

PROYECTO DE TRABAJO DE FIN DE GRADO

Apellidos y nombre del alumno: FARIÑA BARRERA, ALBERTO
DNI: 43836555-G

Apellidos y nombre del tutor: GARCÍA BÁEZ, PATRICIO
DNI: 43356987-D

Apellidos y nombre de la co-tutora: SUÁREZ ARAUJO, CARMEN PAZ
DNI: 43640373-N

Título del Proyecto: “Detección de distintos tipos de cáncer mediante Redes Neuronales Artificiales”

1. Introducción

En este proyecto se plantea el desarrollo de una Red Neuronal Artificial, o RNA según sus siglas, que identifique de la manera más precisa posible la presencia de distintos tipos de cáncer en una persona, ya sea cáncer de pulmón, cáncer de piel o cáncer de mama.

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. Cada una de estas contendrá datos de pacientes con cáncer y sujetos sanos para que la red pueda ser entrenada y validada. En concreto más de 2000 casos por cada tipo de cáncer que incluyen datos de análisis de sangre, como colesterol, distintas proteínas, etc.

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. Este proyecto se encuadra en la línea de trabajo "Redes Neuronales Artificiales aplicadas a la medicina" propuesta por los tutores, que consta de dos TFGs aceptados y en realización. Ambos se diferenciarán en los preprocesamientos, el software de RNAs a usar y los modelos de RNAs a utilizar, teniendo en común los problemas abordados y las bases de datos para el aprendizaje.

A su vez, se señala la intención de incluir este proyecto en el Programa de Apoyo a Trabajos Finales Libres de la Oficina de Software Libre de la ULL.

2. Antecedentes y estado actual del tema

El campo de las redes neuronales tuvo gran repercusión en los años 50 y 60, hasta que la capacidad de los ordenadores de la época hizo imposible continuar con su expansión. En los años 80 se reanuda con fuerza su investigación, constituyendo actualmente uno de los contenidos más

importantes de la computación inteligente. Uno de los campos de gran proliferación de las redes neuronales es la medicina, en el ámbito de los sistemas de ayuda a la decisión clínica, entre otros. Ello es debido a su capacidad de aprendizaje y su habilidad para procesar datos multidimensionales, entornos complejos, cambiantes, ruidosos, pudiendo alcanzar un diagnóstico aproximado en tiempo real.

El diagnóstico de cáncer con RNA se han seguido distintas aproximaciones, por ejemplo el caso de los doctores Rolando Flores Carapia y Benjamín Luna Benoso y el alumno Hugo Flores Gutiérrez de la Universidad Politécnica Nacional de México, que desarrollaron una red neuronal para detectar cáncer de mama a partir de imágenes de mamografías.

El mismo laboratorio que ha cedido la base de datos de pacientes ha desarrollado un sistema similar al que proponemos, pero utilizando árboles de decisión genéticos en lugar de redes neuronales. Los resultados de este sistema se compararán con los obtenidos en la propuesta en este Trabajo de Fin de Grado.

3. Actividades a realizar

1. Preparación

- 1.1. Búsqueda de información acerca de RNAs
- 1.2. Búsqueda y elección de herramientas para la realización de las RNA
- 1.3. Estudio del trabajo realizado por el laboratorio austriaco (HEAL)

2. Preprocesado

- 2.1. Análisis de las bases de datos que se utilizarán en el proyecto para detectar posibles dudas y resolverlas.
- 2.2. Elección de herramienta de preprocesado para las bases de datos
- 2.3. Preprocesamiento de la base de datos para eliminar datos nulos, campos innecesarios y ajustar valores
- 2.4. Selección de las particiones de la base de datos que se usarán para entrenamiento, validación y testeo

3. Diseño y Desarrollo

- 3.1. Elección del modelo de RNA que se utilizará en el proyecto. (Redes de base radial, Perceptrón Multicapa,...)
- 3.2. Diseño de la red
- 3.3. Entrenamiento y validación de la RNA.
- 3.4. Optimización de los parámetros de la RNA

4. Evaluación de la red neuronal

- 4.1. Realización de pruebas con los datos de la partición de testeo
- 4.2. Creación de gráficas de errores y precisión para la presentación de la RNA
- 4.3. Comparación de resultados con el laboratorio HEAL

5. Finalización

5.1. Preparación de la memoria del TFG

5.2. Preparación de la defensa de la memoria del TFG

4. Plan de Trabajo

A continuación se muestran las tareas especificadas en el apartado anterior junto con unas fechas estimadas en las que se llevarán a cabo:

Nombre	Duración	Inicio	Terminado
Preparación	21 days?	27/01/16 8:00	24/02/16 17:00
Búsqueda de información acerca de RNAs	8 days?	27/01/16 8:00	5/02/16 17:00
Búsqueda y elección del programa para las RNA	11 days?	5/02/16 8:00	19/02/16 17:00
Estudio del trabajo realizado por el HEAL	4 days?	19/02/16 8:00	24/02/16 17:00
Preprocesado	25 days?	24/02/16 8:00	29/03/16 17:00
Análisis de las bases de datos cedidas por el HEAL	6 days?	24/02/16 8:00	2/03/16 17:00
Elección de la herramienta de preprocesado	8 days?	2/03/16 8:00	11/03/16 17:00
Preprocesamiento de la base de datos	11 days?	11/03/16 8:00	25/03/16 17:00
Selección de las particiones de la base de datos	3 days?	25/03/16 8:00	29/03/16 17:00
Diseño y desarrollo	24 days?	29/03/16 8:00	29/04/16 17:00
Elección del modelo de RNA	6 days?	29/03/16 8:00	5/04/16 17:00
Diseñar la red	6 days?	5/04/16 8:00	12/04/16 17:00
Entrenamiento y validación de la red	4 days?	12/04/16 8:00	15/04/16 17:00
Optimización de la red	11 days?	15/04/16 8:00	29/04/16 17:00
Evaluación de la red neuronal	20 days?	29/04/16 8:00	26/05/16 17:00
Realización de pruebas con los datos para testeo	11 days?	29/04/16 8:00	13/05/16 17:00
Creación de gráficas de errores y precisión	5 days?	13/05/16 8:00	19/05/16 17:00
Comparación con los resultados del HEAL	6 days?	19/05/16 8:00	26/05/16 17:00
Finalización	17 days?	26/05/16 8:00	17/06/16 17:00
Preparación de la memoria del TFG	12 days?	26/05/16 8:00	10/06/16 17:00
Preparación de la defensa de la memoria del TFG	6 days?	10/06/16 8:00	17/06/16 17:00

5. Propuesta de evaluación

En este punto se hará una propuesta de evaluación teniendo en cuenta las tareas que se han expuesto en los puntos anteriores. Por lo tanto, la evaluación consistirá en tres módulos diferenciados que se repartirán el porcentaje de la nota de la siguiente manera:

Módulo	Porcentaje
Preparación	20%
Preprocesado	20%
Diseño y Desarrollo	30%
Evaluación de la red	10%
Finalización	10%

Para la distribución de pesos, se ha dado mayor importancia al desarrollo de la aplicación por considerar que su finalización es esencial para superar la asignatura. La suma de todos los módulos supondría un 90% en la nota final del Trabajo Fin de Grado, dejando un 10% para la evaluación que hace el Tutor del proyecto.

La Laguna, 29 de abril de 2016

Fdo.:

A handwritten signature in black ink, appearing to be 'Alonso', written over a horizontal line.
