

Estado del arte de clasificación de nódulos pulmonares utilizando Aprendizaje Automático

Cecilia Irene Loeza Mejía, Rajesh Roshan Biswal, Eddy Sánchez-DelaCruz

Instituto Tecnológico Superior de Misantla, Departamento de Posgrado,
Misantla, Veracruz, México

cecilialoeza@yahoo.com, rroshanb@itsm.edu.mx

Resumen. Los nódulos pulmonares son manchas o lesiones que se diagnostican en los pulmones y pueden tener tanto causas benignas como malignas, principalmente relacionadas con cáncer de pulmón. En México, de acuerdo con estadísticas, el cáncer de pulmón se ubica en el séptimo lugar y es el más letal. Diversas instituciones médicas internacionales trabajan para mejorar el diagnóstico del cáncer de pulmón, debido a que la principal causa de muerte es el diagnóstico tardío. Por esta razón, el análisis de nódulos pulmonares es un desafío en el procesamiento de imágenes médicas para determinar el tratamiento adecuado, se proponen nuevas metodologías y técnicas no sólo por expertos en medicina sino también por otros científicos. El objetivo de este estudio es mostrar técnicas de aprendizaje automático aplicadas en la clasificación de nódulos pulmonares, específicamente se revisarán redes neuronales artificiales y máquina de vector soporte. Se concluye que el aprendizaje automático es efectivo para clasificar nódulos pulmonares en malignos (cancerosos) y benignos (no cancerosos).

Palabras clave: cáncer de pulmón, clasificación de nódulos pulmonares, aprendizaje automático

A Review on Classification of Pulmonary Nodules Using Machine Learning

Abstract. Pulmonary nodules are stains or lesions that are diagnosed in the lungs and can have both benignant and malignant causes, mainly related to lung cancer. In Mexico, according to statistics, lung cancer is in seventh place and is the most lethal. Several international medical institutions continue to work to improve the diagnosis of lung cancer, as delayed diagnosis leads to the principal cause of the mortality. For this reason, the processing of medical images and the diagnosis of pulmonary nodules can be life saving, to determine the appropriate treatment techniques as proposed by medical experts and scientists in the field. The aim of this study is to review the application of state-of-the-art machine learning techniques for the classification of pulmonary nodules, specifically

artificial neural networks and support vector machine. It is concluded that machine learning is very much effective in classifying pulmonary nodules as malignant (cancerous) and benign (non-cancerous).

Keywords: lung cancer, classification of pulmonary nodules, machine learning.

1. Introducción

El cáncer de pulmón es un problema mundial de salud, en México cada año fallecen aproximadamente ocho mil personas a causa del mismo, siendo el principal problema el diagnóstico tardío [1]. En el 99 % de los casos se detecta el cáncer en fase III o IV, existen pocas iniciativas para detectarlo antes [2]. La detección temprana puede prolongar la supervivencia y facilitar la recuperación de las personas. Además, el diagnóstico exacto es clave para el tratamiento adecuado, por lo que es necesario implementar medidas de diagnóstico oportuno.

Las imágenes médicas son muy importantes, ya que ayudan a emitir diagnósticos, determinar el tratamiento y evaluar los cambios que ocurren con el tiempo [3], ejemplo de éstas son la radiografía, la tomografía computarizada (TC) y el ultrasonido. TC es un estudio radiológico que utiliza rayos X para generar múltiples imágenes transversales del interior del cuerpo. Para la detección del cáncer de pulmón, un radiólogo interpreta una tomografía computarizada del tórax, a través de su experiencia y conocimiento médico, dicha interpretación la recibe el oncólogo para emitir el tratamiento de cáncer de pulmón, sin embargo, la visualización de múltiples tomografías puede provocar cansancio al radiólogo y hacer que cometa errores.

Mediante el aprendizaje automático se pueden detectar y predecir patrones [4], ha mostrado ser efectivo en aplicaciones del ámbito médico, tales como: detección temprana de enfermedades, planificación quirúrgica, análisis de patologías, oncología y radiología. El aprendizaje automático es una de las disciplinas de la inteligencia artificial, que emplea estadística, probabilidad y técnicas de optimización, para comprender reglas complicadas y características de alta dimensión; en el tema de cáncer de pulmón se ha aplicado para detectarlo [5], clasificar la gravedad del cáncer [6], predecir la supervivencia de pacientes [7], clasificar el cáncer de pulmón [8,9] y clasificar nódulos pulmonares [10–19].

A continuación, se muestra la revisión de la literatura en la sección 2, mientras que en la sección 3 se exponen los datasets y métodos que generalmente se utilizan en la clasificación de nódulos, en la sección 4 se resumen los resultados que han obtenido diversos autores utilizando redes neuronales artificiales y máquina de vector soporte; mientras que en la sección 5 se proporciona una visión general del aprendizaje automático en la clasificación de nódulos pulmonares.

2. Revisión de la literatura

Liu y Kang [10] utilizan la arquitectura de redes neuronales multivista, la cual toma múltiples vistas de cada nódulo ingresado, permitiendo realizar la clasificación binaria de nódulos (benigno y maligno) y una clasificación ternaria (benigno, maligno primario y maligno metastásico). Para validar su arquitectura utilizan el dataset LIDC-IDRI.

Zhu *et al.* [11] proponen DeepLung el cual es un sistema automático de diagnóstico de cáncer. DeepLung consta de dos fases: detección de nódulos y clasificación de nódulos (en benignos o malignos), utiliza redes convolucionales basadas en redes de doble vía 3D y potenciación del gradiente. Las redes de doble vía son más compactas y puede tener mejor rendimiento que redes residuales. Para entrenar y evaluar la detección de nódulos utilizaron el dataset LUNA16, en cambio, para validar la clasificación utilizan el dataset LIDC-IDRI.

Abbas [12] presenta el sistema Nodular-Deep que permite clasificar nódulos pulmonares, contempla tres etapas: (i) extracción de características mediante una red neuronal convolucional (ii) optimización de características utilizando red neuronal recurrente, la cual es una técnica de aprendizaje supervisado (iii) clasificación de nódulos (benignos y malignos) optimizadas a través del clasificador lineal softmax. Debido a la arquitectura presentada, Nodular-Deep no requiere preprocesamiento ni postprocesamiento. Para validar sus resultados utiliza el dataset LIDC-IDRI.

Yan *et al.* [13] proponen utilizar una red neuronal convolucional para clasificar nódulos pulmonares (en benignos y malignos), así mismo, implementan tres redes neuronales: (i) 2D a nivel de división (ii) 2D a nivel de nódulos (iii) 3D a nivel de nódulos. Para comparar las redes neuronales utilizan el dataset LIDC-IDRI, obteniendo mejores resultados en la red 3D a nivel de nódulos.

Shukla *et al.* [14] muestran un sistema de diagnóstico automático, el cual consta de cuatro etapas (i) preprocesamiento, en la cual utilizan un filtro difuso (ii) segmentación utilizando el algoritmo Watershed (iii) extracción de características mediante una matriz de co-ocurrencia de nivel de gris (iv) clasificación utilizando máquina de vector soporte.

Parven y Kavitha [15] proponen un sistema que considera cuatro etapas (i) preprocesamiento donde extraen la región del pulmón de la TC (ii) segmentación para identificar regiones del pulmón con nódulos (iii) extracción de características utilizando una matriz de co-ocurrencia de nivel de gris (iv) clasificación de nódulos en benignos y malignos utilizando máquina de vector soporte con distintos kernels (lineal, cuadrático polinomial y función de base radial). Utilizan un dataset con TC de 11 pacientes de un hospital reconocido. Las pruebas las realizan en tiempo real.

Mabrouk *et al.* [16] exponen el desarrollo de un sistema de diagnóstico asistido por computadora para clasificar nódulos grandes. La metodología que utilizan es (i) extracción de regiones pulmonares mediante umbralización (ii) extracción de las características más relevantes (iii) selección de características, en esta etapa eligen las características más importantes para diferenciar entre tejidos normales y tejidos cancerosos (iv) clasificación de tejidos normales y cancerosos utilizando

máquina de vector soporte con diferentes kernels (lineal, cuadrático y función de base radial). Utilizan un dataset de 12 TC recuperadas de la Universidad de Cornell.

Kuruville y Gunavathi [17] clasifican nódulos pulmonares, la metodología que utilizan es (i) segmentación mediante operaciones morfológicas y el método de OTSU (ii) extraen parámetros estadísticos de las regiones segmentadas de los pulmones (iii) utilizan redes neuronales para la clasificación de nódulos cancerosos y nódulos no cancerosos. Las tomografías que utilizan son las del dataset LIDC-IDRI y de un hospital.

Madero *et al.* [18] proponen una metodología para la clasificación de nódulos compuesta por las etapas (i) extracción de la región de interés (ii) transformada Wavelet (iii) extracción de características (iv) selección de atributos y sub-bandas, en esta etapa buscan el mejor subconjunto de atributos en el vector de características para diferenciar entre nódulos cancerosos y no nódulos (v) clasificación utilizando máquina de vector soporte. Los datasets que utilizan son LIDC-IDRI y ELCAP.

Son *et al.* [19] proponen un método de aprendizaje profundo para mejorar la precisión de clasificación de nódulos pulmonares en la TC. Utilizan una red neuronal convolucional profunda 2D de 15 capas para la extracción y clasificación de nódulos y no nódulos. Además, aplican la función de pérdida focal en el proceso de entrenamiento para aumentar la precisión de clasificación del modelo de aprendizaje automático. Para validar sus resultados utilizan el dataset LUNA16.

3. Materiales y métodos de clasificación utilizando Aprendizaje Automático

El aprendizaje automático permite la resolución de tareas complejas, en la literatura existen diversos enfoques de aprendizaje automático para clasificar nódulos pulmonares, a continuación, se hace un compendio de los materiales y métodos utilizados en la literatura, se muestran los datasets que utilizan y el procesamiento de la TC aplicando aprendizaje automático.

3.1. Dataset

Para entrenar y evaluar el rendimiento de algoritmos se debe obtener un dataset de imágenes de TC del tórax. Dichas imágenes son en formato DICOM, el cual es un estándar que se creó para promover la comunicación de información asociada a imágenes médicas entre dispositivos fabricados por diferentes proveedores [20]. Una imagen DICOM contiene un encabezado con información para identificarla en el dataset, datos personales del paciente, parámetros de adquisición y datos de la imagen. Cada TC de un paciente se agrupa en una carpeta y contiene múltiples imágenes transversales de diferentes planos o cortes del tórax: (i) sagital, es paralelo al plano lateral del paciente (ii) coronal, es paralelo al plano frontal del paciente, (iii) axial, es perpendicular al eje longitudinal del paciente.

En la literatura se han utilizado datasets privados los cuales se generan en hospitales y en universidades, tales como la de Michigan, de Harvard y de Toronto. También se han utilizado datasets públicos como LIDC-IDRI, LUNA16 y ELCAP. Cada uno de los datasets contienen distinto número de carpetas que corresponde

Datasets privados. En [8] se mencionan 3 datasets privados: (i) Universidad de Michigan contiene 96 muestras, de las cuales 86 son adenocarcinoma primario (donde 67 están en etapa I y 19 en etapa III) y 10 de tejidos no neoplásicos. (ii) Brigham and Women's Hospital, Harvard contiene 181 muestras de tejidos malignos, clasificados en mesotelioma pleural y adenocarcinoma. (iii) Universidad de Toronto contiene 39 casos de cáncer de células no pequeñas.

Datasets públicos. LIDC-IDRI (del inglés *Lung Image Database Consortium and Image Database Resource Initiative*) es un repositorio público con tomografías de 1018 casos recuperadas de diversos centros médicos e instituciones de imágenes médicas. Dentro de cada caso se incluyen imágenes de una TC y un archivo XML asociado que registra los resultados de un proceso de anotación de imagen de dos fases realizado por cuatro radiólogos torácicos experimentados [21]. (ii) LUNA16 contiene 888 TC, es un subconjunto de tomografías del dataset LIDC-IDRI que excluye los escaneos con un grosor de corte superior a 2.5 mm [22]. (iii) ELCAP consta de un conjunto imágenes de 50 TC de pulmón completo documentadas en dosis bajas [23].

3.2. Preprocesamiento

El preprocesamiento permite mejorar la calidad de las imágenes y eliminar la redundancia de los datos. Además, puede realzar características para su posterior procesamiento, permitiendo visualizar detalles que no se veían en la imagen original [24]. El preprocesamiento generalmente es el paso que ocupa más tiempo. En esta etapa se convierten las tomografías a escala de grises y se elimina el ruido, el cual se puede observar en las imágenes de TC como un granulado de la misma, es decir, aparecen como si tuvieran manchas [25]. Para realizar el preprocesamiento [14] aplica un filtro difuso, el cual comúnmente se utiliza para disminuir el ruido de imágenes. Por otra parte [15] emplea técnicas de binarización, etiquetado, encogimiento y expansión.

3.3. Segmentación

La segmentación permite obtener objetos o regiones de interés significativas de la imagen, para ello, se utilizan distintos métodos, los cuales trabajan sobre las variaciones de intensidad o textura [24]. La segmentación en tomografías identifica las regiones pulmonares. El nódulo pulmonar es un tejido anormal que puede ser benigno o maligno, máximo mide 3 centímetros de diámetro y su forma es redonda u ovalada [26]. La segmentación de la TC de tórax es un

desafío ya que hay falta de homogeneidad en las estructuras pulmonares con densidades similares, tales como venas, arterias, bronquiolos, además, tienen diferentes protocolos de escaneo y escáneres [14].

Para segmentar la TC, [9] utiliza la umbralización y el algoritmo Watershed, en cambio [14] aplica el algoritmo Watershed, por otra parte [17] utiliza operaciones morfológicas y el método de OTSU, mientras que [16] emplea la transformada de Radón.

3.4. Extracción de características

Es una etapa crucial e importante, obtiene las características más relevantes en las cuales se tienen sospechas de la presencia de nódulos, para que posteriormente se puedan diferenciar entre benignos y malignos. La calidad de la extracción de características es de suma importancia para clasificar correctamente los nódulos. Para extraer características [14–16, 18] emplean una matriz de co-ocurrencia de nivel de gris (GLCM) la cual contiene información sobre los píxeles con valores de nivel de gris similares. GLCM es de los métodos más utilizados en el análisis de textura de imágenes biológicas. Por otra [11] parte utiliza red profunda 3D, mientras que [12] utiliza una red neuronal convolucional.

3.5. Clasificación de nódulos

La clasificación automática es fundamental en el diagnóstico asistido por computadora cuando se emplean imágenes médicas. En esta etapa se clasifican los nódulos en benignos (no cancerosos) o malignos (cancerosos), con base en su tamaño, densidad, contorno y forma. Para clasificar los nódulos se utilizan dos conjuntos de datos uno para entrenamiento y el otro para validación. Las técnicas de aprendizaje automático que se han aplicado para la clasificación de nódulos son redes neuronales artificiales (ANN) y máquina de vector soporte (SVM), las cuales se utilizan ampliamente en el campo médico para la detección de enfermedades.

Redes neuronales artificiales. Las ANN son neuronas entrelazadas que se inspiran en las redes biológicas y se procesan de forma paralela para obtener un resultado. Las ANN tienen la propiedad de aprender a partir de unas pocas representaciones de los datos, su estructura es jerárquica, así mismo, tienen flexibilidad lo que permite reconocer patrones con ruido o incompletos. Además, tienen redundancia intrínseca, lo cual permite un funcionamiento perfecto del sistema aun estando dañado alguno de sus componentes [27]. Para clasificar nódulos pulmonares se han utilizado redes multivista, backpropagation, 2D, 3D y 3D con potenciación del gradiente.

Máquina de vector soporte. Es una técnica de aprendizaje supervisado que permite analizar e identificar categorías o clases utilizando los datos entrenados y son robustas al ruido en los datos, así mismo, se pueden aplicar en problemas

de alta dimensionalidad. Para separar las clases analizadas se utilizan kernels, los cuales permiten transformar un problema no linealmente separable a uno linealmente separable. Además, la SVM tiene una gran capacidad de generalización incluso cuando el conjunto de entrenamiento es pequeño [28]. Los kernels que utilizan en la literatura son: lineal, cuadrático, polinomial y función de base radial.

En la sección de resultados se muestran los algoritmos que se han utilizado en diversos proyectos de clasificación de nódulos pulmonares. Los algoritmos se han programado en diversas herramientas como Python, R y Matlab; empleando distribuciones del sistema operativo Windows y Linux. Algunos autores también han empleado librerías como TensorFlow en Python.

3.6. Validación de resultados

Se verifica que el algoritmo clasifica correctamente los nódulos. En la literatura se han utilizado: matriz de confusión y validación cruzada para evaluar el modelo. En la matriz de confusión en cada columna se representa el número de predicciones de cada clase y cada fila representa los valores reales. Los parámetros que se utilizan son TP (verdadero positivo), TN (verdadero negativo), FP (falso positivo) y FN (falso negativo). La matriz de confusión permite calcular la precisión (*accuracy*), como la proporción entre las predicciones correctas que ha hecho el modelo y el total de predicciones, la ecuación 1 permite obtener la precisión. La ecuación 2 obtiene la sensibilidad, la cual es la fracción de nódulos clasificados correctamente. En la ecuación 3 se muestra la especificidad la cual es la fracción de no nódulos clasificados correctamente. También se puede obtener la curva ROC, la cual que describe la fracción de identificación correcta (sensibilidad) contra la tasa de falsas positivos (1 - especificidad):

$$ACC = \frac{TP + TN}{TP + FP + FN + TN}, \quad (1)$$

$$SE = \frac{TP}{TP + FN}, \quad (2)$$

$$SP = \frac{TN}{TN + FP}. \quad (3)$$

Para validar sus resultados [18] utiliza sensibilidad, especificidad, precisión y curva ROC. En cambio [16, 17, 19] calculan precisión, especificidad y sensibilidad. Por otra parte [10] calcula sensibilidad, especificidad y curva ROC. En contraste [15] utiliza sensibilidad y especificidad. Mientras que [14] calcula la precisión.

Otros autores utilizan la validación cruzada para evaluar la clasificación, en la cual, el dataset original se divide en k subconjuntos o pliegues, para cada pliegue, el algoritmo se entrena en la unión de los otros pliegues y luego el error de su salida se estima utilizando el pliegue. Finalmente, el promedio de todos estos

errores es la estimación del error verdadero [29]. En la ecuación 4 se muestra la fórmula para el cálculo del error de validación cruzada en k iteraciones:

$$E = \frac{1}{K} \sum_{i=1}^K E_i. \quad (4)$$

El número de iteraciones k depende del tamaño del dataset, [11] utiliza validación cruzada de 10 iteraciones en el dataset LUNA16, así mismo [12] aplica validación de 10 iteraciones en el dataset LIDC-IDRI, por otra parte [13] emplea validación de 5 iteraciones en LIDC-IDRI.

3.7. Resumen de metodología de clasificación de nódulos

En la figura 1 se expone de manera general la metodología que utilizan diversos autores para clasificar los nódulos: (i) se preprocesa el dataset de TC de tórax (ii) se identifican regiones con nódulos (iii) se obtienen las características más relevantes (iv) clasifican los nódulos pulmonares en benignos o malignos y (v) validan que los resultados hayan sido los correctos.

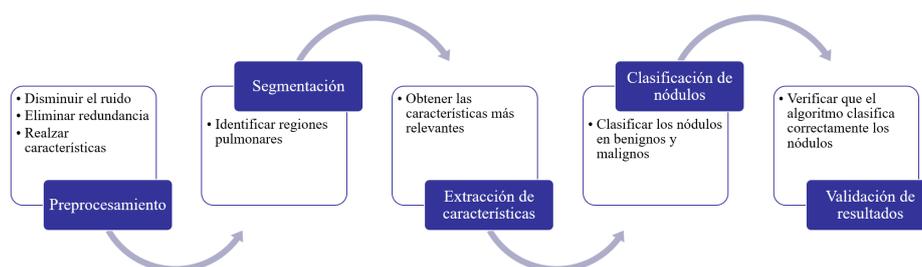


Fig. 1. Metodología para Clasificación de Nódulos Pulmonares utilizando Aprendizaje Automático

4. Resultados y discusión

En el estado del arte generalmente se clasifican los nódulos pulmonares en benignos y malignos. Para entrenar y validar la clasificación mediante redes neuronales los autores han trabajado con los datasets LUNA16 y LIDC-IDRI, los cuales contienen 888 y 1018 TC respectivamente. Mientras que para el entrenamiento y validación utilizando máquina de vector soporte han utilizado datasets privados que contienen desde 11 TC hasta los datasets públicos como ELCAP con 50 TC y LIDC-IDRI con 1018 TC.

Los resultados que obtuvieron distintos autores se muestran en la tabla 1 se mencionan los métodos de aprendizaje automático utilizados (específicamente

redes neuronales artificiales y máquina de vector soporte) aplicados en proyectos de clasificación de nódulos pulmonares, la precisión que tienen y los tipos de nódulos que clasifican.

El porcentaje de acierto varía desde 81.41 % utilizando redes neuronales convolucionales 3D hasta el 97.2 % utilizando red neuronal convolucional 2D y función de pérdida focal. Como se puede observar el porcentaje de precisión puede variar, aunque se utilice la misma técnica de clasificación, esto se debe a que utilizan diversas técnicas para el preprocesamiento, segmentación y extracción de características.

Tabla 1. Métodos de aprendizaje automático para clasificar nódulos pulmonares.

Referencia	Método utilizado	Acierto	Clasificación de nódulos
[10]	Redes neuronales convolucionales multivista	94.59 % 86.09 %	Benignos y malignos Benigno, maligno primario y maligno metastásico
[11]	Red neuronal convolucional 3D y potenciación del gradiente	81.41 %	Benignos y malignos
[12]	Redes neuronales y softmax	95 %	Benignos y malignos
[13]	Red neuronal convolucional: 2D a nivel de división 2D a nivel de nódulos 3D a nivel de nódulos	86.7 % 87.3 % 87.4 %	Malignos
[14]	SVM	92.5 %	Benignos y malignos
[15]	SVM: Kernel Lineal Kernel Polinomial Kernel función de base radial	83.45 % 85.79 % 91.38 %	Normales, benignos y malignos
[16]	SVM: Kernel Lineal Kernel Cuadrático Kernel función de base radial	91.9 % 94.9 % 91.9 %	Normales, benignos y malignos
[17]	Red neuronal backpropagation utilizando pendiente de gradiente con velocidad de aprendizaje variable	91 %	Benignos y malignos
[18]	SVM	82 %	De 2mm a 30mm de diámetro
[19]	Red neuronal convolucional 2D y función de pérdida focal	97.2 %	Nódulos y no nódulos

5. Conclusión

El aprendizaje automático ha mostrado diversas aplicaciones en el tema de cáncer de pulmón. En el estudio realizado se expone la metodología común que siguen diversos autores para clasificar nódulos pulmonares, lo cual da una pauta para la detección y clasificación del cáncer de pulmón. Para la clasificación de nódulos generalmente se realiza lo siguiente: (i) preprocesamiento de las tomografías para disminuir el ruido y la redundancia, donde emplean diversas técnicas como binarizar y aplicar filtro difuso (ii) segmentación para identificar regiones pulmonares, las técnicas que utilizan son: umbralización, algoritmo Watershed, operaciones morfológicas y transformada de Radón (iii) extracción de las características más relevantes, generalmente los autores emplean la matriz de co-ocurrencia de nivel de gris (iv) clasificación de nódulos en benignos y malignos, aplican métodos de aprendizaje automático como redes neuronales artificiales y máquina de vector soporte (v) validación de los resultados mediante matriz de confusión o validación cruzada, haciendo uso de dataset de TC de tórax, el cual proviene de universidades y hospitales. En el estudio se muestra una visión general de los datasets de tomografías del tórax que utilizan, los métodos que aplican y los resultados que obtienen. Se concluye que el aprendizaje automático se puede aplicar en la detección y clasificación de nódulos pulmonares en tomografías computarizadas para ayudar en la detección y determinación del tratamiento.

Referencias

1. Secretaría de salud: Cada año mueren cerca de ocho mil mexicanos por cáncer de pulmón (2018) <https://www.gob.mx/salud/prensa/145-cada-ano-mueren-cerca-de-ocho-mil-mexicanos-por-cancer-de-pulmon?idiom=es>. Último acceso: 2019/03/09
2. The Economist Intelligence Unit Limited: Lung cancer in Latin America: Time to stop looking away (2018) <https://perspectives.eiu.com>. Último acceso: 2019/03/27
3. Restrepo, A.: Procesamiento de imágenes médicas. *Iatreia* 12(1) (1999)
4. Russell, S., Norvig, P.: *Artificial Intelligence: A Modern Approach*. 3rd edn. Prentice Hall, New Jersey (2010)
5. Zhuo, L., Chenhui, Y., Hang Y., Taihua W.: Deep Reinforcement Learning with Its Application for Lung Cancer Detection in Medical. *Internet of Things*. vol. 97, pp. 1–9. <https://doi.org/10.1016/j.future.2019.02.068>
6. Bergquist, S., Brooks, G., Keating, N., Landrum, M., Rose, S.: Classifying Lung Cancer Severity with Ensemble Machine Learning in Health Care Claims Data. In: *2nd Machine Learning for Healthcare Conference*, pp. 25–38 (2017)
7. Pradeep, K., Naveen, N.: Lung Cancer Survivability Prediction based on Performance Using Classification Techniques of Support Vector Machines, C4.5 and Naive Bayes Algorithms for Healthcare Analytics. *Procedia Computer Science*, vol. 132, pp. 412–420 <https://doi.org/10.1016/j.procs.2018.05.162> (2018)
8. Barchuck, A. (et al.): Evaluation of Machine Learning Algorithm Utilization for Lung Cancer Classification Based on Gene Expression Levels. *Asian Pac J Cancer Prev*. 17(2), 835–838. <https://doi.org/10.7314/APJCP.2016.17.2.835> (2016)

9. Alakwaa, W., Nassef, M., Badr, A.: Lung Cancer Detection and Classification with 3D Convolutional Neural Network (3D-CNN) vol. 11 <https://doi.org/10.14569/IJACSA.2017.080853> (2017)
10. Liu, K., Kang, G.: Multiview Convolutional Neural Networks for Lung Nodule Classification. *International Journal of Imaging Systems and Technology* 27(11), 12–22 (2017)
11. Zhuo, W., Liu, C., Fan, W., Xie, X.: DeepLung: Deep 3D Dual Path Nets for Automated Pulmonary Nodule Detection and Classification. <https://doi.org/10.1109/WACV.2018.00079>
12. Abbas, Q.: Nodular-Deep: Classification of Pulmonary Nodules using Deep Neural Network. *International Journal of Medical Research & Health Sciences* 6(8), pp. 111–118 (2017)
13. Yan, X. (et al.): Classification of Lung Nodule Malignancy Risk on Computed Tomography Images Using (2017)Convolutional Neural Network: A Comparison Between 2D and 3D Strategies. https://doi.org/10.1007/978-3-319-54526-4_7
14. Shukla, N., Narayane, A., Nigade, A., Yadav, K., Mhaske, H.: Lung cancer detection and classification using Support Vector Machine. *International Journal of Engineering And Computer Science* 4(11), 14983–14986 (2015)
15. Parveen, S., Kavitha, C.: Classification of Lung Cancer Nodules using SVM Kernels. *International Journal of Computer Applications* 95(25) (2014)
16. Mabrouk, M., Karrar, A., Sharawy, A.: Support Vector Machine Based Computer Aided Diagnosis System for Large Lung Nodules Classification. *Journal of Medical Imaging and Health Informatics*, vol. 3, pp. 214–220, <https://doi.org/10.1166/jmihi.2013.1154> (2013)
17. Kuruvilla, J., Gunavathi, K.: Lung cancer classification using neural networks for CT images. *Computer Methods and Programs in Biomedicine* 113(1), 202–209. <https://doi.org/10.1016/j.cmpb.2013.10.011> (2014)
18. Madero, H., Vergara, O., Cruz, V., Ochoa, H., Nandayapa, M.: Automated system for lung nodules classification based on wavelet feature descriptor and support vector machine. *BioMedical Engineering OnLine* 14. <https://doi.org/10.1186/s12938-015-0003-y> (2015)
19. Son, G., Phuong, T., Thi, V., Mai, C., Burie, J: Improving Accuracy of Lung Nodule Classification Using Deep Learning with Focal Loss. *Journal of Healthcare Engineering*, vol. 2019, pp. 1–9. <https://doi.org/10.1155/2019/5156416> (2019)
20. Pianykh, O.: *Digital Imaging and Communications in Medicine (DICOM): A Practical Introduction and Survival Guide*. Springer, Berlin (2012)
21. Armato, S. et al.: The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans. *Med Phys* 38(2), 915–931 <https://doi.org/10.1118/1.3528204> (2011)
22. Grand Challenges (s.f.): Lung nodule analysis 2016 <https://luna16.grand-challenge.org/Data/>. Último acceso: 2019/04/09
23. Vision and Image Analysis Group (s.f.): ELCAP Public Lung Image Database <http://www.via.cornell.edu/databases/lungdb.html>. Último acceso: 2019/04/09
24. Bronzino, J.: *Handbook of Medical Imaging*. Academic Press, USA (2000)
25. Gil, J. et al.: *Técnico especialista en laboratorio*. MAD, España (2006)
26. Clavero, J.: Nódulos pulmonares. *Revista Médica Clínica Las Condes* 26(3), pp. 302–312 (2015)
27. Viñuela, I., Galván, I.: *Redes de Neuronas Artificiales un enfoque práctico*. Pearson educación, Madrid (2004)

Cecilia Irene Loeza Mejía, Rajesh Roshan Biswal, Eddy Sánchez-DelaCruz

28. Maldonado, S., Weber, R.: Modelos de selección de atributos para Support Vector. *Revista Ingeniería de Sistemas*, vol. 26 (2012)
29. Shalev, S., Ben, S.: *Understanding Machine Learning: From Theory to Algorithms*. 3rd edn. Cambridge University Press, New York (2014)