

Trabajo de Fin de Grado

Detección de distintos tipos de cáncer mediante Redes Neuronales Artificiales

*Cancers detection using Artificial Neural Networks*

Alberto Fariña Barrera

\*La Laguna, 4 de marzo de 2015

D. **Patricio García Báez**, con N.I.F. 43.356.987-D profesor Titular de Universidad adscrito al Departamento de \*Nombre del Departamento de la Universidad de La Laguna, como tutor

D. **Carmen Paz Suárez Araujo**, con N.I.F. 43.640.373-N profesor Titular de Universidad adscrito al Departamento de \*Nombre del Departamento de la Universidad de La Laguna, como cotutor

**C E R T I F I C A (N)**

Que la presente memoria titulada:

*“Detección de distintos tipos de cáncer mediante Redes Neuronales Artificiales”*

ha sido realizada bajo su dirección por D. **Alberto Fariña Barrera**, con N.I.F. 43.836.555-G.

Y para que así conste, en cumplimiento de la legislación vigente y a los efectos oportunos ﬁrman la presente en La Laguna a \*4 de marzo de 2015.

Agradecimientos

Me gustaría agradecer a mis tutores Patricio García Báez y Carmen Paz Suárez Araujo por el apoyo y la ayuda prestada durante el desarrollo del trabajo y de esta memoria. También me gustaría agradecer a mi compañero Óscar Trujillo Acosta por haberme ayudado en algunos aspectos del proyecto. Agradecer a mi amiga Rita Hernández Pérez, estudiante del Grado de Medicina por el apoyo y la ayuda a entender la parte médica del proyecto. Por último, a mi familia por haberme ayudado a llegar hasta aquí.

Licencia

C:\Users\Usuario\Desktop\memoriaTFG_1415_LaTeX\tfg\images\by-nc_88x31.png

© Esta obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial 4.0 Internacional.

**Resumen**

El objetivo de este trabajo ha sido el desarrollo de tres Redes Neuronales Artificiales (RNA), una para cada tipo de cáncer tratado, que son: Cáncer de mama, Melanoma y Cáncer de pulmón.

Se dispone para el proyecto de una base de datos proporcionada por el Heuristic and Evolutionary Algorithms Laboratory (HEAL). Dicha base de datos tiene tres tablas, una para cada uno de los tres tipos de cáncer citados anteriormente que contendrán datos de pacientes con cáncer y pacientes sanos para que la red pueda ser entrenada y validada. Se llevarán a cabo estudios con tres redes o algoritmos distintos: Back-propagation para entrenamiento supervisado, Deep Belief Network para entrenamiento no supervisado y Support Vector Machine como problema de clasificación.

El uso de este sistema no pretende ser el de un diagnosticador final, sino más bien una herramienta de soporte para los oncólogos, que les permita hacer un diagnóstico inicial del paciente y derivar a pruebas más concienzudas.

**Palabras clave:** *Red Neuronal Artificial, Back-Propagation, Deep Belief Network, Support Vector Machine, Detección de cáncer.*

**Abstract**

The goal of this project has been the development of three Artificial Neural Networks, or ANN as its acronym, that identifies the most accurate way possible the presence of different types of cancer in a person, such as breast cancer, melanoma or respiratory system cancer.

A database is provided for the project by Heuristic and Evolutionary Algorithms Laboratory (HEAL). This database contains three tables, one for each type of cancer named before with data about patients with and without cancer so the network can be trained and validated. There will be studies for three different types of network or algorithms: Back-propagation for supervised training, Deep Belief Network for unsupervised training and Support Vector Machines as classification problem.

This system doesn’t pretend to be a final test, but more like a support tool for oncologists, that allow them to do an initial test and refer to more important tests if necessary.

***Keywords:*** *Artificial Neural Network, Back-Propagation, Deep Belief Network, Support Vector Machine, Cancer detection.*

Índice General

[Capítulo 1. Introducción 1](#_Toc452577522)

[Capítulo 2. Antecedentes 2](#_Toc452577523)

[Capítulo 3. Estado del arte 3](#_Toc452577524)

[Capítulo 4. Objetivos 4](#_Toc452577525)

[Capítulo 5. Fases y desarrollo del proyecto 5](#_Toc452577526)

[Capítulo 6. Conclusiones y líneas futuras 6](#_Toc452577527)

[Capítulo 7. Summary and Conclusions 7](#_Toc452577528)

[7.1 First Section 7](#_Toc452577529)

[Capítulo 8. Presupuesto 8](#_Toc452577530)

[8.1 Sección Uno 8](#_Toc452577531)

[Apéndice A. Título del Apéndice 1 9](#_Toc452577532)

[A.1. Algoritmo XXX 9](#_Toc452577533)

[A.2. Algoritmo YYY 9](#_Toc452577534)

[Apéndice B. Título del Apéndice 2 11](#_Toc452577535)

[B.1. Otro apendice: Seccion 1 11](#_Toc452577536)

[B.2. Otro apendice: Seccion 2 11](#_Toc452577537)

[Bibliografía 12](#_Toc452577538)

Índice de figuras

**No se encuentran elementos de tabla de ilustraciones.**

Índice de tablas

[Tabla 7.1. Tabla resumen de los Tipos. 8](#_Toc452563717)

## Introducción

En este trabajo se propone la elaboración de tres Redes Neuronales Artificiales, una para cada tipo de cáncer propuesto (Cáncer de Mama, Melanoma, Cáncer de Pulmón), que detecten la presencia de la enfermedad en un paciente con una alta probabilidad de acierto. Para ello se llevarán a cabo estudios sobre varios tipos de redes neuronales: RNA con Back-Propagation como algoritmo de entrenamiento, Support Vector Machine y Deep Belief Network para entrenamiento no supervisado.

Se utiliza como herramienta PyBrain, una librería de RNA sobre el lenguaje Python, que permite que la creación y entrenamiento de redes se haga de manera sencilla e intuitiva

Para elaborar la red neuronal más eficiente posible se realizan análisis en profundidad para determinar el valor de las variables características de cada uno de los algoritmos, para finalmente realizar una comparativa entre los tres algoritmos, para seleccionar el que mejor determine si un paciente tiene cáncer o no.

El resultado de este proyecto tiene como objetivo servir de ayuda a médicos a la hora de obtener un primer diagnóstico rápido de un paciente, para, si es necesario, derivar a pruebas más concienzudas con la mayor brevedad posible.

## Antecedentes

Existen varios casos de software de diagnosis de cáncer a partir de bases de datos, tal vez el más conocido sea el “Global Neural Network Cloud Service for Breast Cancer” de Brittany Wenger, que es capaz de detectar si un tumor es maligno con un 99% de acierto. En el caso de este proyecto, las diferencias son que se hará para tres tipos de cáncer y que las bases de datos usadas no serán tan grandes, por lo que los resultados pueden no ser tan altos.

A su vez, el Heuristic and Evolutionary Algorithm Laboratory (HEAL), el mismo laboratorio de investigación que cedió las bases de datos para este proyecto, elaboró un diagnosticador basado en las mismas bases de datos, pero elaborado con árboles de decisión genéticos, cuyos resultados se compararán con los resultados de este trabajo en uno de los capítulos.

Otra de las variables posibles del uso de redes neuronales la vemos en el caso de dos profesores y un alumno del Instituto Politécnico Nacional de México, que crearon una red neuronal que detecta tumores malignos a partir de las imágenes de mamografías, por lo que no solo con datos empíricos puede entrenarse una red neuronal aplicada a la medicina, también con datos extraídos de una imagen digital.

En conclusión, las muchas y variadas incursiones de las redes neuronales en el mundo de la medicina, y más concretamente en el tema de la oncología, proporciona una gran base con la que comparar los resultados de este proyecto para evaluar su éxito.

## Estado del arte

## Objetivos

Los capítulos intermedios servirán para cubrir los siguientes aspectos: antecedentes, problemática o estado del arte, objetivos, fases y desarrollo del proyecto.

En el capítulo 1 se describió bla, bla, bla.....

## Fases y desarrollo del proyecto

## Conclusiones y líneas futuras

Este capítulo es obligatorio. Toda memoria de Trabajo de Fin de Grado debe incluir unas conclusiones y unas líneas de trabajo futuro.

## Summary and Conclusions

This chapter is compulsory. The memory should include an extended summary and conclusions in English.

### First Section

## Presupuesto

Este capítulo es obligatorio. Toda memoria de Trabajo de Fin de Grado debe incluir un presupuesto.

### Sección Uno

|  |  |
| --- | --- |
| Tipos | Descripción |
| AAA | BBB |
| CCC | DDD |
| EEE | FFF |
| GGG | HHH |

|  |  |
| --- | --- |
| Tipos | Descripción |
| AAA | BBB |
| CCC | DDD |
| EEE | FFF |
| GGG | HHH |

Tabla 7.1. Tabla resumen de los Tipos.

## Apéndice A. Título del Apéndice 1

### A.1. Algoritmo XXX

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*

\* Fichero .h

\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*

\* AUTORES

\*

\*

\* FECHA

\*

\*

\* DESCRIPCION

\*

\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

### A.2. Algoritmo YYY

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*

\* Fichero .h

\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*

\* AUTORES

\*

\*

\* FECHA

\*

\*

\* DESCRIPCION

\*

\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

## Apéndice B. Título del Apéndice 2

### B.1. Otro apendice: Seccion 1

Texto

### B.2. Otro apendice: Seccion 2

Texto

## Bibliografía

1. ACM LaTeX Style. http://www.acm.org/publications/latex style/.
2. FACOM OS IV SSL II USER’S GUIDE, 99SP0050E5. Technical report, 1990.
3. D. H. Bailey and P. Swarztrauber. The fractional Fourier transform and applications. *SIAM Rev.*, 33(3):389–404, 1991.
4. A. Bayliss, C. I. Goldstein, and E. Turkel. An iterative method for the Helmholtz equation. *J. Comp. Phys.*, 49:443–457, 1983.
5. C. Darwin. *The Origin Of Species*. November 1859.
6. C. Goldstein. Multigrid methods for elliptic problems in unbounded domains. SIAM, *J. Numer. Anal.*, 30:159–183, 1993.
7. P. Swarztrauber. *Vectorizing the FFTs*. Academic Press, New York, 1982.
8. S. Taásan. *Multigrid Methods for Highly Oscillatory Problems*. PhD thesis, Weizmann Institute of Science, Rehovot, Israel, 1984.
9. Berndtsson, M., Hansson, J., Olsson, B., Lundell, B. (2008), *A Guide for Students in Computer Science and Information Systems*, Springer.