

Selección de un algoritmo para la clasificación de Nódulos Pulmonares Solitarios

Algorithm selection for classifying Solitary Pulmonary Nodules

MSc. Arellys Rivero Castro,^I Ing. Luis Manuel Cruz Correa,^{II} Ing. Jeffrey Artiles Lezcano^{III}

I Universidad de las Ciencias Informáticas. Carretera a San Antonio de los Baños, km 2 ½, Boyeros, Ciudad de La Habana, Cuba. E-mail: arcastro@uci.cu

II Universidad de las Ciencias Informáticas. Cuba. E-mail: lmcorrea@uci.cu

RESUMEN

En los últimos años la comunidad científica internacional ha dedicado considerables recursos a la investigación y desarrollo de sistemas de diagnóstico asistidos por ordenador, utilizados por los médicos en el proceso de diagnóstico. Se ha prestado especial atención en algunas áreas médicas, como las especialidades oncológicas, por los altos índices de mortalidad provocados por algunas enfermedades como el cáncer de pulmón. El diagnóstico temprano de este padecimiento puede reducir en gran medida estos indicadores y mejorar la calidad de vida de los pacientes. El objetivo que se pretende con el desarrollo de esta investigación, es la selección adecuada de un algoritmo de clasificación, para ser utilizado en la fase que lleva el mismo nombre como parte de un sistema de diagnóstico asistido por ordenador para la clasificación de nódulos pulmonares solitarios. Para la selección adecuada del algoritmo de clasificación, se realiza un experimento utilizando las herramientas Weka v3.7.10 y Matlab 2013. Para determinar cuál de las técnicas estudiadas arroja mejores resultados de rendimiento, se utilizó el mismo conjunto de datos para las fases de entrenamiento, prueba y validación del clasificador, disponible en la base de datos internacional The Lung Image Database Consortium Image Collection.

Palabras Clave: algoritmo de clasificación, aprendizaje automatizado, nódulos pulmonares solitarios, precisión.

ABSTRACT

In recent years the international scientific community has devoted considerable resources to research and development of systems for computer-aided diagnosis used by physicians in the diagnostic process. Special attention has been provided in some medical areas, such as oncology specialties, by high mortality rates caused by some diseases like lung cancer. Early diagnosis of this condition can greatly reduce these indicators and improve quality of life of patients. The objective pursued with the development of this research is the proper selection of a classification algorithm, to be used in the phase that has the same name, as part of a system of computer-aided diagnosis for classification of solitary pulmonary nodules. For the selection of the appropriate classification algorithm, an experiment was performed using the tools Weka v3.7.10 and Matlab 2013. To determine which of the techniques studied produces better performance results, the same data set was used for the phases of training, testing and validation of the classifier, available in the international database The Lung Image Database Consortium Image Collection.

Key Words: classification algorithm, machine learning, solitary pulmonary nodules, accuracy.

INTRODUCCIÓN

El diagnóstico médico se ha beneficiado por los adelantos alcanzados en la computación y la electrónica.¹ Estos avances puestos en función de los procedimientos médicos, han revolucionado los procesos de atención al paciente ofreciendo alternativas que permiten alcanzar una mejor calidad de vida.

Investigadores a nivel mundial se han dado a la tarea de estudiar la estructura y el funcionamiento de los sistemas de diagnóstico asistido por ordenador (CAD, por sus siglas en inglés), con el fin de ayudar a los especialistas en la toma de decisiones durante el proceso de diagnóstico médico. Un sistema CAD realiza una caracterización de la información, con el objetivo de detectar patrones en los datos y emitir un diagnóstico secundario al del experto. Tiene como objetivo final ayudar a que el profesional mejore su rendimiento diagnóstico.²

Especial atención se le ha prestado al desarrollo de CAD para el estudio y diagnóstico de aquellas afecciones que provocan mayores índices de prevalencia y mortalidad en la población mundial, entre las que se encuentran las enfermedades oncológicas.^{3,4} Entre los diferentes tipos de cáncer, el de pulmón constituye la primera causa de muerte en el varón y la tercera, después del de colon y mama, en la mujer.³ Por tal motivo en la pasada década, ha sido extensivamente evaluado el potencial de los CAD, para aumentar la habilidad de los radiólogos a la hora de detectar lesiones y específicamente clasificar nódulos pulmonares solitarios (NPS) en TC de tórax.^{5,6}

1.1 Sistemas de Diagnóstico Asistido por Ordenador

Los CAD tienen una estructura dividida en cuatro etapas: preprocesamiento, identificación de formas, reconocimiento de formas y clasificación.² La clasificación es la etapa en la que a partir de las características morfológicas o rasgos obtenidos de las estructuras nodulares se realiza una decisión sobre la clase a la que

pertencen (benigna o maligna).⁷ El resultado del proceso proporciona una segunda opinión a los especialistas, mejorando así el rendimiento y la eficacia del diagnóstico médico.^{2,8}

Los resultados obtenidos en la etapa de clasificación están influenciados por el entrenamiento que haya recibido el clasificador. El tipo de entrenamiento y la estructura del clasificador condicionan las características de la clasificación; por lo que su correcto diseño es relevante para el funcionamiento del CAD. La salida de esta etapa es evaluada en el análisis de rendimiento del clasificador. Es en este momento donde se obtienen los indicadores que permiten evaluar el rendimiento del proceso diagnóstico.⁹

La etapa de clasificación se compone de dos fases implícitas: el entrenamiento del clasificador y la evaluación del rendimiento del clasificador.²

El entrenamiento del clasificador evalúa cuáles son los valores de umbral que separan las clases, para esto se apoya en las reglas de decisión previamente definidas. Por lo general cuando se cuenta con un conjunto de objetos previamente clasificados por algún método preciso, estos se utilizan como conjunto de entrenamiento del clasificador.⁹

La evaluación del rendimiento del clasificador da una medida de cuán bueno es el algoritmo clasificando nuevas instancias de objetos. La precisión del clasificador puede ser estimada tabulando su rendimiento sobre un conjunto de objetos de prueba. Para ello el conjunto debe ser representativo, y estar libre de errores.⁹

Para el desarrollo de la fase de clasificación durante la creación de un sistema CAD, se han empleado con éxito varios algoritmos de inteligencia artificial, obteniéndose diversos resultados de precisión. El diagnóstico final emitido por el especialista debe ser el más acertado posible con el objetivo de establecer los procedimientos médicos y tratamientos adecuados para mejorar la condición del paciente.

Teniendo en cuenta este elemento es necesario seleccionar como algoritmo para realizar el proceso de clasificación, aquel que sea capaz de propiciar los mejores indicadores de precisión. El objetivo del presente trabajo es determinar qué algoritmo puede ser más factible para realizar la clasificación de nódulos pulmonares solitarios luego de su identificación en imágenes médicas de tomografía.

2. Nódulo Pulmonar Solitario

Un nódulo pulmonar solitario (NPS) es aquella lesión única, redondeada, menor o igual de 30 mm de diámetro, que se encuentra rodeada completamente de parénquima pulmonar normal, sin ninguna otra lesión acompañante.¹⁰ La mayoría de los NPS son benignos, sin embargo pueden representar una etapa primaria del cáncer de pulmón. Los índices de supervivencia del cáncer de pulmón son muy bajos, apenas el 14% en un intervalo de 5 años. Si el tumor es detectado y clasificado cuando aún se encuentra en la fase de NPS, los índices de supervivencia se elevan hasta un 70-80%. Por lo que el diagnóstico temprano del NPS puede ser la única oportunidad de cura.⁷

2.1 Características morfológico-radiográficas de los nódulos pulmonares solitarios

En las imágenes de Tomografía Computarizada (TC), los NPS presentan un grupo de características morfológico-radiográficas que son utilizadas para su clasificación. En la tabla 1 se muestran estas características y los valores que pudieran tomar cada una de ellas.

Tabla 1. Valores que pueden tomar los nódulos pulmonares solitarios por cada una de sus características morfológico-radiográficas^{11, 12}

Características morfológico-radiográficas	Valores
sutileza	extremadamente sutil - obvio
estructura interna	tejido blando, fluido, grasa o aire
patrón de calcificación	palomita de maíz, laminada, sólida, no central, central o ausente
esfericidad	lineal, ovoide o redonda
bordes	pobremente definidos- bien definidos
lobulación	marcada o no lobulada
espiculación	marcada o no espiculado
textura	no sólida, vidrio deslustrado, parcialmente sólida o sólida

2.2 Selección del Clasificador

La selección del clasificador depende de las características de los datos a utilizar: puede ser supervisado, parcialmente supervisado o sin supervisar. Cada uno corresponde a un estado de conocimiento acerca de las clases objetivos; conocimiento total acerca de la pertenencia de los objetos, parcial o total desconocimiento respectivamente.^{9,13}

Para el desarrollo de la presente investigación se cuenta con un conjunto de estudios médicos de tomografía diagnosticados y se conocen de antemano las clases establecidas en criterios de malignidad que varían desde 1 hasta 5. En este grupo de datos las estructuras nodulares ya están identificadas y sus rasgos han sido cuantificados siguiendo el estándar XML aprobado por el Lung Image Database Consortium (LIDC)¹⁴ del año 2009, actualizado en 2010. Teniendo esta información los autores deciden emplear el enfoque de clasificación supervisada.

La clasificación supervisada trabaja con dos hipótesis bien definidas:⁹

- a. las clases son de naturaleza determinística pues se cuenta con un vector que representa a todos los objetos de una clase, y se conoce como vector prototipo.
- b. toda la información necesaria y suficiente para su diseño se encuentra disponible de antemano.

2.3 Tendencias en el desarrollo de la fase de clasificación de CAD para nódulos pulmonares solitarios

Varias de las investigaciones llevadas a cabo por la comunidad científica internacional se enmarcan en el enfoque de clasificación supervisada. Entre las técnicas más utilizadas están las Redes Neuronales Artificiales (ANN, por sus siglas en inglés)¹⁵ por su capacidad de hacer generalizaciones para resolver futuros problemas, Máquinas de Soporte Vectorial (SVM, por sus siglas en inglés)¹⁶ por su facilidad de entrenamiento y k Vecinos Más Cercanos (kNN, por sus siglas en inglés) por sus fortalezas en la inducción ante datos ruidosos y funciones objetivo complejas.^{17,18} El estudio de tendencias aparece relacionado en la tabla 2.

Tabla 2. Distribución de datos por conjuntos de entrenamiento, prueba y validación

Artículo	Base de Datos	Datos	Algoritmo	Resultados
Neural network-based computer-aided diagnosis in distinguishing malignant from benign solitary pulmonary nodules by computed tomography ¹⁹	<i>The Lung Image Database Consortium Image Collection (LIDC)</i>	Conjunto de entrenamiento: 140 nódulos (selección aleatoria) de conjunto de prueba: 60 nódulos	Red Neuronal Artificial (retroalimentación con propagación trasera) Entrenamiento: Levenberg-Marquardt ²⁰ Condición de parada: suma cuadrada del error menor que 0.001	Sensibilidad: 97.6% Especificidad: 84.2% Precisión promedio: 93.3%.

Artículo	Base de Datos	Datos	Algoritmo	Resultados
Multiple kernel MTLSSVM and its application in lung nodule recognition ²¹	<i>Japanese Society of Radiological Technology (JSRT)</i>	Conjunto de 20 estudios: 80 nódulos.	Máquina de Soporte Vectorial. Kernels: híbrido y Radial Basis Function	Sensibilidad, especificidad y precisión balanceadas. Az=0.96
A Computer Aided Diagnosis System for Lung Cancer based on Statistical and Machine Learning Techniques ²²	<i>Japanese Society of Radiological Technology (JSRT)</i>	Conjunto de 154 nódulos: 100 malignos y 54 benignos.	Combinación de 2 algoritmos, el <i>k Means clustering</i> y el <i>k Vecinos más cercanos</i>	Precisión: 98.7%
A new fusion model for classification of the lung diseases using genetic algorithm ²³	<i>The Lung Image Database Consortium Image Collection (LIDC)</i>	Muestra compuesta por 400 nódulos	K Vecinos más cercanos	Sensibilidad: 88%. Especificidad: 87%.

Una vez identificado que existe una tendencia en el uso de técnicas como kNN, ANN y SVM en el desarrollo de clasificadores para nódulos pulmonares solitarios, los autores de la presente investigación, analizaron los resultados obtenidos por otros investigadores, en estudios comparativos realizados entre estos tres algoritmos.

2.3.1 Comparación entre kNN y ANN

En¹⁵ los autores realizaron un experimento para poder elaborar una comparación entre los algoritmos kNN y ANN. De dichos algoritmos, ANN es más complejo, porque dispone de varios parámetros que deben ser establecidos antes de diseñar el modelo de la red neuronal. El modelo, tamaño, función de activación, parámetros de aprendizaje y el número de muestras de entrenamientos se encuentran entre estos parámetros. Luego de realizada las pruebas de un diagnóstico de imágenes de espectrograma, se concluyó que kNN arrojaba mejores resultados en términos de especificidad.

En¹⁷ realizaron un estudio cuyo objetivo es analizar las diferencias y similitudes en las dos reglas de aprendizaje supervisado y determinar si un clasificador es más adecuado para ciertos problemas de clasificación. Las pruebas realizadas demostraron que el conjunto de entrenamiento tiene una gran influencia en las capacidades de entrenamiento del clasificador. El parámetro comparado en ambos algoritmos es la optimización y el experimento demostró que kNN presenta mejores resultados en sensibilidad y especificidad que ANN, para el mismo tamaño de muestra.

2.3.2 Comparación entre kNN y SVM

En²⁴ decidieron comparar versiones optimizadas de los algoritmos kNN y SVM. Los resultados mostraron que SVM a pesar de tener un buen rendimiento en general, no sobrepasaba a kNN en especificidad y precisión. Si un procesamiento previo de la información es empleado con kNN, el algoritmo mantiene altos valores de precisión incluso al ser aumentado el tamaño del conjunto de datos. No ocurre de la misma forma para la SVM que al aumentar la cantidad de datos, su tiempo de respuesta aumenta de forma cuadrática.

En¹⁸ se realiza un experimento basándose en el reconocimiento de patrones en un grupo de imágenes dividido en tres categorías: modas, flores y personas africanas. Este estudio compara el rendimiento de los algoritmos basado en el graficado de la curva de aprendizaje. Como resultado se obtiene que kNN presenta valores de precisión por encima del 90%, mientras que SVM presenta un comportamiento inestable, quedando por debajo de kNN en dos de las tres categorías de imágenes probadas.

2.3.3 Resultados del estudio comparativo entre kNN, ANN y SVM

Luego de una revisión bibliográfica de artículos que compararan los distintos algoritmos, los autores de la presente investigación evidencian que no es posible determinar qué algoritmo ofrece mejores resultados de precisión realizando una comparación directa entre los estudios. Esto se debe a que existe diversidad de criterios y a que el origen de los datos utilizados para las pruebas es diferente. Para determinar cuál de las técnicas estudiadas pudiera arrojar mejores resultados de rendimiento al ser utilizada para clasificar nódulos pulmonares solitarios, se decide realizar un experimento, que emplee los mismos datos para entrenar y probar los algoritmos: y así igualar las condiciones. Los datos a utilizar serán obtenidos de una base de datos internacional de estudios médicos.

2.4 Materiales y Métodos empleados

Los métodos científicos que se utilizan durante el desarrollo de la investigación son:²⁵

Análítico-Sintético: al estudiar los diferentes elementos que conforman los algoritmos de clasificación de nódulos pulmonares solitarios. Determinar los elementos importantes a tener en cuenta durante la selección de los algoritmos de clasificación a utilizar.

Inductivo-Deductivo: para analizar la problemática existente mediante la comprensión del funcionamiento del proceso de clasificación de estructuras nodulares en imágenes médicas y definir los elementos característicos, que permitan proponer un algoritmo que alcance altos valores de precisión en la clasificación.

Histórico-Lógico: para investigar los antecedentes y evolución de los algoritmos de clasificación de estructuras nodulares. Permite realizar el estudio del estado del arte de la problemática planteada mediante el análisis de diversas soluciones existentes e identificar las fuentes de información.

Experimento: para determinar qué técnica de inteligencia artificial es la que mejores resultados arroja en la clasificación de nódulos pulmonares solitarios.

Estadística Descriptiva: utilizada para recolectar las características de las estructuras nodulares, necesarias para realizar el proceso de clasificación.

Herramientas y tecnologías informáticas utilizadas:

Weka v3.7.10: es un software de código abierto desarrollado en JAVA y licenciado bajo Licencia Pública General (GPL, por sus siglas en inglés), este contiene una colección de algoritmos de aprendizaje automatizado para realizar tareas de minería de datos. Está dotado de un conjunto de herramientas para el preprocesamiento de los datos, su clasificación, análisis de regresión, clustering, establecer reglas de asociación o visualizar la información en forma de gráficos. Puede ser empleado para desarrollar nuevos esquemas de aprendizaje automatizado. Cuenta con módulos específicos para la realización de experimentos y el graficado del flujo de conocimiento en los algoritmos.²⁶

Matlab 2013: es un poderoso asistente matemático interactivo utilizado para la computación numérica, la visualización y la programación. Cuenta con un lenguaje de programación de alto nivel que permite analizar datos, desarrollar algoritmos y crear modelos de aplicaciones. Las herramientas y las funciones matemáticas que incorpora permiten explorar múltiples enfoques en la resolución de problemas y alcanzar soluciones con mayor rapidez que otros lenguajes. Puede ser utilizado para llevar a cabo experimentos en un ambiente controlado, permitiendo una total manipulación de las variables implicadas.²⁷

RESULTADOS Y DISCUSIÓN

Para realizar el experimento es necesario definir el origen de los datos que serán utilizados en el entrenamiento y prueba de los algoritmos, así como sus características. A nivel internacional han sido desarrolladas bases de datos (BD)

que recopilan información asociadas a procesos médicos con el objetivo de potenciar el desarrollo de sistemas CAD. Estas BD reúnen los resultados de análisis llevados a cabo por especialistas y sirven de guía para el desarrollo de herramientas computacionales que utilicen estos datos para inferir nuevo conocimiento.^{28,29} Para el desarrollo de CAD orientados a la clasificación de nódulos pulmonares solitarios se han destacado las siguientes BD:

- The Lung Image Database Consortium Image Collection - Image Database Resource Initiative (LIDC/IDRI).^{30,31}
- Japanese Society of Radiological Technology (JSRT).^{22,32}
- Early Lung Cancer Action Program Public Lung Image Database (ELCAP).³¹
- CT Image Library (CTIL).³³

Existe un número significativo de investigaciones que hacen uso de la LIDC/IDRI por sus ventajas respecto a las otras BD mencionadas. Entre las ventajas se encuentran el mayor tamaño de la BD y que la misma contiene anexo a las series de imágenes de cada estudio, un fichero XML que describe según la norma del National Institute of Health (NIH) y el The Cancer Imaging Archive (TCIA) aprobada en el año 2006 y actualizada en el año 2010, las características más relevantes de las estructuras nodulares presentes en dichas imágenes.^{31,34} En el XML se especifican los datos asociados al estudio en general y las descripciones de las estructuras nodulares encontradas por los radiólogos. Entre el conjunto de parámetros, el XML almacena la probabilidad de malignidad asociada a cada estructura nodular variando desde 1 hasta 5.

Empleando los datos contenidos en 30 series de imágenes escogidas aleatoriamente, los autores de la presente investigación desarrollaron un experimento, para determinar cuál de los algoritmos SVM, ANN y kNN sería utilizado como modelo para la implementación del algoritmo para la clasificación de nódulos pulmonares solitarios. Para lograrlo, modelaron el funcionamiento de los algoritmos encontrados en ambiente Matlab. La implementación de los algoritmos en este asistente matemático permite modificar sus parámetros para simular la igualdad de condiciones. Los datos empleados en cada iteración de prueba fueron los mismos para cada uno de los algoritmos. La distribución de los datos por conjuntos se muestra en la tabla 3.

Tabla 3. Distribución de datos por conjuntos de entrenamiento, prueba y validación

Entrenamiento	Prueba	Validación	Total
68	14	14	96
70%	15%	15%	100%

De los algoritmos, la SVM fue empleada con el kernel Radial Basis Function. La ANN era un perceptrón multicapa entrenada siguiendo el algoritmo de Levenberg-Marquardt. El algoritmo kNN fue probado para distintos valores de k. Los resultados del experimento se evidencian en la tabla 4. La prueba realizada arrojó como resultado que kNN con k=3 obtuvo mejores resultados en cuanto a precisión que los otros algoritmos, para el juego de datos utilizados en el experimento.

Tabla 4. Resultados del experimento para determinar el modelo a seguir para la implementación del algoritmo de clasificación de NPS

Algoritmo	Precisión promedio
kNN k=3	<u>0.87856</u>
kNN k=5	0.86427
RNA	0.81428
SVM	0.83571

CONCLUSIONES

Una vez realizado el análisis de los diferentes algoritmos empleados en la clasificación de estructuras nodulares, se arribó a las siguientes conclusiones:

- Los pasos básicos seguidos por los desarrolladores de algoritmos para la clasificación de nódulos pulmonares solitarios son: entrenamiento del clasificador, clasificación de estructuras nodulares y evaluación del rendimiento del clasificador.
- Existe una tendencia al empleo de bases de datos internacionales para entrenar y validar los algoritmos de clasificación de nódulos pulmonares solitarios, la más relevante de ellas es The Lung Image Database Consortium Image Collection.
- Se seleccionó como algoritmo más apropiado para la clasificación de nódulos pulmonares solitarios el kNN debido a que en el marco de la investigación arrojó mejores resultados de precisión que otras técnicas de clasificación como SVM y ANN al alcanzar valores cercanos al 88%.

REFERENCIAS BIBLIOGRÁFICAS

1. Aguilar C. "IMPACTO DE LA TECNOLOGÍA EN LA MEDICINA," IMPACTO DE LA TECNOLOGÍA EN LA MEDICINA, 2012. [Online]. Available: <http://impactodelatecnologiaenlamedicina.blogspot.com/>
2. Bogoni L, Ko J. P. "Impact of a Computer-Aided Detection (CAD) System Integrated into a Picture Archiving and Communication System (PACS) on Reader Sensitivity and Efficiency for the Detection of Lung Nodules in Thoracic CT Exams," J Digit Imaging. vol. 25, pp. 771-781, 2012.
3. "Cancer Fact Sheet," GLOBOCAN, 2008.
4. Stewart B. W, Wild C. P. "Cancer facts sheet," World Health Organization, Scientific, 2015.
5. Summers R. M. "Road maps for advancement of radiologic computer-aided detection in the 21st century". Radiology, vol. 229, no. 1, pp. 11-13, 2003.

6. Masoomeh B. "Computer-aided dermoscopy for diagnosis of melanoma". *BMC Dermatology*. vol. 5, 2005.
7. Alavi A, Mosenifar Z. "Solitary Pulmonary Nodule". *Medscape*, vol. 1, pp. 1-4, 2013.
8. Suárez J. J. "Desarrollo de un sistema de diagnóstico asistido por computador para detección de nódulos pulmonares en tomografía computarizada multicorte". Tesis PhD, 2009.
9. Ruiz J, Guzmán A, Martínez J. F. *Enfoque Lógico Combinatorio al Reconocimiento de Patrones. Selección de variables y clasificación supervisada.*, Colección de Ciencias de la Computación. Instituto Politécnico Nacional de México: Departamento de Ingeniería Eléctrica de México, 1999.
10. Conde A, Navascués E. "Estudio del nódulo pulmonar solitario". *Manual de Neumología Neumosur*, vol. 3, pp. 233-242, 2009.
11. SEPÚLVEDA C, SEPÚLVEDA A, FUENTES E. "Nódulo Pulmonar Solitario". *Rev. Chilena de Cirugía*. vol. 60, no. 1, pp. 71-78, 2008.
12. Jeong Y. J, Yi C. A, Lee K. S. "Solitary Pulmonary Nodules: Detection, Characterization, and Guidance for Further Diagnostic Workup and Treatment". *Radiology*. vol. 50, no. 3, pp. 2-8, 2008.
13. Martín B, Sanz A. *Redes Neuronales y Sistemas Borrosos*. 2nd ed. Zaragoza, España: Alfaomega, 2001.
14. "Lung Image Database Consortium (LIDC)," Lung Image Database Consortium (LIDC). 2014. [Online]. Available: <http://imaging.cancer.gov/programsandresources/informationsystems/lidc>
15. Dennis M.C. Ideler, "Comparison of Artificial Neural Network and k-Nearest Neighbor for Classification," Brock University, Computer Science Department, Scientific 1, 2010.
16. BETANCOURT G. A. "LAS MÁQUINAS DE SOPORTE VECTORIAL (SVMs)". *Scientia et Technica*. vol. 11, no. 27, 2005.
17. Mustafa M, Taib M, Murat Z, Sulaiman N. "Comparison between KNN and ANN Classification in Brain Balancing Application via Spectrogram Image". *Journal of Computer Science & Computational Mathematics*. vol. 2, no. 4, 2012.
18. Song Z, Gu Q. "Image Classification Using SVM, KNN and Performance Comparison with Logistic Regression". *Ciencia*, Universidad de Dartmouth, 2011.
19. Chen H, Wang X, Ma D, Ma B. "Neural network-based computer-aided diagnosis in distinguishing malignant from benign solitary pulmonary nodules by computed tomography". *Chin Med J (Engl)*. vol. 120, no. 14, p. 1211:1215, 2007.
20. Moré J. J. "The Levenberg-Marquardt algorithm: Implementation and theory," in *Numerical Analysis*, G. A. Watson, Ed. Springer Berlin Heidelberg, 1978, pp. 105-116.

21. Liu L, Wang K, Wen D-W, Li Y. "Multiple kernel MtLSSVM and its application in lung nodule recognition". Journal of Jilin University (Engineering and Technology Edition). vol. 44, no. 2, pp. 508-515, 2014.
22. Hamada S, Brahim B, Suziah S. "A Computer Aided Diagnosis System for Lung Cancer based on Statistical and Machine Learning Techniques". Journal of Computers. vol. 9, no. 2, pp. 425-431, 2014.
23. Bhuvaneswari C, Aruna P, Loganathan D. "A new fusion model for classification of the lung diseases using genetic algorithm". Egyptian Informatics Journal. vol. 15, no. 2, pp. 69-77, 2014.
24. Colas F, Brazdil P. "Comparison of SVM and Some Older Classification Algorithms in Text Classification Tasks". Artificial Intelligence in Theory and Practice. vol. 217, pp. 172-177, 2006.
25. Hernández R, Fernández C, Baptista P. Metodología de la investigación. 4th ed, vol. 4. México: Mc Graw Hill, 2010.
26. "Weka 3: Data Mining Software in Java". 2013. [Online]. Available: <http://www.cs.waikato.ac.nz/%7Eml/weka/index.html>
27. "Matlab. The Language of technical Computing". 2014. [Online]. Available: <http://www.mathworks.com/products/matlab/>
28. Meyer C, Armato S-I, Fenimore C, McLennan G, Bidaut L, et-al. "Quantitative imaging to assess tumor response to therapy: common themes of measurement, truth data, and error sources". Translational Oncology. vol. 2, no. 4, pp. 198-210, 2009.
29. Armato S-I, Meyer C, McNitt-Gray M, McLennan G, Reeves A. "The Reference Image Database to Evaluate Response to therapy in lung cancer (RIDER) project: A resource for the development of change analysis software". Clinical pharmacology and therapeutics, vol. 84, no. 4, pp. 448-456, 2008.
30. Kuravati R, Sasidhar B, Ramesh D-R. "A Novel Method for Classification of Lung Nodules as Benign and Malignant using Artificial Neural Network". International Journal of Engineering and Computer Science. vol. 3, no. 8, pp. 7641-7645, 2014.
31. Meyer C, Armato S-I, Fenimore C, McLennan G, Bidaut L, et-al. "The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI). A completed reference database of lung nodules on CT scans". Medical Physics. vol. 38, no. 2, pp. 915-931, 2011.
32. Shiraishi J, Katsuragawa S, Ikezoe J, Matsumoto T, Kobayashi T, Komatsu K. "Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules". American Journal of Roentgenology. vol. 174, no. 1, pp. 71-7, 2000.
33. Clark K, Gierada D, Moore S, Maffitt D. "Creation of a CT Image Library for the Lung Screening Study of the National Lung Screening Trial". Journal of Digital Imaging. vol. 20, no. 1, pp. 23-31, 2010.

34. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P. "The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository". Journal of Digital Imaging. vol. 26, no. 6, pp. 1045-1057, 2013.

Recibido: 22 de marzo de 2016.

Aprobado: 12 de mayo de 2016.