid false positives when downloading medical images from said server.

B. Preliminary Stage

As a preliminary stage, a background search was conducted, where we elaborated a state of the art with different authors that have studied Cloud Computing, data analysis and IoT focused on the health sector was made. We documented 20 authors with the corresponding publication tile, year of publication, summary, key words, conclusions and references for their review, in order to identify the metrics for Cloud Computing focused in health, identify which of these have developed research processes and determine the research gaps, which can be examined further for future works.

C. Selection of the Monitoring Tool

We evaluated different network monitoring tools that were able to approach the tests and evaluation processes for the offered cloud service. A comparative table was made using the specific characteristics from different tools (Monitis, Nagios, Zbbix, etc.), for the development and fulfillment of the objectives from the work. The selected tool was "Pandora FMS", since it met the characteristics required by the project, such as the measurement of QoS metrics, free software, Linux implementation, etc.

D. Work Environment

The TIGUM telemedicine group provided a work architecture or environment, where most of the development of the study was focused in. This architecture is shown in Fig. 2.

Fig. 1 PandoraFMS monitoring tool logo

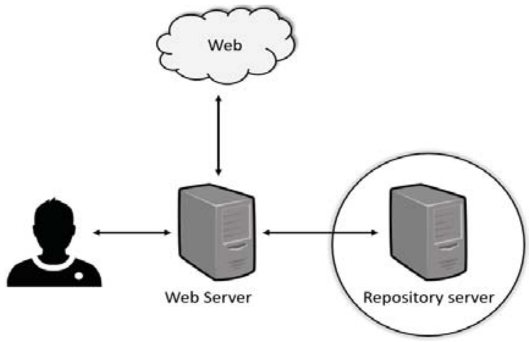


Fig. 2 Work environment from the TIGUM group

Mainly, the environment is constituted by a web server, connected directly to the repository or Cloud service. These servers are based and configured in Linux/Debian and contain characteristics related to specific processing, RAM memory, storage, etc., that can fulfill the objectives and purposes they were acquired for and implemented in the TIGUM group. This focused on the management of the repository server (cloud), which was responsible for the management, handling and storage of the information, as well as for its backup system. The existence and importance of the repository cloud server justified on the fact that any process is related to the service offered to the final user through the web server, is conducted through in it. Additionally, the TIGUM group has a database, where it stores all the necessary information for its subsequent treatment and fulfillment of the project's objectives. This database is designed in MySQL and supported by PhpMyAdmin via the route http://tigum.umng.edu.co/phpmyadmin/

III. Development

A. Parameters

For the development, we guaranteed the understanding and importance of the QoS metrics and the parameters that had to be considered, to conduct the accurate implementation of the management methodologies, thus, fulfilling the objectives from this project. Hereunder, said parameters and metrics are explained with their corresponding impact:

1) Availability: The information managed and administered must fulfill the requirement of availability, which means that doctors, patients and authorized personnel must be able to access the information in a quick way and it must also fulfill the ubiquity criteria (present at all times and places).
2) Confidentiality: The information handled in the TIGUM repository server must fulfill the confidentiality criterion,

since the access to a patient's information must be exclusive to the competent personnel and explicitly required in the evaluation and analysis process to avoid its inappropriate use, disclosure and even manipulation in a harmful way, which will affect its integrity

3) Integrity: As is well known, any information that is required and handled must comply with the integrity criterion, since it must be accurate for its analysis and subsequent decision-making process. In the health sector, the integrity of the information can determine, for example, the difference between the medical image from a healthy patient and the one from a sick patient.

B. QoS Metrics

For the fulfillment of the management model for cloud services that will be implemented, we worked with a set of metrics that guaranteed its correct functioning, implementation and handling. The metrics we worked with were:

Delay: It is the relation between the times taken for the transmission of the information or packages within a network. It is calculated with the difference between the total reception time (distribution and processing) and the total transmission time for the information. This metric determines how fast or how slow is the traffic of the information (since, the slower the process, the bigger is the probability of loss of information). Along with the concept of latency, this metric guarantees the reliability of the service (should be near 0).

$$\text{Delay} = Arrival\ time - Exit\ time \quad (1)$$

Latency: It is the expected value for how long it takes for a package to conduct the transmission. It is also defined as the delay average (delay). This metric determines the quickness of the service (must be near 0).

$$\text{Latency} = \sum_{j=0}^{n} \frac{Delay(j)}{n} \text{ or } E(Delay) \quad (2)$$

Bandwidth: It determines the amount of information that can be transmitted simultaneously on the channel. It is worth mentioning that for this metric, there is a relation between the bandwidth that is theoretically offered by the service (agreed in the SLA) and the one that is available. For the accurate selection of the bandwidth required by the system, we conducted the following procedure:

$$\text{BW Required} = \frac{\#\, Packages * Lenght\ of\ the\ package}{Time} \quad (3)$$

Bandwidth metric determines the efficiency of the service that will be delivered and is calculated with:

$$\text{Efficiency} = \frac{Theoretical\ bandwidth}{Maximum\ bandwidth} * 100\% \quad (4)$$

The efficiency of the service is agreed on the service level agreement and must be guaranteed in an over 80% capacity with respect to the maximum it can have (according to the maximum bandwidth).

Throughput: It is the quantity of data that effectively arrive in a unit of time, in other words, the packages that were correctly received. This metric guarantees the efficiency of the offered service and must be the same as the amount of data generated in the transmission

$$\text{TH} = \frac{Received\ packages}{Time} \quad (5)$$

Generated traffic: The generated traffic is the sum of the exit packages and the lost packages. Since it is medical information, the generated traffic must be the same as the received traffic, indicating, in the same way, the efficiency of the service.

$$\text{Nall} = \#Exit\ pack. + \#Lost\ pack. \quad (6)$$

Lost: It is the number of lost packages during the information transmission process. Said metric must be near 0 to guarantee an efficient and reliable system.

$$\text{Lost} = \frac{\#Lost\ packages}{Time} \quad (7)$$

Jitter: It is the difference in the absolute delay value of a package and the expected transmission time (latency), which is calculated in the following way:

$$\text{Jitter} = |Package\ delay - Latency| \quad (8)$$

This metric determines the stability of the service or system and must be a value near 0, in other words, the offered service must guarantee a perfect stability (or an optimal one in a worst-case scenario).

C. Test Scenarios

Once the metrics needed for this first stage of the Cloud service management model were determined, different test scenarios were established to determine the speed, reliability, efficiency, effectiveness and stability criteria for the telemedicine service that will be implemented, as well as the development and the level of offering handled within the service. The corresponding measurement for each of the tests was conducted using PandoraFMS and Wireshark to verify the correct performance of each of the metrics.

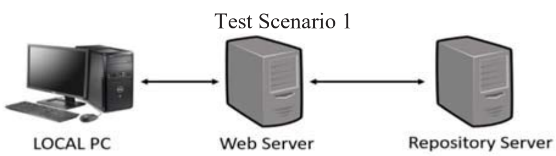


Fig. 3 Architecture for test scenario 1

To analyze the performance WITHOUT the load of users to the system, we used a PC connected LOCALLY to the web and repository servers (through LAN), with the purpose of conducting a series of Echo Request and Echo Reply tests, which provide delay values for the system. This "PING" test was conducted, in the first instance, in the following way:

1) PING from LOCAL PC to PRIVATE IP from the

Repository server (Internal: 172.16.9.26).
2) PING from the WEB Server to PRIVATE IP from the Repository server (Internal: 172.16.9.26).
3) PING from LOCAL PC to the PRIVATE IP from the web server (Internal: 172.16.9.28).
4) PING from the LOCAL PC to the PUBLIC IP from the web server (External: 190.60.233.52).
5) PING from the WEB Server to the PUBLIC IP from the web server (External: 190.60.233.52).
6) PING from the WEB Server to the PRIVATE IP from the web server (External: 190.60.233.52).

These tests were conducted with the purpose of analyzing the performance of the servers under 'perfect' conditions, it means, without user traffic or data flow. It is worth mentioning that the test was not conducted on the public address of the repository server, precisely because it does not have access to said address, since it is a repository server.

The tests conducted using 'Echo' had a duration of 10 minutes, which provided values for the size, time and number of the package. 600 samples were obtained, which allowed us to calculate the delay, latency, jitter, general and instantaneous throughput, and efficiency and we calculated an average for each of the QoS metrics for the analysis of the servers' performance.

Test Scenario 2 (JMETER)

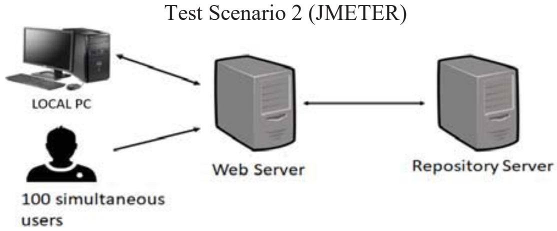


Fig. 4 Architecture for test scenario 2

With the JMETER software, the load and stress test of the server was carried out through the simultaneous connection of users to the servers this in order to perform a visualization, connection and conducting of the request (using a script) to the servers simultaneously. These load and stress tests were conducted with 1, 10, 100, 200 and 300 users, thus determining that with 100 users, the error percentage was less than 1% in the servers, concluding that 100 users are the maximum user load that the servers were able to stand.

Once the maximum amount of users that the server can stand simultaneously was determined, the same 'echo request' and 'echo reply' tests were conducted using 'ping' as in scenario 1 (Fig. 1). The load of 100 users was applied simultaneously in order to determine the performance and behavior of the server in loaded way (which must guarantee an efficiency over 80% to be considered as an optimal service). In the same way as in test scenario 1, the test lasted 10 minutes and the values corresponding to the QoS metrics were calculated.
1) PING from the WEB Server to the PRIVATE IP from the Repository server (Internal: 172.16.9.26).
2) PING from the LOCAL PC to the PRIVATE IP from the web server (Internal: 172.16.9.28).
3) PING from the LOCAL PC to the PUBLIC IP from the web server (External: 190.60.233.52).
4) PING from the WEB Server to the PRIVATE IP from the web server (External: 190.60.233.52).
5) PING from the WEB Server to the PUBLIC IP from the web server (External: 190.60.233.52).
6) PING from the LOCAL PC to the PRIVATE IP from the Repository server (Internal: 172.16.9.26).

Test Scenario 3

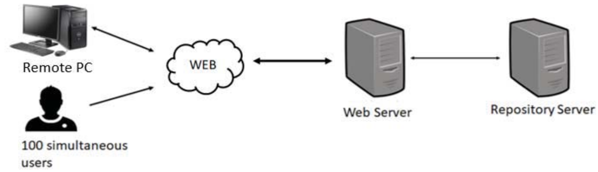


Fig. 5 Architecture for test scenario 3

In this test scenario, the same procedure from previous tests was conducted (Fig. 4). This time, the tests were conducted from a PC connected through internet, thus determining the behavior and performance of the server with a remote user. In this case the conducted tests were the following:
1) PING from the WEB Server to the PRIVATE IP from the repository server (Internal: 172.16.9.26).
2) PING from the REMOTE PC to the PUBLIC IP from the web server (External: 190.60.233.52).

Said remote test provided time, package number and package size values, which allowed us to determine the QoS metrics mentioned before. Since the connection to the servers was conducted remotely, there were losses regarding the performance of the server. Nevertheless, we obtained results that fulfilled the minimum standards needed to guarantee the availability, accessibility and strength parameters according to the service being offered.

Test Scenario 4

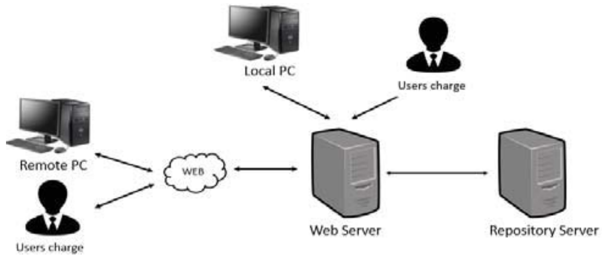


Fig. 6 Architecture for test scenario 4

For this scenario, the corresponding Echo tests were conducted, but this time, the load of users to the servers was conducted simultaneously, in other words, the users were loaded remotely and locally, with the purpose of determining the performance and behavior of the servers when the user wants to do a consult and wants to download medical information. The corresponding PING tests were conducted with the help of a LOCAL PC and a REMOTE PC, to obtain the QoS metrics and subsequently conduct the calculation, graphs and documentation for the analysis of the results. The conducted tests were the following:

1) PING from the LOCAL PC to the PRIVATE IP from the web server (Internal: 172.16.9.28).
2) PING from the LOCAL PC to the PUBLIC IP from the web server (External: 190.60.233.52).
3) PING from the LOCAL PC to the PRIVATE IP from the Repository server (Internal: 172.16.9.26).
4) PING from the WEB Server to the PRIVATE IP from the web server (External: 190.60.233.52).
5) PING from the WEB Server to the PUBLIC IP from the web server (External: 190.60.233.52).
6) PING from the WEB Server to the PRIVATE IP from the Repository server (Internal: 172.16.9.26).
7) PING from the REMOTE PC to the PUBLIC IP from the web server (External: 190.60.233.52).

Cloud Test Scenario

In the cloud test scenario, the refusal and offer level from the service provided by Cloud in the repository server was determined. The search of the 10 DICOM '.dcm' images from different specialties from the medicine sector, such as mastology, neurology, chiropractic, etc., was conducted through DICOM image banks available online (example: http://dicomlibrary.com). Once the images were obtained, they were loaded to the repository service (cloud) in specific routes with the purpose of storing them and subsequently manage and manipulate them (Fig. 7).

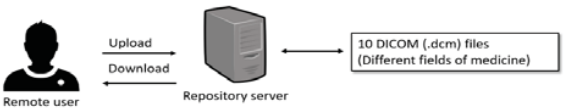


Fig. 7 Architecture from the cloud scenario

Parent Directory		-
1.dcm	2018-06-14 08:37	203K
2.dcm	2018-06-14 08:37	515K
3.dcm	2018-06-14 08:37	514K
4.dcm	2018-06-14 08:37	89K
5.dcm	2018-06-14 08:37	55K
6.dcm	2018-06-14 08:37	2.0M
7.dcm	2018-06-14 08:37	130K
8.dcm	2018-06-14 08:37	798K
9.dcm	2018-06-14 08:37	130K
10.dcm	2018-06-14 08:37	513K

Fig. 8 Script for the download of images from the repository server

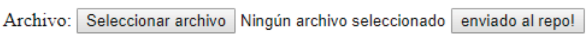


Fig. 9 Script for the upload of images to the repository server

Through the creation of two scripts based on PHP, 2 instances were created, where the last test scenario was implemented. This scenario consists of the measurement of QoS metrics, pretending to be a remote user that accesses the service and makes the upload and download request for the medical images mentioned before. Hereunder, the scripts for the upload and download of images from the repository server are shown, which was accessed through the webpage: www.tigum.umng.edu.co/tigum/pruebascloud/imagenes (Figs. 8 and 9)

Once the scripts were created, the download of 10 DICOM images was conducted, while simultaneously, the QoS metrics were measured with Wireshark, which pretended to be a doctor that needed to consult and download the images located in the repository server from the TIGUM group. This process was repeated with the upload images script, where it pretended to be a doctor that wanted to upload diagnostic images from a patient to the server.

1) TCP – File upload
2) HTTP – File upload
3) TCP – File download
4) HTTP - File download

Once the results for the time, package size and package number metrics were obtained, the corresponding QoS metrics and the averages obtained in the upload and download of the 10 images were calculated for their subsequent graphic representation and documentation. This process provided satisfactory results regarding the level of offer from the repository server, considering the type of information being handled within it (diagnostic images). It is worth mentioning that the conducted tests were made in different time slots to obtain results with different parameters such as, network traffic, user of resources, etc.

D.Data Mining

Fig. 10 WEKA data mining tool

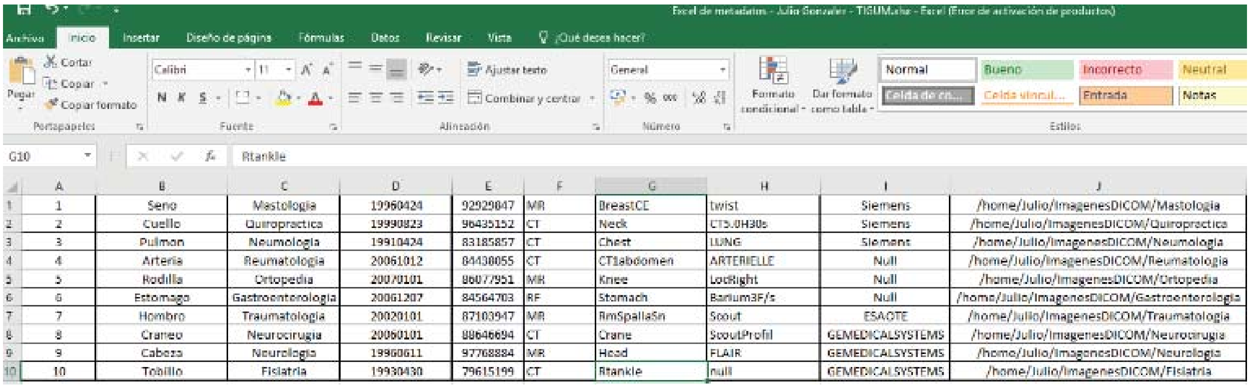
For the development and implementation of a data mining technique that was able to adapt and work in the Cloud server, a study was made on different data mining techniques such as k-NN (k-Nearest Neighbors), decision tree and linear regression. It was determined that the proper data mining technique was the decision tree, due to its prediction model based on artificial intelligence that uses logical construction diagrams. This is a technique focused on the representation of the behavior and making successive decisions for the solution of a problem and obtaining an expected result. For the implementation of this technique, the WEKA data mining tool was used, which is a software platform from Machine Learning with a Java basis, developed by the University of Waikato.

The metadata from the DICOM images previously stored in the Could server, through an online visualizer (https://www.dicomlibrary.com/) from said format were extracted. The metadata extraction process made way for their subsequent documentation using an Excel table (Fig. 11), which identified the following aspects:

1) ID
2) Type of image
3) Specialty

4) Date of the study
5) Citizenship card (Fictional, since the used images belong to anonymous people)
6) Modality
7) Description of the study
8) Description of the set
9) Manufacturer
10) Link route in the repository server (location)

These metadata were recorded in WEKA through an '.arff' file, modified with specific parameters that stipulate the instances and attributes that should be evaluated by the decision tree. Once they were registered in WEKA, it shows us the table information (Fig. 12), for the subsequent calculation of the decision tree.



A	B	C	D	E	F	G	H	I	J
1	Seno	Mastologia	19960424	92929847	MR	BreastCE	twist	Siemens	/home/Julio/ImagenesDICOM/Mastologia
2	Cuello	Quiropractica	19990823	96435152	CT	Neck	CT5.0H30s	Siemens	/home/Julio/ImagenesDICOM/Quiropractica
3	Pulmon	Neumologia	19910424	83185857	CT	Chest	LUNG	Siemens	/home/Julio/ImagenesDICOM/Neumologia
4	Arteria	Reumatologia	20061012	84438055	CT	CT1abdomen	ARTERIELLE	Null	/home/Julio/ImagenesDICOM/Reumatologia
5	Rodilla	Ortopedia	20070101	86077951	MR	Knee	LocRight	Null	/home/Julio/ImagenesDICOM/Ortopedia
6	Estomago	Gastroenterologia	20061207	84564703	RF	Stomach	Barium3F/s	Null	/home/Julio/ImagenesDICOM/Gastroenterologia
7	Hombro	Traumatologia	20020101	87103947	MR	RmSpallaSn	Scout	ESAOTE	/home/Julio/ImagenesDICOM/Traumatologia
8	Craneo	Neurocirugia	20060101	88646694	CT	Crane	ScoutProfil	GEMEDICALSYSTEMS	/home/Julio/ImagenesDICOM/Neurocirugia
9	Cabeza	Neurologia	19960611	97768884	MR	Head	FLAIR	GEMEDICALSYSTEMS	/home/Julio/ImagenesDICOM/Neurologia
10	Tobillo	Fisiatria	19930430	79615199	CT	Rtankle	null	GEMEDICALSYSTEMS	/home/Julio/ImagenesDICOM/Fisiatria

Fig. 11 Table with metadata from the DICOM images

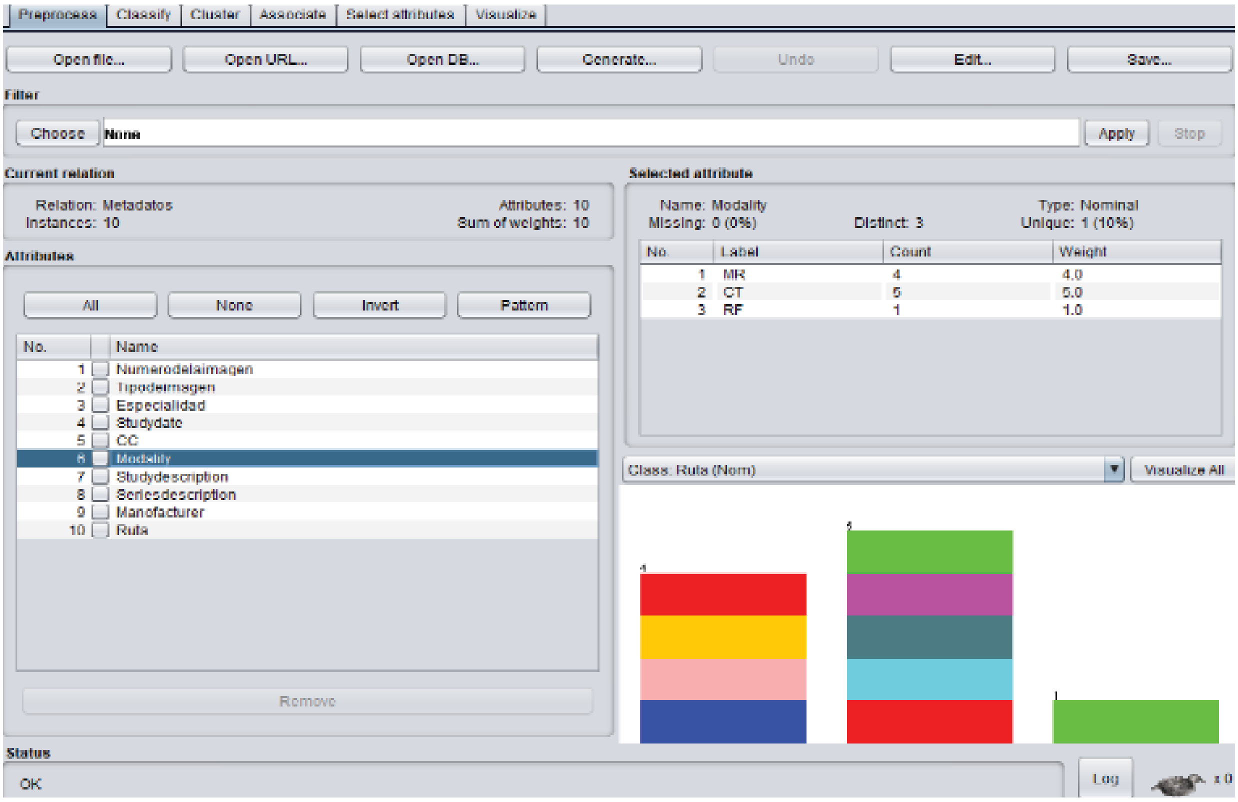


Fig. 12 WEKA with the metadata from the 10 DICOM images

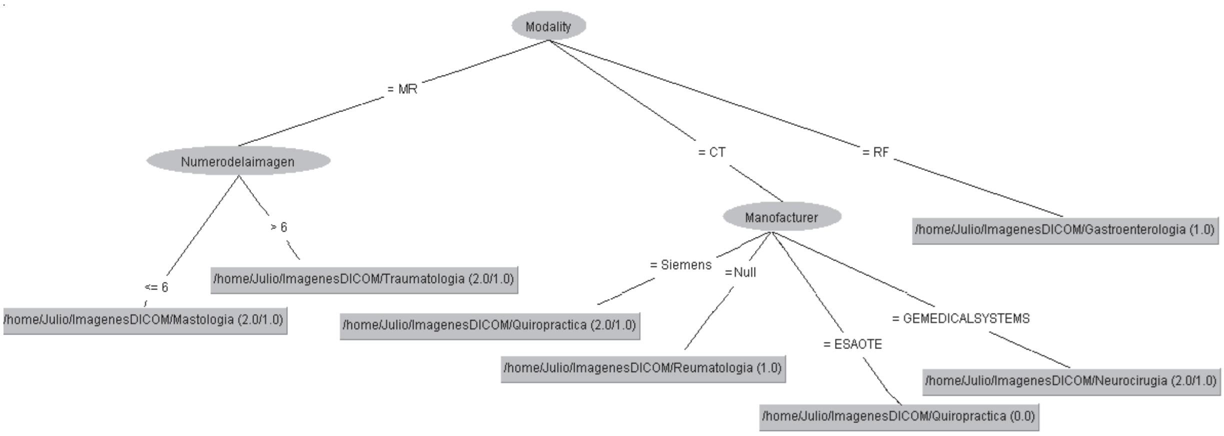
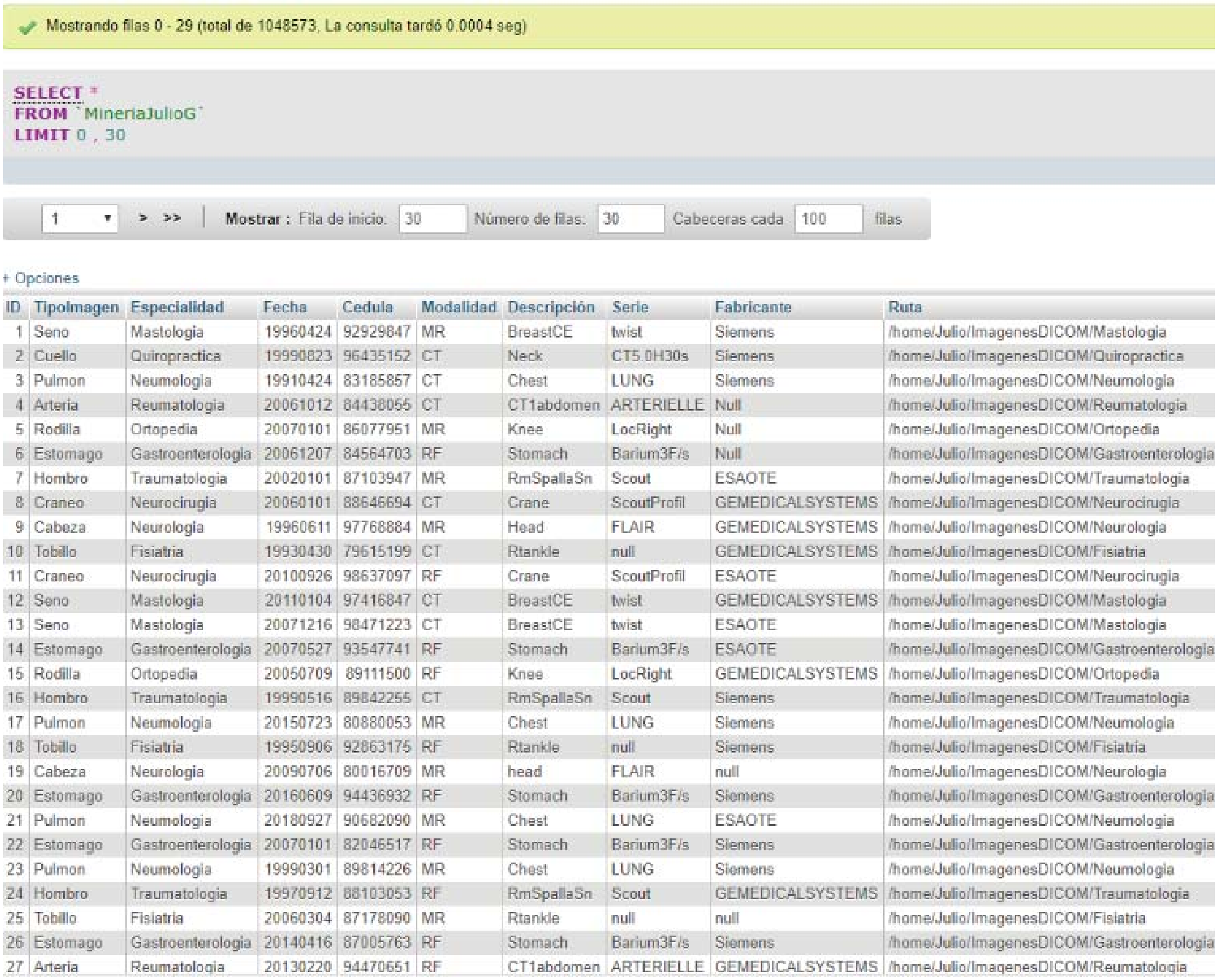


Fig. 13 Decision tree provided by Weka



ID	TipoImagen	Especialidad	Fecha	Cedula	Modalidad	Descripción	Serie	Fabricante	Ruta
1	Seno	Mastologia	19960424	92929847	MR	BreastCE	twist	Siemens	/home/Julio/ImagenesDICOM/Mastologia
2	Cuello	Quiropractica	19990823	96435152	CT	Neck	CT5.0H30s	Siemens	/home/Julio/ImagenesDICOM/Quiropractica
3	Pulmon	Neumologia	19910424	83185857	CT	Chest	LUNG	Siemens	/home/Julio/ImagenesDICOM/Neumologia
4	Arteria	Reumatologia	20061012	84438055	CT	CT1abdomen	ARTERIELLE	Null	/home/Julio/ImagenesDICOM/Reumatologia
5	Rodilla	Ortopedia	20070101	86077951	MR	Knee	LocRight	Null	/home/Julio/ImagenesDICOM/Ortopedia
6	Estomago	Gastroenterologia	20061207	84564703	RF	Stomach	Barium3F/s	Null	/home/Julio/ImagenesDICOM/Gastroenterologia
7	Hombro	Traumatologia	20020101	87103947	MR	RmSpallaSn	Scout	ESAOTE	/home/Julio/ImagenesDICOM/Traumatologia
8	Craneo	Neurocirugia	20060101	88646694	CT	Crane	ScoutProfil	GEMEDICALSYSTEMS	/home/Julio/ImagenesDICOM/Neurocirugia
9	Cabeza	Neurologia	19960611	97768884	MR	Head	FLAIR	GEMEDICALSYSTEMS	/home/Julio/ImagenesDICOM/Neurologia
10	Tobillo	Fisiatria	19930430	79615199	CT	Rtankle	null	GEMEDICALSYSTEMS	/home/Julio/ImagenesDICOM/Fisiatria
11	Craneo	Neurocirugia	20100926	98637097	RF	Crane	ScoutProfil	ESAOTE	/home/Julio/ImagenesDICOM/Neurocirugia
12	Seno	Mastologia	20110104	97416847	CT	BreastCE	twist	GEMEDICALSYSTEMS	/home/Julio/ImagenesDICOM/Mastologia
13	Seno	Mastologia	20071216	98471223	CT	BreastCE	twist	ESAOTE	/home/Julio/ImagenesDICOM/Mastologia
14	Estomago	Gastroenterologia	20070527	93547741	RF	Stomach	Barium3F/s	ESAOTE	/home/Julio/ImagenesDICOM/Gastroenterologia
15	Rodilla	Ortopedia	20050709	89111500	RF	Knee	LocRight	GEMEDICALSYSTEMS	/home/Julio/ImagenesDICOM/Ortopedia
16	Hombro	Traumatologia	19990516	89842255	CT	RmSpallaSn	Scout	Siemens	/home/Julio/ImagenesDICOM/Traumatologia
17	Pulmon	Neumologia	20150723	80880053	MR	Chest	LUNG	Siemens	/home/Julio/ImagenesDICOM/Neumologia
18	Tobillo	Fisiatria	19950906	92863175	RF	Rtankle	null	Siemens	/home/Julio/ImagenesDICOM/Fisiatria
19	Cabeza	Neurologia	20090706	80016709	MR	head	FLAIR	null	/home/Julio/ImagenesDICOM/Neurologia
20	Estomago	Gastroenterologia	20160609	94436932	RF	Stomach	Barium3F/s	Siemens	/home/Julio/ImagenesDICOM/Gastroenterologia
21	Pulmon	Neumologia	20180927	90682090	MR	Chest	LUNG	ESAOTE	/home/Julio/ImagenesDICOM/Neumologia
22	Estomago	Gastroenterologia	20070101	82046517	RF	Stomach	Barium3F/s	Siemens	/home/Julio/ImagenesDICOM/Gastroenterologia
23	Pulmon	Neumologia	19990301	89814226	MR	Chest	LUNG	Siemens	/home/Julio/ImagenesDICOM/Neumologia
24	Hombro	Traumatologia	19970912	88103053	RF	RmSpallaSn	Scout	GEMEDICALSYSTEMS	/home/Julio/ImagenesDICOM/Traumatologia
25	Tobillo	Fisiatria	20060304	87178090	MR	Rtankle	null	null	/home/Julio/ImagenesDICOM/Fisiatria
26	Estomago	Gastroenterologia	20140416	87005763	RF	Stomach	Barium3F/s	Siemens	/home/Julio/ImagenesDICOM/Gastroenterologia
27	Arteria	Reumatologia	20130220	94470651	RF	CT1abdomen	ARTERIELLE	GEMEDICALSYSTEMS	/home/Julio/ImagenesDICOM/Reumatologia

Fig. 14 Database with the metadata from the DICOM images (1 million)

Afterwards, the decision tree J48 was generated, which is a free implementation in Java of the C4.5 algorithm, which uses the concept of entropy of the information for the selection of variables that conduct the best classification process (in this case, the metadata from the medical images). Through the instances and attributes stipulated in WEKA, the ways to arrive to a specific characteristic from the metadata were determined. The objective of applying this mining data technique is to allow

International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:13, No:3, 2019

us to arrive to the route where the DICOM images are located in the Cloud server, thus, the tree provided by WEKA shows how to conduct the search based on the metadata to arrive to said route. It is worth mentioning that WEKA worked as a guide to apply the decision tree, but the filtering system was arbitrarily modified to arrive to the specific required route.

To verify and determine the effectiveness of the factors that were going to be evaluated with the data mining technique, it was necessary to extend the amount of information to which the mining technique was applied to. For this purpose, the documentation of 10 DICOM images was expanded to 1 million images, each one with its respective metadata generated randomly, keeping the sense of certain metadata, such as route, specialty, description of the study, route, etc. If, for example, the image is a BREAST image, it keeps metadata, such as the route where it is located, the specialty, the description and the series. (Example: "Home/July/ ImagesDICOM/Mastology"). The metadata, such as the citizenship card, date of study and modality of the images, were key for the proper application of the decision tree (given its randomness). This expansion of the information was entered to WEKA through its corresponding configuration and the corresponding decision tree was generated to verify its functioning in the development of the project. Subsequently, the metadata were entered into the database from the TIGUM group for its corresponding treatment and handling (Fig. 14).

Once the metadata were placed in the database, the design of the search method for the DICOM images (Fig. 15) was made based on the metadata extracted before. This design was made based on the results provided by WEKA and decision tree J48.

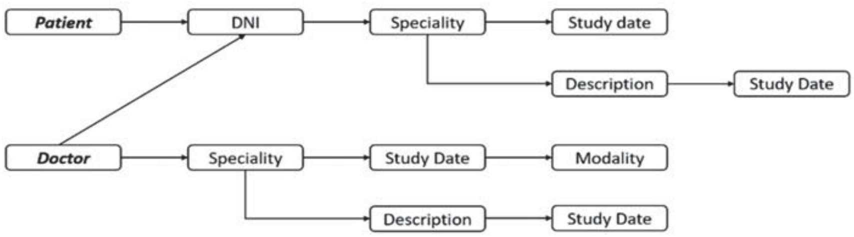


Fig. 15 Design of the search method for DICOM images

A PHP script was developed for the search of DICOM images located in the Cloud server, following the design of the search method mentioned before (Fig. 15). The script was based on a main menu where the user chose if he was a patient or a doctor that wants to search for the information. Said file was called 'search.php' and was constituted by 2 main buttons, which are responsible for redirecting the user to the corresponding search menu, as shown Z (Fig. 16).

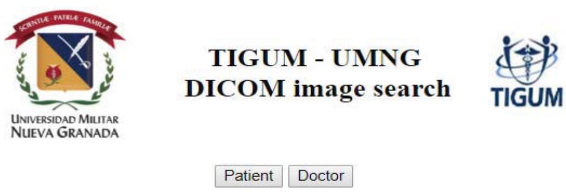


Fig. 16 Main menu for the DICOM images search script

As shown in Fig. 15, the search menu for patients 'patient.php' is constituted by four criteria for the search of information, which are:

1) DNI
2) Specialty
3) Study date
4) Description

These fields correspond to the metadata needed to fulfill the search of DICOM images, applying the 'decision trees' mining technique (Fig. 17). Likewise, the search menu for doctors 'doctor.php' is constituted by five criteria for the search of information, which are:

1) DNI
2) Specialty
3) Study date
4) Description
5) Modality

These fields correspond to the metadata needed to fulfill the search of DICOM images applying the 'decision tree' data mining technique, as shown hereunder (Fig. 18).

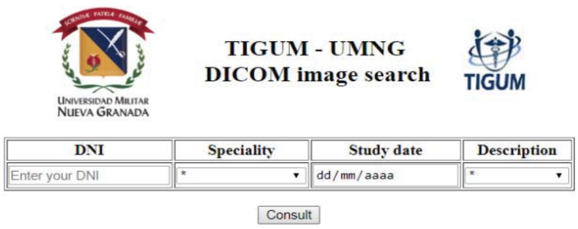


Fig. 17 Script Patient.php for the search of DICOM images

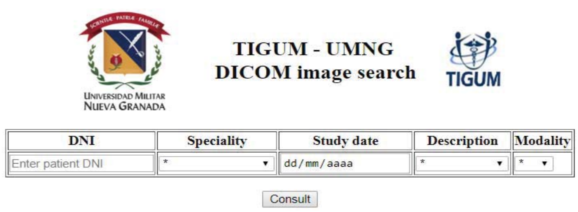


Fig. 18 Script doctor.php for the search of DICOM images

After this, the tests corresponding to the data mining technique were conducted.

The objective of implementing decision trees is to make the comparison regarding the search time and the download of the requested information by evaluating the server with the applied mining technique and without it. The factors to be considered by the comparison were the following:

Search time: When conducting the consult of a medical image located between thousands of images in a database, a decrease in the search time can present itself if a data mining technique, such as the decision tree, is applied, since it will filter the information according to its metadata, instead of conducting a sequential search without any background process.

Optimization of resources and decrease of false positives: When conducting the download of a medical history, medical images and files belonging to a specific patient, the decision tree technique allows to filter it through metadata, with the purpose of downloading only the required information, without having to download all the information stored in the database, which optimized physical and network resources (RAM, processing, bandwidth) and improved the efficiency of the service, since it minimizes the possibility of false positives (download of unwanted images in a specific process).

It is worth mentioning that to determine the improvement of the efficiency from the service regarding the decrease of false positives, the following metrics were used, which characterize said condition of the system:

Sensibility: it is the capacity the system has to detect how many searches were correct regarding the total amount of obtained result. Said metric is measured in the following way:

$$Sensitivity = \frac{TP}{TP+FN} \tag{9}$$

where TP: Number of true positives; FN: Number of false negatives.

Specificity: It is the capacity the system has to detect how many searches were wrong regarding the total amount of obtained results. Said metric is measured in the following way:

$$Specificity = \frac{TN}{TN+FP} \tag{10}$$

where TN: Number of true negatives; FP: Number of false positives.

For the implementation of the comparison tests, the sequential search of 20 random images with specific metadata (type of image, date of study, description, modality, etc.) was conducted. Since the search was conducted in a sequential way with the DNI criterion, the results of the database belonging to this criterion were obtained. This process was conducted for both the 'patient.php' script and the 'doctor.php' script, creating a total of 40 sequential search tests, which are illustrated in Figs. 19 and 20.

As shown before, when we conducted the sequential search for DICOM images located in the Cloud server, we obtained multiple results that fulfill the requested DNI criterion, but we did not obtain a specific result as expected for the consult of an image from a specific study from a patient, this is the reason why false positives appear. It is also possible to observe the search time used when conducting the consult, which allowed us to conduct the subsequent analysis of the search times applying the decision trees and the sequential search.

Finally, the search of the same 20 DICOM images located in the cloud server is conducted, this time applying the data mining technique decision trees, in other words, fulfilling the four criteria for the search of information for the 'patient.php' script and the 5 criteria for the search for the 'doctor.php' script. Said process can be observed in Figs. 21 and 22.


TIGUM - UMNG
DICOM image search

DNI	Speciality	Study date	Description
79000433	*	dd/mm/aaaa	*

Consult

Type image	Speciality	Date	DNI	Modality	Description	Serie	Manufacturer	Path
Seno	Mastologia	20150918	79000433	CT	BreastCE	twist	GEMEDICALSYSTEMS	home/Julio/ImagenesDICOM/Mastologia
Rodilla	Ortopedia	20090820	79000433	MR	Knee	LocRight	GEMEDICALSYSTEMS	home/Julio/ImagenesDICOM/Ortopedia
Seno	Mastologia	20051103	79000433	CT	BreastCE	twist	Siemens	home/Julio/ImagenesDICOM/Mastologia
Hombro	Traumatologia	20020418	79000433	RF	RmSpallaSn	Scout	Siemens	home/Julio/ImagenesDICOM/Traumatologia
Pulmon	Neumologia	20060303	79000433	RF	Chest	LUNG	ESAOTE	home/Julio/ImagenesDICOM/Neumologia
Hombro	Traumatologia	19960708	79000433	CT	RmSpallaSn	Scout	GEMEDICALSYSTEMS	home/Julio/ImagenesDICOM/Traumatologia
Hombro	Traumatologia	20010105	79000433	CT	RmSpallaSn	Scout	null	home/Julio/ImagenesDICOM/Traumatologia
Hombro	Traumatologia	20040214	79000433	RF	RmSpallaSn	Scout	Siemens	home/Julio/ImagenesDICOM/Traumatologia
Hombro	Traumatologia	20061110	79000433	MR	RmSpallaSn	Scout	GEMEDICALSYSTEMS	home/Julio/ImagenesDICOM/Traumatologia
Craneo	Neurocirugia	20061222	79000433	MR	Crane	ScoutProfil	GEMEDICALSYSTEMS	home/Julio/ImagenesDICOM/Neurocirugia
Hombro	Traumatologia	19970808	79000433	RF	RmSpallaSn	Scout	null	home/Julio/ImagenesDICOM/Traumatologia
Cabeza	Neurologia	20150820	79000433	CT	head	FLAIR	null	home/Julio/ImagenesDICOM/Neurologia
Pulmon	Neumologia	20080708	79000433	CT	Chest	LUNG	ESAOTE	home/Julio/ImagenesDICOM/Neumologia
Pulmon	Neumologia	20000109	79000433	MR	Chest	LUNG	GEMEDICALSYSTEMS	home/Julio/ImagenesDICOM/Neumologia
Seno	Mastologia	20060313	79000433	MR	BreastCE	twist	ESAOTE	home/Julio/ImagenesDICOM/Mastologia
Pulmon	Neumologia	20170402	79000433	RF	Chest	LUNG	GEMEDICALSYSTEMS	home/Julio/ImagenesDICOM/Neumologia
Seno	Mastologia	20021130	79000433	MR	BreastCE	twist	ESAOTE	home/Julio/ImagenesDICOM/Mastologia
Tobillo	Fisiatria	20070909	79000433	CT	Rtankle	null	GEMEDICALSYSTEMS	home/Julio/ImagenesDICOM/Fisiatria
Rodilla	Ortopedia	19990501	79000433	MR	Knee	LocRight	ESAOTE	home/Julio/ImagenesDICOM/Ortopedia
Cabeza	Neurologia	20001013	79000433	RF	head	FLAIR	GEMEDICALSYSTEMS	home/Julio/ImagenesDICOM/Neurologia
Cuello	Quiropractica	20170521	79000433	CT	Neck	CT5.0H30s	GEMEDICALSYSTEMS	home/Julio/ImagenesDICOM/Quiropractica
Rodilla	Ortopedia	19950223	79000433	MR	Knee	LocRight	GEMEDICALSYSTEMS	home/Julio/ImagenesDICOM/Ortopedia
Tobillo	Fisiatria	19961224	79000433	RF	Rtankle	null	ESAOTE	home/Julio/ImagenesDICOM/Fisiatria
Hombro	Traumatologia	20161228	79000433	MR	RmSpallaSn	Scout	null	home/Julio/ImagenesDICOM/Traumatologia

Time search: 0.00016403198242188

Fig. 19 Sequential search of DICOM images through the patient.php script


TIGUM - UMNG
DICOM image search

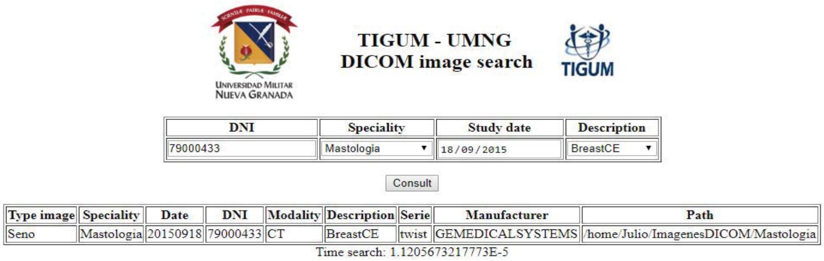
DNI	Speciality	Study date	Description	Modality
79000433	*	dd/mm/aaaa	*	*

Consult

Type image	Especiality	Date	DNI	Modality	Description	Serie	Manufacturer	Path
Seno	Mastologia	20150918	79000433	CT	BreastCE	twist	GEMEDICALSYSTEMS	/home/Julio/ImagenesDICOM/Mastologia
Rodilla	Ortopedia	20090820	79000433	MR	Knee	LocRight	GEMEDICALSYSTEMS	/home/Julio/ImagenesDICOM/Ortopedia
Seno	Mastologia	20051103	79000433	CT	BreastCE	twist	Siemens	/home/Julio/ImagenesDICOM/Mastologia
Hombro	Traumatologia	20020418	79000433	RF	RmSpallaSn	Scout	Siemens	/home/Julio/ImagenesDICOM/Traumatologia
Pulmon	Neumologia	20060303	79000433	RF	Chest	LUNG	ESAOTE	/home/Julio/ImagenesDICOM/Neumologia
Hombro	Traumatologia	19960708	79000433	CT	RmSpallaSn	Scout	GEMEDICALSYSTEMS	/home/Julio/ImagenesDICOM/Traumatologia
Hombro	Traumatologia	20010105	79000433	CT	RmSpallaSn	Scout	null	/home/Julio/ImagenesDICOM/Traumatologia
Hombro	Traumatologia	20040214	79000433	RF	RmSpallaSn	Scout	Siemens	/home/Julio/ImagenesDICOM/Traumatologia
Hombro	Traumatologia	20061110	79000433	MR	RmSpallaSn	Scout	GEMEDICALSYSTEMS	/home/Julio/ImagenesDICOM/Traumatologia
Craneo	Neurocirugia	20061222	79000433	MR	Crane	ScoutProfil	GEMEDICALSYSTEMS	/home/Julio/ImagenesDICOM/Neurocirugia
Hombro	Traumatologia	19970808	79000433	RF	RmSpallaSn	Scout	null	/home/Julio/ImagenesDICOM/Traumatologia
Cabeza	Neurologia	20150820	79000433	CT	head	FLAIR	null	/home/Julio/ImagenesDICOM/Neurologia
Pulmon	Neumologia	20080708	79000433	CT	Chest	LUNG	ESAOTE	/home/Julio/ImagenesDICOM/Neumologia
Pulmon	Neumologia	20000109	79000433	MR	Chest	LUNG	GEMEDICALSYSTEMS	/home/Julio/ImagenesDICOM/Neumologia
Seno	Mastologia	20060313	79000433	MR	BreastCE	twist	ESAOTE	/home/Julio/ImagenesDICOM/Mastologia
Pulmon	Neumologia	20170402	79000433	RF	Chest	LUNG	GEMEDICALSYSTEMS	/home/Julio/ImagenesDICOM/Neumologia
Seno	Mastologia	20021130	79000433	MR	BreastCE	twist	ESAOTE	/home/Julio/ImagenesDICOM/Mastologia
Tobillo	Fisiatria	20070909	79000433	CT	Rtankle	null	GEMEDICALSYSTEMS	/home/Julio/ImagenesDICOM/Fisiatria
Rodilla	Ortopedia	19990501	79000433	MR	Knee	LocRight	ESAOTE	/home/Julio/ImagenesDICOM/Ortopedia
Cabeza	Neurologia	20001013	79000433	RF	head	FLAIR	GEMEDICALSYSTEMS	/home/Julio/ImagenesDICOM/Neurologia
Cuello	Quiropractica	20170521	79000433	CT	Neck	CT5.0H30s	GEMEDICALSYSTEMS	/home/Julio/ImagenesDICOM/Quiropractica
Rodilla	Ortopedia	19950223	79000433	MR	Knee	LocRight	GEMEDICALSYSTEMS	/home/Julio/ImagenesDICOM/Ortopedia
Tobillo	Fisiatria	19961224	79000433	RF	Rtankle	null	ESAOTE	/home/Julio/ImagenesDICOM/Fisiatria
Hombro	Traumatologia	20161228	79000433	MR	RmSpallaSn	Scout	null	/home/Julio/ImagenesDICOM/Traumatologia

Time search: 0.00013303756713867

Fig. 20 Sequential search of DICOM images through the doctor.php script


TIGUM - UMNG
DICOM image search

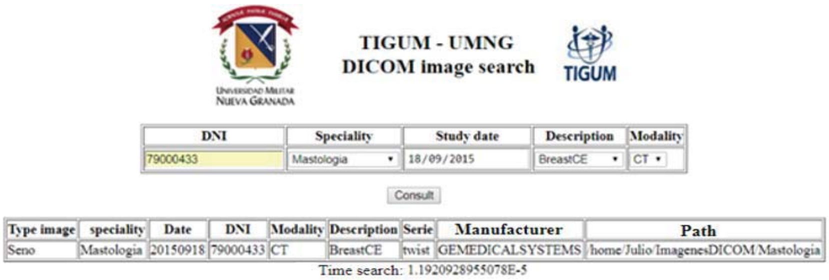
DNI	Speciality	Study date	Description
79000433	Mastologia	18/09/2015	BreastCE

Consult

Type image	Speciality	Date	DNI	Modality	Description	Serie	Manufacturer	Path
Seno	Mastologia	20150918	79000433	CT	BreastCE	twist	GEMEDICALSYSTEMS	/home/Julio/ImagenesDICOM/Mastologia

Time search: 1.1205673217773E-5

Fig. 21 Search using decision trees of the DICOM images for the patient-php script


TIGUM - UMNG
DICOM image search

DNI	Speciality	Study date	Description	Modality
79000433	Mastologia	18/09/2015	BreastCE	CT

Consult

Type image	speciality	Date	DNI	Modality	Description	Serie	Manufacturer	Path
Seno	Mastologia	20150918	79000433	CT	BreastCE	twist	GEMEDICALSYSTEMS	/home/Julio/ImagenesDICOM/Mastologia

Time search: 1.1920928955078E-5

Fig. 22 Search using decision trees of the DICOM images for the doctor.php script

It's observed in Figs. 21 and 22 that only the correct result was obtained according to the query made, thus determining the efficiency of implementing decision trees compared to the sequential search (without data mining technique), since the possibility of false positives was minimized as much as possible. Besides this, the search times for this case can observed, which showed a decrease.

IV. Results

A. Results of the Test Scenarios

As mentioned in the development, each of the test scenarios had a duration of 10 minutes for the extraction of time, package size and package number variables. Once these 3 values were obtained, the calculation of the QoS metrics, such as delay, throughput (general and instantaneous), bandwidth (available and variation), jitter and latency were conducted, which were averaged from the 600 samples obtained in each of the tests (24 tests in total, divided between the 5 corresponding scenarios). Hereunder, the collection and calculation of the QoS metrics are shown for each of the conducted tests (Fig. 23).

Packet Size (Bytes)	Protocol	IP source	IP Destination	# of secuency	Time (ms)	Delay E2E (s)	Delay E2E (ms)	Latency (ms)	Jitter (ms)	BW Available (Mbps)	BW Variation (Mbps)	efficiency	Instant Throughput (bps)
54	TCP	192.168.0.17	190.60.233.52	20	7.7128	0.003856	3.8564	3.8564	0.0000	0.1120	0.1120	6%	0.0026
66	TCP	192.168.0.17	190.60.233.52	21	0.3400	0.000170	0.1700	3.8645	3.6945	3.1059	1.4969	153%	0.0618
60	TCP	190.60.233.52	192.168.0.17	22	24.8380	0.012419	12.4190	4.4290	7.9900	0.0387	0.7291	2%	0.0009
66	TCP	190.60.233.52	192.168.0.17	23	14.7980	0.007399	7.3990	4.7507	2.6483	0.0714	0.3289	4%	0.0016
54	TCP	192.168.0.17	190.60.233.52	24	0.0720	0.000036	0.0360	4.7522	4.7162	12.0000	5.8356	593%	0.3333
1381	TCP	192.168.0.17	190.60.233.52	25	0.9000	0.000450	0.4500	4.7702	4.3202	24.5511	9.3578	1213%	0.0278
13194	TCP	192.168.0.17	190.60.233.52	26	0.2100	0.000105	0.1050	4.7742	4.6692	1005.2571	497.9497	49667%	0.1238
60	TCP	190.60.233.52	192.168.0.17	27	35.6560	0.017828	17.8280	5.4345	12.3935	0.0269	248.9614	1%	0.0008
1514	TCP	192.168.0.17	190.60.233.52	28	0.1280	0.000064	0.0640	5.4368	5.3728	189.2500	29.8557	9350%	0.2188
66	TCP	190.60.233.52	192.168.0.17	29	2.0080	0.001004	1.0040	5.4714	4.4674	0.5259	14.6649	26%	0.0144
66	TCP	190.60.233.52	192.168.0.17	30	1.0000	0.000500	0.5000	5.4881	4.9881	1.0560	6.8044	52%	0.0300
66	TCP	190.60.233.52	192.168.0.17	31	3.7980	0.001899	1.8990	5.5494	3.6504	0.2780	3.2632	14%	0.0082
66	TCP	190.60.233.52	192.168.0.17	32	6.0480	0.003024	3.0240	5.6439	2.6199	0.1746	1.5443	9%	0.0053
66	TCP	190.60.233.52	192.168.0.17	33	6.3600	0.003180	3.1800	5.7402	2.5602	0.1660	0.6891	8%	0.0052
60	TCP	190.60.233.52	192.168.0.17	34	0.6840	0.000342	0.3420	5.7503	5.4083	1.4035	0.3572	69%	0.0497
24874	TCP	192.168.0.17	190.60.233.52	35	0.1480	0.000074	0.0740	5.7524	5.6784	2689.0811	1344.3619	132860%	0.2365
60	TCP	190.60.233.52	192.168.0.17	36	4.2340	0.002117	2.1170	5.8112	3.6942	0.2267	672.0676	11%	0.0085
2974	TCP	192.168.0.17	190.60.233.52	37	0.1000	0.000050	0.0500	5.8126	5.7626	475.8400	98.1138	23510%	0.3700
60	TCP	190.60.233.52	192.168.0.17	38	23.9920	0.011996	11.9960	6.1282	5.8678	0.0400	49.0369	2%	0.0016
60	TCP	190.60.233.52	192.168.0.17	39	0.0040	0.000002	0.0020	6.1283	6.1263	240.0000	95.4816	11858%	9.7500
60	TCP	190.60.233.52	192.168.0.17	40	0.0020	0.000001	0.0010	6.1283	6.1273	480.0000	192.2592	23715%	20.0000
60	TCP	190.60.233.52	192.168.0.17	41	0.0000	0.000000	0.0000	6.1283	6.1283	0.0000	96.1296	0%	0.0000
14654	TCP	192.168.0.17	190.60.233.52	42	0.1660	0.000083	0.0830	6.1303	6.0473	1412.4337	658.1521	69784%	0.2530

Fig. 23 Example of calculation and average of the QoS metrics obtained in different tests scenarios

Once the results for the QoS metrics were obtained, these were placed in graphs using the MATLAB tool for the analysis of the behavior during the 10 minutes of the test. Thus, it was possible to determine the variation in time of each of the metrics in the different tests scenarios and verify that each metric remained in the range of optimal and efficient values, which ensure a level of strong and reliable offer (based on the type of information handled and complying with the standards and objectives from the project). This graphic representation process was documented for its subsequent analysis (Figs. 24 and 25).

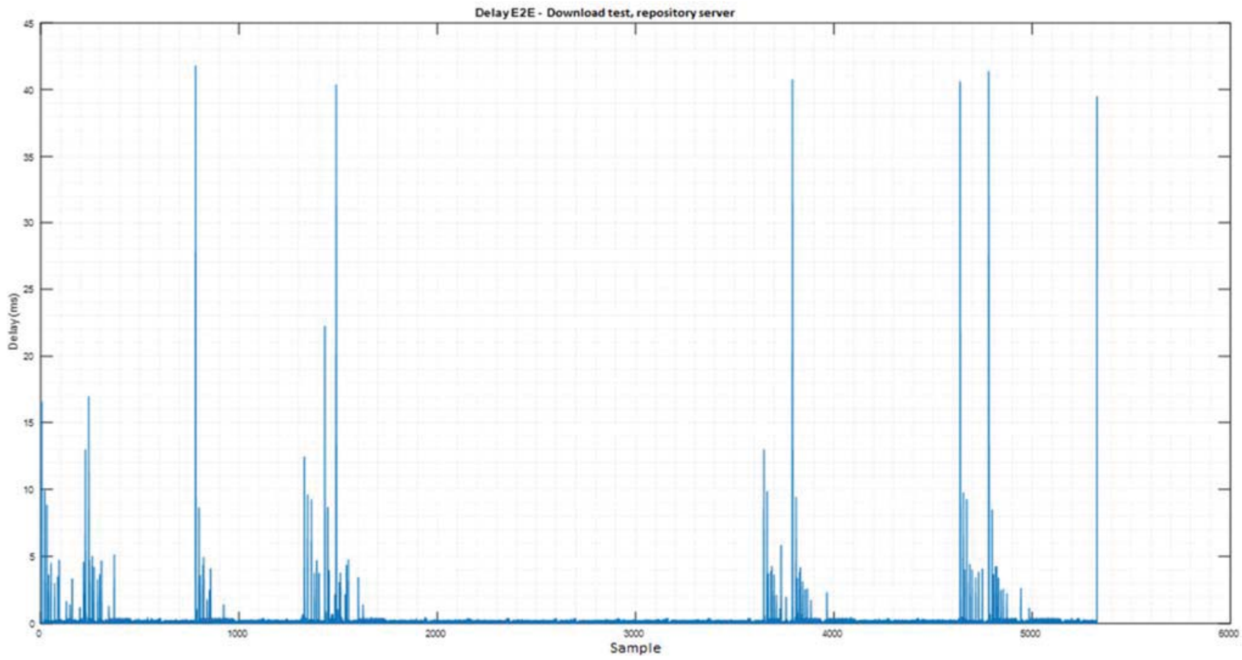


Fig. 24 Example of plot for the QoS metrics for the different test scenarios

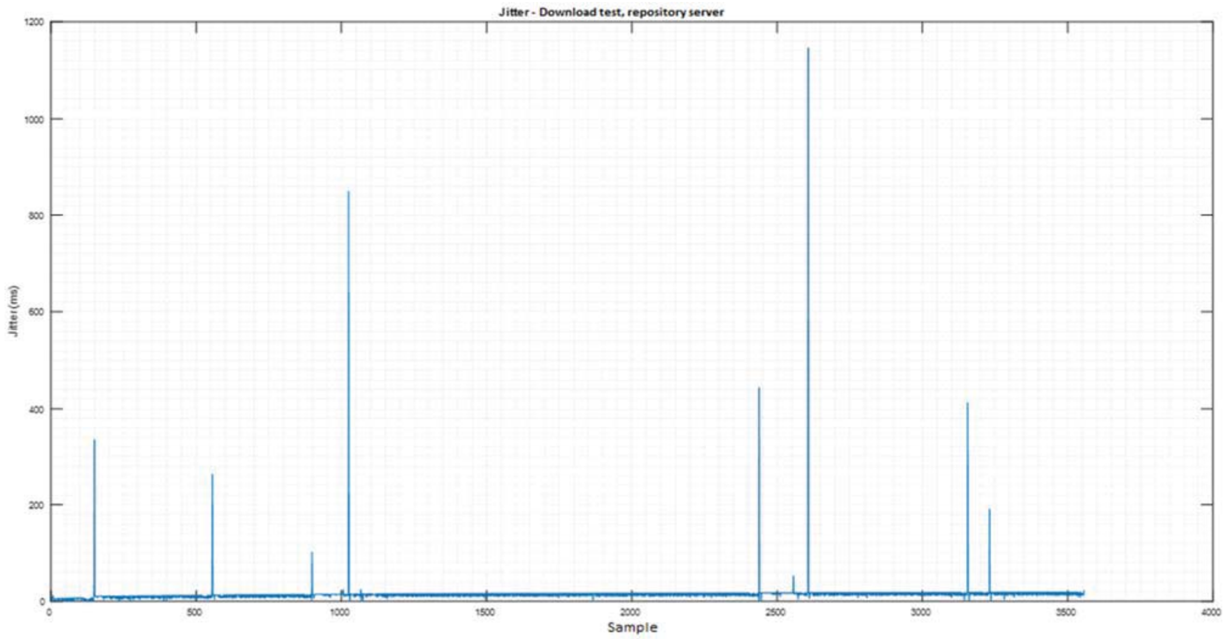


Fig. 25 Example of plot for the QoS metrics for the different test scenarios

In the tests focused on the local measurements (scenarios 1 and 2), it was determined that for the type of information analyzed in the cloud server (DICOM images form a patient), the server exceeded the minimum standards that must be considered regarding the delay, latency, bandwidth, throughput and efficiency, confirming that the TIGUM group has excellent quality parameters that guarantee the efficiency, reliability, quickness, effectiveness and stability in the service. Said results are shown in Tables I and II.

TABLE I
TEST SCENARIO 1 RESULTS

Metric	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6
Delay E2E (ms)	0,284	0,508	0,630	0,478	0,741	0,029
Latency (ms)	1,513	2,317	3,481	4,384	0,291	0,156
Jitter (ms)	1,239	2,171	3,098	3,917	2,600	0,128
BW available (Mbps)	2,198	2,992	2,100	2,166	2,155	19,194
BW variation (Mbps)	0,743	1,052	0,746	0,754	0,781	6,431
Efficiency (%)	108,60	148	104	107	106,5	984,3
Instant TH (bps)	0,629	0,875	1,107	1,136	0,638	5,238
General TH (Kbps)	879,37	492,39	387,25	434,9	337,33	8557,9

BW = Bandwidth, TH = Throughput, E2E = end to end, ms = millisecond, Mbps = Megabits per second, bps = bit per second, Kbps = Kilobits per second.

TABLE II
TEST SCENARIO 2 RESULTS

Metric	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6
Delay E2E (ms)	0,354	0,728	0,775	0,021	0,683	0,497
Latency (ms)	2,225	2,531	3,022	0,116	6,455	2,198
Jitter (ms)	2,054	2,256	2,590	0,095	5,924	1,814
BW available (Mbps)	3,388	2,111	1,996	26,219	2,276	1,848
BW variation (Mbps)	1,178	0,776	0,736	8,816	0,811	0,630
Efficiency (%)	167	104	99	129	112	91
Instant TH (bps)	1,003	1,100	1,041	7,590	0,687	0,969
General TH (Kbps)	706,62	430,49	487,36	11836	366,15	622,0

BW = Bandwidth, TH = Throughput, E2E = end to end, ms = millisecond, Mbps = Megabits per second, bps = bit per second, Kbps = Kilobits per second.

For the tests conducted remotely (scenarios 3, 4 and cloud scenario), the QoS metrics levels and the average of the results for each of them decreased significantly regarding the tests conducted locally. It is worth mentioning that the cloud server guaranteed an efficiency and offering level of over 80% (according to the QoS metrics), thus fulfilling the availability, strength and accessibility parameters that were evaluated throughout the article. These results are shown in Tables III-V.

TABLE III
TEST SCENARIO 3 RESULTS

Metric	Test 1	Test 2
Delay E2E (ms)	19,263	0,412
Latency (ms)	115,241	2,747
Jitter (ms)	96,158	2,491
BW available (Mbps)	0,020	3,028
BW variation (Mbps)	0,007	1,062
Efficiency (%)	97,4	149,61
Instant TH (bps)	0,012	0,893
General TH (Kbps)	6,489	606,83

BW = Bandwidth, TH = Throughput, E2E = end to end, ms = millisecond, Mbps = Megabits per second, bps = bit per second, Kbps = Kilobits per second.

TABLE IV
TEST SCENARIO 4 RESULTS

Metric	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7
Delay E2E (ms)	1,106	1,106	0,347	0,020	1,222	0,933	7,030
Latency (ms)	10,275	7,447	2,208	0,125	7,785	5,543	42,333
Jitter (ms)	9,200	6,459	1,906	0,105	6,849	4,887	35,303
BW available (Mbps)	1,739	1,636	2,023	27,95	1,942	2,720	0,039
BW variation (Mbps)	0,677	0,627	0,692	9,345	0,740	1,029	0,013
Efficiency (%)	86	81	100	138	96	134	191
Instant TH (bps)	0,514	0,496	0,589	8,123	0,568	0,777	0,023
General TH (Kbps)	226,	226,1	720,1	12401	204,5	268,0	17,78

BW = Bandwidth, TH = Throughput, E2E = end to end, ms = millisecond, Mbps = Megabits per second, bps = bit per second, Kbps = Kilobits per second.

TABLE V
TEST SCENARIO CLOUD RESULTS

Metric	Test 1	Test 2	Test 3	Test 4
Packet size (Bytes)	1527	4128	1014	1094
Delay E2E (ms)	3,31	0,61	0,29	380,51
Latency (ms)	16,25	0,05	11,83	5,75
Jitter (ms)	15,09	0,56	11,61	380,35
BW Available (Mbps)	353,98	543,55	111,42	59,14
BW Variation (Mbps)	184,15	358,85	58,57	37,57
Instant TH (bps)	37,17	10,04	49,96	32,44
General TH (Kbps)	3536,46	52737,09	26979,94	22,46

BW = Bandwidth, TH = Throughput, E2E = end to end, ms = millisecond, Mbps = Megabits per second, bps = bit per second, Kbps = Kilobits per second.

B. Results from the Data Mining

Once the consult of 20 images was conducted using two selected search types, the tabulation of the obtained results was conducted regarding the search times obtained in the sequential search and the search with decision trees, obtaining the following results shown in Tables VI and VII for the 20 tests with 'patient.php' script and the 'doctor.php' script.

TABLE VI
TESTS OF SEQUENTIAL SEARCH VS. DECISION TREE SEARCH - PATIENT

Test	DNI	SeqS Time	DTS Time
1	79000207	0,00001692	0,00001406
2	95444191	0,00022697	0,00001597
3	79002145	0,00003409	0,00001811
4	79008980	0,00005006	0,00001311
5	79001261	0,00022506	0,00001502
6	79404495	0,00001406	0,00001406
7	85687381	0,00031208	0,00001406
8	79404918	0,00002408	0,00001502
9	92711292	0,00001502	0,00001287
10	79007138	0,00002098	0,00001192
11	79404813	0,00003195	0,00001406
12	79273261	0,00001287	0,00001382
13	79007771	0,00014059	0,00001375
14	99999063	0,00001716	0,00001287
15	79026311	0,00002098	0,00001215
16	83342000	0,00001597	0,00001406
17	79000433	0,00019097	0,00001408
18	81489583	0,00005316	0,00001602
19	79530632	0,00005483	0,00001311
20	79007371	0,00001315	0,00001502
	Average	***0,000074548***	***0,000014157***

DNI = Document identification number, SeqS = Sequential Search, DTS = Decision tree Search.

We can observe that based on the averages obtained from the 20 conducted tests, the obtained search times for both the 'patient.php' script and the 'doctor.php' script, present a decrease relation when the decision trees are applied. The results show a decrease of approximately 16% when compared with sequential search. This confirms the optimization in the search of information in the database from the TIGUM group's server when using decision trees.

To verify the behavior of the results throughout the tests, we made the corresponding graphs for Tables VI and VII (Figs. 26 and 27).

This result is due to the fact that at the moment of conducting the search of information through different specific criteria, the system does not need to analyze each result in a sequential way based on only one criterion (DNI), but it instead focuses on the search using the 4 and 5 criteria present on the decision trees, something that minimizes the result possibilities and thus, optimizes the search times. In turn, through the amount of obtained results, we were able to determine the decrease regarding the false positives that can be obtained while conducting the sequential search in comparison to the search with decision trees. Said results obtained from the 20 consulted images are shown in Table VIII.

TABLE VII
TESTS OF SEQUENTIAL SEARCH VS. DECISION TREE SEARCH - DOCTOR

Test	DNI	SeqS Time	DTS Time
1	79000207	0,00001292	0,00001006
2	95444191	0,00023869	0,00001397
3	79002145	0,00003309	0,00001411
4	79008980	0,00005606	0,00001111
5	79001261	0,00022506	0,00001202
6	79404495	0,00001406	0,00001406
7	85687381	0,00037208	0,00001206
8	79404918	0,00002452	0,00001302
9	92711292	0,00001522	0,00001287
10	79007138	0,00002098	0,00001092
11	79404813	0,00003195	0,00001306
12	79273261	0,00001384	0,00001382
13	79007771	0,00014059	0,00001375
14	99999063	0,00001711	0,00001287
15	79026311	0,00002098	0,00001215
16	83342000	0,00001597	0,00001706
17	79000433	0,00021091	0,00001208
18	81489583	0,00004578	0,00001202
19	79530632	0,00005846	0,00001311
20	79007371	0,00001333	0,00001402
	Average	***0.000079080***	***0.000012907***

DNI = Document identification number, SeqS = Sequential Search, DTS = Decision tree Search.

Through the obtained averages, we can observe that the quantity of images consulted using the sequential search is substantially greater compared to the search using decision trees. We calculated the sensibility and specificity percentage with (8) and (9), which determined the capacity the system had to detect the amount of correct and incorrect searches regarding the total amount of results obtained in each case. Finally, the percentage of avoidance of false positives was calculated by applying the decision trees, obtaining a result of 58%, thus determining that the efficiency regarding the search of DICOM medical images when applying the data mining technique was successful in each test (since only one result was obtained, which corresponds to the expected result in each test) and thus, guarantees an optimal offer level for the service. These are the results, because at the time of conducting the search with decision trees, the search criteria conducted said process in a detailed way, arriving to the expected result in a consistent and concise manner.

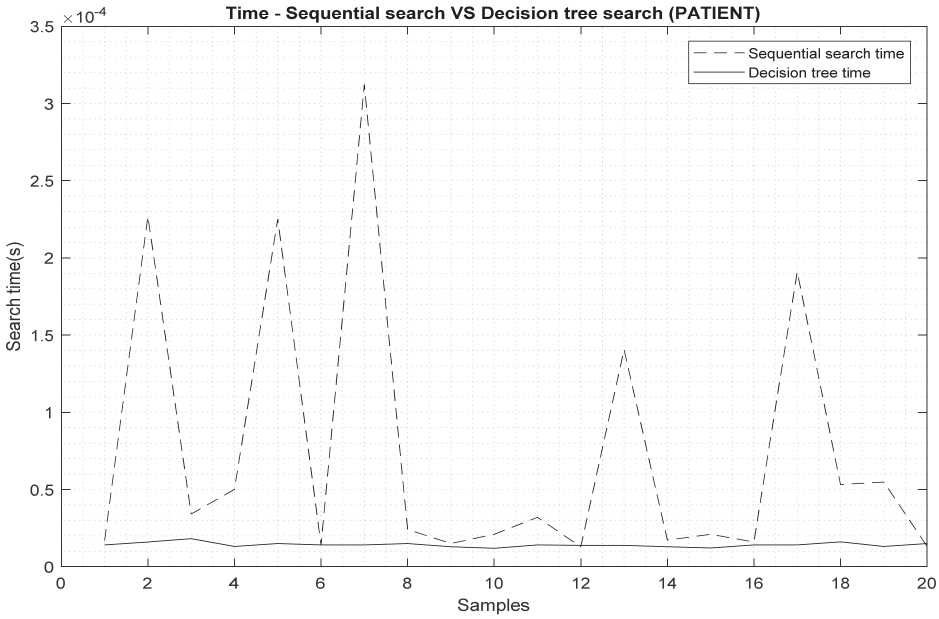


Fig. 26 Plot for time applying the sequential search vs. search with decision trees for the 'patient.php' scrip

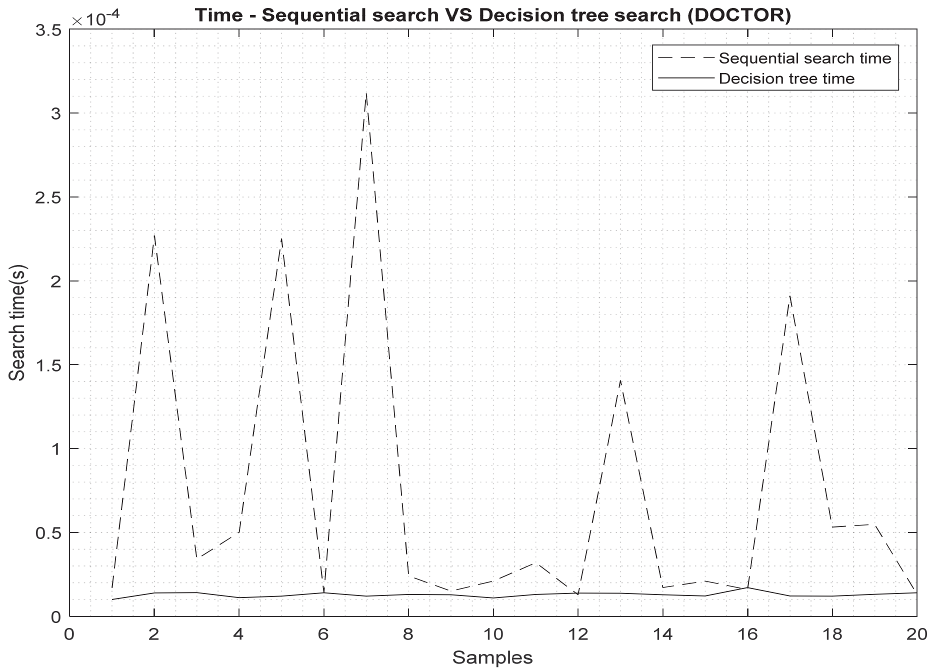


Fig. 27 Plot for time applying the sequential search vs. search with decision trees for the 'doctor.php' script

V. CONCLUSION

Cloud computing has allowed to generate opportunities in different work fields, such as the health sector, the remote monitoring of patients and the management of medical information, which allow the development of a new approach in the IT sector that must be implemented eventually in all health areas. It has generated benefits for different information systems based on the health sector. Cloud has not only focused on the resolution of technical incidents, such as the storage of data, the bandwidth and the reduction of costs through the optimization of resources, it has also allowed to establish a bond between technology and healthcare. Its implementation has directly helped to achieve better results in the treatment of patients and has provided new services focused on the health sector. The fulfillment of the QoS metrics is an important factor for the deployment and operation of the cloud services, since these determine the level of offer that a service must provide, and they must guarantee that the technical aspects, such as system efficiency effectiveness, reliability, speed and stability are fulfilled. It is worth mentioning that the information generated in the medical field, is a type of information completely different to the one that can be generated in other aspects from the IT sector. The medical information is classified as semiprivate and classified, which is why it must be treated with a completely different priority compared to the type of information that exists, for example, in the economic/financial field, since the medical one does not only possess numerical or statistical data. The wrong use, managing, administration and handling of medical information may result in a bad procedure that can subsequently harm the health and life of a patient.

Decision trees as a data mining technique allow to conduct the management and administration of the information in an optimal and precise manner. The implementation of said technique allows us to improve aspects related to the times taken to search for information and minimize the waste of resources on the network and hardware level, while also avoiding the appearance of false positives in consult and download of information processes, thus providing a more efficient and reliable service available for the user.

Cloud computing has allowed to generate benefits in different information systems based on the health sector. It has not only focused on the storage of data, the bandwidth and the reduction of costs through the optimization of resources, but also allowed to establish a bond between technology and healthcare. Its implementation has directly helped to achieve better results in the treatment of patients and has provided new services to the health sector.

TABLE VIII
TEST OF SEQUENTIAL SEARCH VS DECISION TREE SEARCH – FALSE POSITIVES

Test	DNI	SeqS QoR	DTS QoR	% Sensitivity	% Specificity	% AFP
1	*79000207*	*2*	*1*	*33%*	*67%*	*33%*
2	*95444191*	*29*	*1*	*3%*	*97%*	*93%*
3	*79002145*	*6*	*1*	*14%*	*86%*	*71%*
4	*79008980*	*12*	*1*	*8%*	*92%*	*85%*
5	*79001261*	*33*	*1*	*3%*	*97%*	*94%*
6	*79404495*	*1*	*1*	*50%*	*50%*	*0%*
7	*85687381*	*37*	*1*	*3%*	*97%*	*95%*
8	*79404918*	*5*	*1*	*17%*	*83%*	*67%*
9	*92711292*	*1*	*1*	*50%*	*50%*	*0%*
10	*79007138*	*4*	*1*	*20%*	*80%*	*60%*
11	*79404813*	*9*	*1*	*10%*	*90%*	*80%*
12	*79273261*	*1*	*1*	*50%*	*50%*	*0%*
13	*79007771*	*4*	*1*	*20%*	*80%*	*60%*
14	*99999063*	*19*	*1*	*5%*	*95%*	*90%*
15	*79026311*	*4*	*1*	*20%*	*80%*	*60%*
16	*83342000*	*2*	*1*	*33%*	*67%*	*33%*
17	*79000433*	*4*	*1*	*20%*	*80%*	*60%*
18	*81489583*	*24*	*1*	*4%*	*96%*	*92%*
19	*79530632*	*15*	*1*	*6%*	*94%*	*88%*
20	*79007371*	*1*	*1*	*50%*	*50%*	*0%*
Average		***10.65***	***1***	***21%***	***79%***	***58%***

DNI = Document identification number, SeqS = Sequential Search, DTS = Decision tree Search, QoR = Quantity of Results, AFP = Avoid false positives.

ACKNOWLEDGMENT

This project has been funded by the Office of the Vice Rector for Research at Universidad Militar Nueva Granada – Project code: IMP-ING-2660.

REFERENCES

[1] Sultan, N. (2014). Making use of cloud computing for healthcare provision: Opportunities and challenges. International Journal of Information Management, 34(2), 177-184.
[2] Deng, M., Petkovic, M., Nalin, M., & Baroni, I. (2011, July). A Home Healthcare System in the Cloud--Addressing Security and Privacy Challenges. In Cloud Computing (CLOUD), 2011 IEEE International Conference on (pp. 549-556). IEEE.
[3] Devadass, L., Sekaran, S. S., & Thinakaran, R. (2017). Cloud Computing in Healthcare. International Journal of Students' Research in Technology & Management, 5(1), 25-31.
[4] Padhy, R. P., Patra, M. R., Ch, S., & Satapathy, R. Design and Implementation of a Cloud based Rural Healthcare Information System Model.
[5] Doukas, C., & Maglogiannis, I. (2012, July). Bringing IoT and cloud computing towards pervasive healthcare. In Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2012 Sixth International Conference on (pp. 922-926). IEEE.
[6] Pandey, S., Voorsluys, W., Niu, S., Khandoker, A., & Buyya, R. (2012). An autonomic cloud environment for hosting ECG data analysis services. Future Generation Computer Systems, 28(1), 147-154.
[7] Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of "big data" on cloud computing: Review and open research issues. Information Systems, 47, 98-115.
[8] Al-Shehri, S. M., Loskot, P., Numanoglu, T., & Mert, M. (2017). Common Metrics for Analyzing, Developing and Managing Telecommunication Networks. arXiv preprint arXiv:1707.03290.
[9] Amanatullah, Y., Lim, C., Ipung, H. P., & Juliandri, A. (2013, June). Toward cloud computing reference architecture: Cloud service management perspective. In ICT for Smart Society (ICISS), 2013 International Conference on (pp. 1-4). IEEE.
[10] Ambulkar, B., & Borkar, V. (2012, April). Data mining in cloud computing. In MPGI National Multi Conference (Vol. 2012).
[11] Raghupathi, W. (2010). Data Mining in Healthcare. Healthcare Informatics: Improving Efficiency and Productivity, 211.
[12] Kuo, A. M. H. (2011). Opportunities and challenges of cloud computing to improve health care services. Journal of medical Internet research, 13(3).
[13] Lee, I., & Lee, K. (2015). The Internet of Things (IoT): Applications, investments, and challenges for enterprises. Business Horizons, 58(4), 431-440.
[14] Guan, Q., Zhang, Z., & Fu, S. (2012). Ensemble of bayesian predictors and decision trees for proactive failure management in cloud computing systems. Journal of Communications, 7(1), 52-61.