

MODELO HÍBRIDO PARA DIAGNÓSTICO DE CÁNCER PULMONAR

Josimar Chire Saire

Tesis profesional, UNSA

Abril 2016

Contents

Introducción

Cáncer Pulmonar

Trabajos relacionados

Marco teórico

Propuesta

Experimentación

Conclusión

Introducción

- 59 millones de muertes en el mundo(2004)
- 35 millones enfermedades no transmisibles(2004)
- El cáncer fue declarado como la primera causa de mortalidad(2008) a nivel mundial con 7.6 millones de muertes un 13 % del total.
- El cáncer de pulmón causó 1.5 millones de muertes(2011).
- Las proyecciones indican que el número de muertes sobrepasara los 11 millones para el 2030 [WHO, 2008].
- El cáncer está relacionado a las células anormales que se dividen sin control en un tejido [INC, 2015].
- Las tomografías computarizadas axiales son utilizadas para realizar el diagnóstico. El área de procesamiento de imágenes e inteligencia artificial son una herramienta para tareas de detección de cáncer pulmonar.

Motivación

- Para diagnosticar cáncer pulmonar, el radiólogo necesita analizar un conjunto de imágenes basado en su conocimiento y experiencia.
- Existen algoritmos como Patrón binario local(PBL) para extracción de características, clasificadores como redes neuronales Perceptrón Multicapa(PMC) y algoritmos de optimización como Optimización por enjambre de Partículas(OEP). Utilizados en diversos trabajos [Terdale and Kulhalli, 2012, Ada, 2013, Zamani M., 2009, Ayman El-Baz, 2013, Elshinawy et al., 2010, Mughal and Ikram, 2004] con buenos resultados.

Definición de Problema

Existen diferentes trabajos relacionados y usan diferentes algoritmos para extraer características y técnicas de inteligencia artificial para clasificar las imágenes. El problema radica en realizar una extracción de características apropiada de las imágenes y realizar una clasificación de las mismas.

Objetivos

Objetivo General

Este trabajo modela, implementa y evalúa un clasificador de tomografías axiales computarizadas utilizando umbralización y criterios geométricos para la segmentación, partición binaria local y medidas estadísticas para la extracción de características, optimización de enjambre de partículas para optimizar el entrenamiento de la red Perceptrón Multicapa.

Objetivo Específicos

- Analizar segmentación utilizando umbralización y criterios geométricos.
- Utilizar PBL y medidas estadísticas para la extracción de características de las imágenes TAC.
- Utilizar optimización de enjambre de partículas para optimizar el entrenamiento de la red.
- Utilizar la red neuronal PMC para la clasificación de las imágenes.

Justificación

- Cáncer declarado como primera causa mundial de mortalidad por la OMS(2008).
- Cáncer de pulmón causó 1.5 millones de muertes(2011).
- Un gran volumen de información es generado y procesado día a día.
- Esta área es soportada por el área de computación(procesamiento de imágenes, inteligencia artificial).
- Las dos etapas más relevantes son: pre-procesamiento y extracción de características.

Delimitación del trabajo

- Se analizan imágenes(TAC) en formato .jpeg en escala de grises.
- Se utiliza la red perceptrón multicapa en la clasificación.
- La entrada es real y la salida binaria.

Modelo de investigación

Al revisar diversos trabajos de tesis de diversas universidades internacionales, nacionales y ver diversos formatos para redactar la tesis, se optó por seguir el modelo propuesto en [Mikael Berndtsson, 2008].

Title page

Abstract

Acknowledgements

Table of contents

List of figures and tables (optional)

Part 1: Introduction and Background

Chapter 1: Introduction

Chapter 2: Background

Chapter 3: Problem description and problem statement

Part 2: Main body – Describing your work

Chapter 4: Theory (advanced preliminaries and/or theoretical part of solution).

Chapter 5: Description of approach and method(s) to solve the problem

Chapter 6: Result analysis

Part 3: Wrap-up

Chapter 7: Related work

Chapter 8: Conclusions

References (or Bibliography)

Appendices (optional)

Contents

Introducción

Cáncer Pulmonar

Trabajos relacionados

Marco teórico

Propuesta

Experimentación

Conclusión

Definición

- Enfermedades en las que células anormales se dividen sin control y pueden invadir otros tejidos[INC, 2015].
- Hay más de 100 diferentes tipos de cáncer. La mayoría de los cánceres toman el nombre del órgano o de las células en donde empiezan

Cáncer Pulmonar

- Cáncer que se forma en los tejidos del pulmón, por lo general, en las células que recubren las vías respiratorias.
- Los dos tipos más importantes de cáncer de pulmón son el cáncer de pulmón de células pequeñas y el cáncer de pulmón de células no pequeñas.
- El cáncer de pulmón de células no pequeñas, es el tipo más común de cáncer de pulmón y conforma el 87% de todos los tumores de pulmón.

Tomografía Axial Computarizada

- Tecnología de exploración de rayos X que produce imágenes detalladas de cortes axiales del cuerpo, obtiene múltiples imágenes al rotar alrededor del cuerpo[Herman, 2009].



Figure 2: TAC muestra un tumor canceroso indicado por la flecha [Lange, 2004].

Importancia TAC

- Determinación del grado de extensión de la neoplasia.
- Diferenciación entre nódulos benignos y malignos.
- Proporcionan valiosa información sobre características morfológicas del tumor.

Contents

Introducción

Cáncer Pulmonar

Trabajos relacionados

Marco teórico

Propuesta

Experimentación

Conclusión

Estado del arte

Diferentes etapas son propuestas

[Mughal and Ikram, 2004, Terdale and Kulhalli, 2012, Ada, 2013, Ayman El-Baz, 2013, Mukherjee et al., 2014, Ignatious and Joseph, 2015] para el diagnóstico de cáncer pulmonar y tienen en común:

- Segmentación.
- Extracción de características.
- Clasificación

Estado del arte

- Se utiliza umbralización en la segmentación [Terdale and Kulhalli, 2012, Ashwin et al., 2012, Ada, 2013, Mukherjee et al., 2014].
- Se utiliza PBL en el análisis de las imágenes [Kachouie and Fieguth, 2007, Elshinawy et al., 2010].
- Se utilizan medidas estadísticas para extraer información [Mughal and Ikram, 2004, Assefa et al., 2013, Mukherjee et al., 2014, Punithavathy et al., 2015, Ignatious and Joseph, 2015].

Estado del arte

- Se utiliza una red perceptrón multicapa para la clasificación [Zamani M., 2009, Terdale and Kulhalli, 2012, Ada, 2013, Ashwin et al., 2012].
- Optimización de enjambre de partículas es utilizado en problemas de optimización [Nielsen, 1987, Zamani M., 2009].

Contents

Introducción

Cáncer Pulmonar

Trabajos relacionados

Marco teórico

Propuesta

Experimentación

Conclusión

Red Neuronal

- Modelo de aprendizaje y procesamiento automático inspirado en sistemas nerviosos biológicos como el cerebro que procesa información[Haykin, 1998].
- Una red neuronal puede “pensar” como un experto en la categoría que ésta ha sido especializada, puede brindar respuesta a nuevos casos [Stergiou C., 2015].

Neurona

Unidad fundamental de procesamiento de información en la red neuronal. En la figura 3 se muestra la base para el diseño de las redes neuronales.

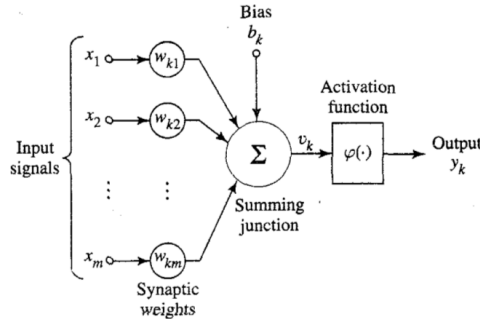


Figure 3: Modelo no lineal de una neurona [Haykin, 1998]

Arquitectura Redes

Redes mono-capa con pre-alimentación

- Una capa de entrada que se proyecta sobre la capa de salida, pero no viceversa.
- Llamadas redes mono-capa debido a la única capa de salida donde se realizan los cálculos.

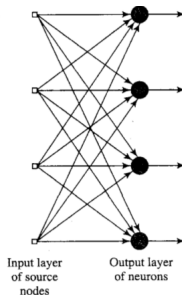


Figure 4: Red acíclica o de pre-alimentación con una sola capa [Havkin, 1998]

Arquitectura Redes

Redes recurrentes

- Tiene al menos un ciclo de retro-alimentación.
- La presencia de ciclos de retro-alimentación tiene un profundo impacto en la capacidad de aprendizaje de la red y en su desempeño.
- Los ciclos involucran el uso de elementos de unidades de retraso (denotados por z^{-1} , que resulta en un compartamiento dinámico no lineal.

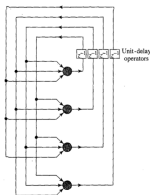


Figure 6: Red acíclica o totalmente conectada con una capa oculta y una capa de salida [Haykin, 1998].

Paradigmas de aprendizaje

Aprendizaje supervisado

- El profesor tiene el conocimiento del entorno, que es representado por ejemplos entrada-salida.
- Los parámetros de la red neuronal son ajustados bajo la influencia del vector de entrenamiento y la señal de error.

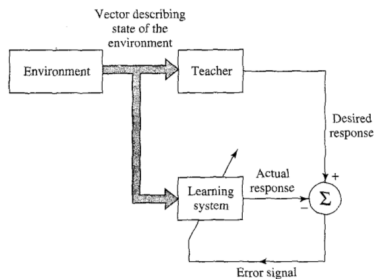


Figure 7: Diagrama de bloques de Aprendizaje Supervisado [Haykin, 1998].

Paradigmas de aprendizaje

Aprendizaje no supervisado

- La red es afinada a las regularidades estadísticas de los datos de entrada, desarrolla la habilidad para formar representaciones internas para codificar las características de la entrada y por tanto crear nuevas clases automáticamente.

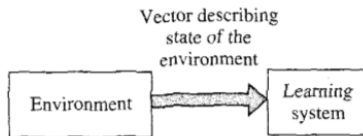


Figure 8: Diagrama de bloques de Aprendizaje No Supervisado [Haykin, 1998].

Perceptrón Multicapa

- Una capa de entrada, una o más capas ocultas de nodos de cálculo y una capa de salida de nodos de cálculo.
- La señal de entrada se propaga a través de la red hacía adelante.
- Han sido aplicadas satisfactoriamente para solver algunos problemas difíciles y diversos al entrenarlas con un algoritmo muy utilizado *algoritmo back-propagation*.

Computación evolutiva

Métodos de optimización global con un carácter metaheurístico u optimización estocástica, donde se utiliza una población de soluciones candidatas (iterando sobre el espacio de búsqueda). Son aplicados a problemas de caja negra en contextos de optimización costosa.

- Optimización de colonia de hormigas
- Colonia de abejas artificiales
- Evolución diferencial
- Algoritmos genéticos
- Optimización por enjambre de partículas

Optimización por Enjambres de Partículas

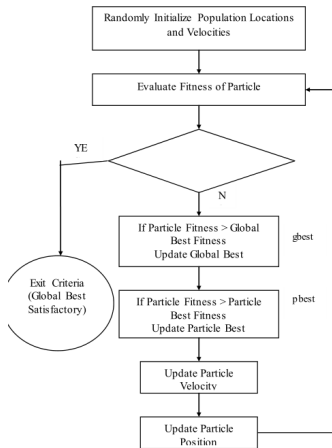


Figure 9: Algoritmo OEP tomado de [King K., 2009].

Segmentación

Umbralización

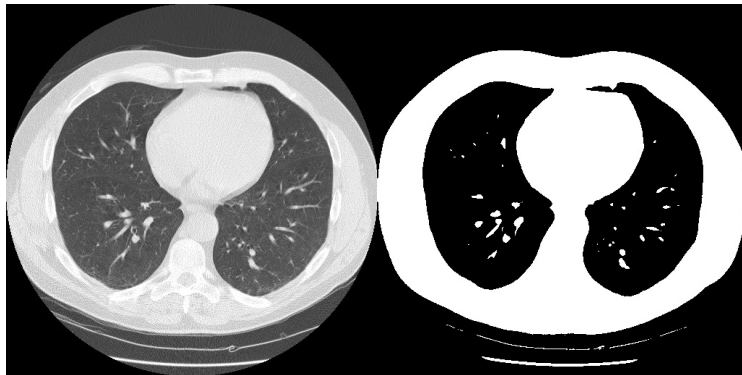


Figure 10: Imagen original e imagen binaria

Extracción de características

Patrón binario local

- Este operador fue definido originalmente para vecindades de 3×3 , dando códigos de 8 bits basados en los 8 píxeles alrededor del píxel central.

$$LBP(x_c, y_c) = \sum_{n=0}^7 2^n s(g_n - g_c) \quad (1)$$

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \quad (2)$$

Extracción de características

Histograma BPL

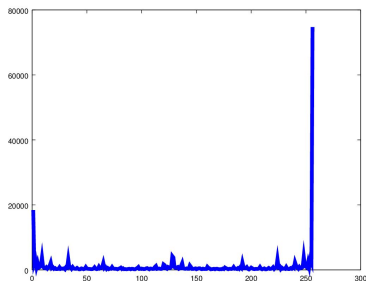


Figure 11: Histograma BPL de imagen 000038

Medición de la clasificación

- True Positive : cuando se predice a un maligno como maligno.
- False Negative: cuando se predice a un maligno como benigno.
- True Negative: cuando se predice a un benigno como benigno.
- False Positive: cuando se predice a un benigno como maligno.

$$\text{Sensibilidad} = \frac{TP}{TP + FN} \times 100\% \quad (3)$$

$$\text{Especificidad} = \frac{TN}{TN + TP} \times 100\% \quad (4)$$

$$\text{Sensibilidad} = \frac{TP + TN}{TN + TP + FN + FP} \times 100\% \quad (5)$$

Contents

Introducción

Cáncer Pulmonar

Trabajos relacionados

Marco teórico

Propuesta

Experimentación

Conclusión

¿Por qué híbrido?

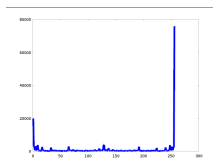
De acuerdo a la RAE:
(Del lat. hybrīda).

1. adj. Dicho de un animal o de un vegetal: Procreado por dos individuos de distinta especie. U. t. c. s.
2. adj. Biol. Dicho de un individuo: Cuyos padres son genéticamente distintos con respecto a un mismo carácter.
3. adj. Se dice de todo lo que es producto de elementos de distinta naturaleza.

Modelo híbrido

La propuesta que se plantea para la clasificación de las imágenes tomográficas utiliza dos técnicas de inteligencia artificial: perceptrón multicapa y optimización por enjambre de partículas. La primera para realizar la clasificación y la segunda para mejorar el tiempo de entrenamiento e influir en la clasificación. Al usar dos técnicas de un área se planteó denominarlo modelo híbrido.

Modelo 1



0.0038 0.0003 14.1938 211.41 0.0928 5.9253

Modelo 2



2.206 403.816 9.322 87.357 1.071e+08 -4.337e+06 3.898e-05 46.807

Contents

Introducción

Cáncer Pulmonar

Trabajos relacionados

Marco teórico

Propuesta

Experimentación

Conclusión

Comparación de resultados

A continuación la tabla que resume los resultados de la propuesta y los de otros autores.

Table 1: Tabla resultados - Nro Características(NC) Sensibilidad(SE), Especificidad(ES), Precisión(PR)

Autor	Total	NC	SE(%)	ES(%)	PR(%)
[Terdale and Kulhalli, 2012]	65	2	82.14	66.6	66
[Azian Azamimi Abdullah, 2012]	30	3	-	-	80
[Ada, 2013]	909	16	-	-	96.04
[Ashwin et al., 2012]	40	-	92.1	94.3	96.7
[Assefa et al., 2013]	50	-	-	-	81
[Ignatious and Joseph, 2015]	-	3	-	-	90.1
Modelo 1	93	6	0	100	98.92
Modelo 2	93	8	100	98.91	98.92

Convergencia

Modelo 1

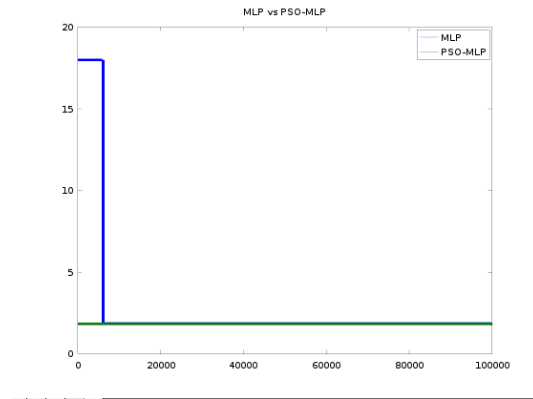


Figure 12: Convergencia de PMC y OEP-PMC

Convergencia

Modelo 2

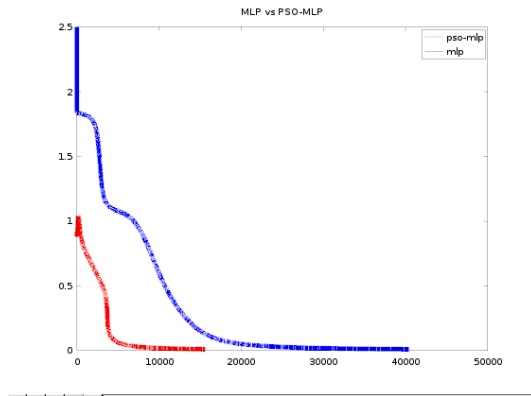


Figure 13: Convergencia de PMC y OEP-PMC

Contents

Introducción

Cáncer Pulmonar

Trabajos relacionados

Marco teórico

Propuesta

Experimentación

Conclusión

Conclusiones

Se logra modelar, implementar y evaluar un clasificador de tomografías axiales computarizadas utilizando umbralización y criterios geométricos, partición binaria local y medidas estadísticas, enjambre de partículas y redes neuronales PMC.

- Utilizar umbralización y criterios geométricos nos permite aislar el área de interés para su posterior análisis.
- Utilizar partición binaria local para generar una nueva imagen y usar medidas estadísticas extrae información relevante.
- Optimización de enjambre de partículas logra una mejor convergencia de la red PMC, disminuyendo el número de etapas en el entrenamiento.
- El uso de una red PMC en la clasificación de imágenes(TAC), nos brinda un resultado satisfactorio y confirma su capacidad de clasificación.
- El número de capas y nodos por capa de la red neuronal son definidos de acuerdo a la naturaleza del problema, los mismos influirán en el entrenamiento y en la clasificación posterior, se utilizó el criterio $2n + 1$.

Recomendaciones

- Usar formatos de imagen que tengan menor pérdida, para tener la imagen con mayor cantidad de información.
- Utilizar un método de extracción de características que sea lo suficientemente discriminativo para obtener una buena precisión en la clasificación.
- Tener cuidado con los tamaños de los punteros, las declaraciones de estructuras de opencv y el uso de sus diferentes funciones; revisando su manual o la página misma.





Comentarios finales

El presente trabajo cumple con su cometido, experimentar con técnicas de inteligencia artificial como la perceptrón multicapa, optimización de enjambre de partícula y alcanzar una clasificación satisfactoria de las imágenes tomográficas. Permite ahondar en los conocimientos sobre procesamiento de imágenes y aprendizaje de máquina, de esta forma sentar las bases para proyectos posteriores relacionados.

Trabajos Futuros

Se pueden realizar muchos trabajos partiendo de [WHO, 2011] , pues hay varias investigaciones sobre cáncer pulmonar y otros tipos de cáncer; recordando que cada uno tiene una naturaleza diferente.

Usar otros métodos para extraer características, encontrar las regiones y otro clasificador. OPF (optimum path forest) es un clasificador que ha tenido buenos resultados con imágenes y videos, fue utilizado en diversos trabajos presentados en conferencias internacionales. Existen algoritmos de computación evolutiva que pueden mejorar el entrenamiento de la red (algoritmos evolutivos cuánticos, etc.)

-  **Ada, R. K. (2013).**
Early detection and prediction of lung cancer survival using neural network classifier.
-  **Ashwin, S., Ramesh, J., Kumar, S. A., and Gunavathi, K. (2012).**
Efficient and reliable lung nodule detection using a neural network based computer aided diagnosis system.
In Emerging Trends in Electrical Engineering and Energy Management (ICETEEEM), 2012 International Conference on, pages 135–142.
-  **Assefa, M., Faye, I., Malik, A. S., and Shoaib, M. (2013).**
Lung nodule detection using multi-resolution analysis.
In Complex Medical Engineering (CME), 2013 ICME International Conference on, pages 457–461.
-  **Ayman El-Baz, Garth M. Beache, G. G. (2013).**
Computer-aided diagnosis systems for lung cancer: Challenges and methodologies.

International Journal of Biomedical Imaging.



Azian Azamimi Abdullah, S. M. S. (2012).

Lung cancer cell classification method using artificial neural network.

Information Engineering Letters.



Elshinawy, M., Abdelmageed, W., Badawy, A.-H., and Chouikha, M. (2010).

Pre-cad system for normal mammogram detection using local binary pattern features.

In Computer-Based Medical Systems (CBMS), 2010 IEEE 23rd International Symposium on, pages 352–357.



Haykin, S. (1998).

Neural Networks: A Comprehensive Foundation.





Prentice Hall PTR, Upper Saddle River, NJ, USA, 2nd edition.












Herman, G. (2009).

Fundamentals of computerized tomography: Image reconstruction from projection (2nd edition).

Springer.

-  Ignatious, S. and Joseph, R. (2015).
Computer aided lung cancer detection system.
In Communication Technologies (GCCT), 2015 Global Conference on, pages 555–558.
-  INC (2015).
¿qué es el cáncer?
-  Kachouie, N. and Fieguth, P. (2007).
A medical texture local binary pattern for trus prostate segmentation.
In Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE, pages 5605–5608.
-  King K., Harun S., M. S. (2009).
Particle swarm optimization feedforward neuralnetwork for hourly rainfall-runoff modeling in bedup basin, malaysia.
International Journal of Civil and Environmental Engineering.

-  [Lange \(2004\).](#)
Thorax ct peripheres bronchial carcinom image.
-  [Mikael Berndtsson, Jörgen Hansson, B. O. B. L. \(2008\).](#)
Thesis projects: a guide for students in computer science and information systems.
[Springer.](#)
-  [Mughal, M. N. and Ikram, W. \(2004\).](#)
Early lung cancer detection by classifying chest ct images: a survey.
In Multitopic Conference, 2004. Proceedings of INMIC 2004. 8th International,
pages 67–72.
-  [Mukherjee, J., Chakrabarti, A., Shaikh, S. H., and Kar, M. \(2014\).](#)
Automatic detection and classification of solitary pulmonary nodules from lung ct images.
In Emerging Applications of Information Technology (EAIT), 2014 Fourth International Conference of, pages 294–299.

-  Nielsen, R. H. (1987).
Kolmogorov's mapping neural network existence theorem.
In Proceedings of the IEEE First International Conference on Neural Networks
(San Diego, CA), volume III, pages 11–13. Piscataway, NJ: IEEE.
-  Punithavathy, K., Ramya, M. M., and Poobal, S. (2015).
Analysis of statistical texture features for automatic lung cancer detection in
pet/ct images.
*In Robotics, Automation, Control and Embedded Systems (RACE), 2015
International Conference on*, pages 1–5.
-  Stergiou C., S. D. (2015).
Neural networks.
-  Terdale, S. B. and Kulhalli, K. (2012).
Cad system for lung cancer detection using ann.
-  WHO (2008).
The global burden disease: 2004 update.



WHO (2011).

Noncommunicable diseases country profiles 2011.



Zamani M., S. A. (2009).

A variation of particle swarm optimization for training of artificial neural networks.

International Journal of Application or Innovation in Engineering and Management.