



Uso de redes BiLSTM para predecir la confusión

Universidad del Quindío
Facultad de Ciencias Básicas y Tecnologías
Programa de Física

Belky Alejandra¹, Cristian F. Zapata²

¹belky.marulandac@uqvirtual.edu.co, ²cfzapata@uqvirtual.edu.co,
23 de Mayo del 2023

Introducción

La electroencefalografía (EEG) es una técnica neurofisiológica que consiste en el registro de actividad cerebral. El cerebro es un órgano constituido por neuronas, las cuáles se comunican por medio de impulsos eléctricos. Utilizando electrodos se puede medir las variaciones de los potenciales en las membranas de las neuronas, y así medir la actividad cerebral. Esta técnica ha resultado en un sin fin de aplicaciones: desde el tratamiento y diagnóstico de enfermedades, hasta la profundización del conocimiento de la neurología de los seres humanos [0].

En 2013 Haohan Wang y otros, publicaron un artículo llamado *Using EEG to improve massive open online courses feedback interaction* [1]. En este artículo plantean la idea de estudiar la confusión de estudiantes que toman clases virtuales, mediante EEG.

Es conocido que aspectos como la atención sostenida, la vigilia y los estados de relajación se pueden medir directamente con EEG. Pero, no se conocía evidencia de que aspectos tan subjetivos como la confusión pudiera medirse. En su artículo, Wang y su equipo, mostraron que sí es posible hacerlo. Y en añadidura, indican que un clasificador computacional podría predecir cuando un estudiante estaba confundido o no. Dicha predicción era comparable a la que podría hacer una persona que 'leía' el lenguaje corporal (del sujeto de estudio).

Este trabajo está inspirado en el de Wang y compañía: se tomó la base de datos que ellos construyeron en su experimento y se aplicó una red neuronal **LSTM bidireccional**, con el **objetivo** de predecir si un estudiante estaba confundido o no.

Este documento se divide en 4 secciones:

En la **primera**: se explica los cimientos científicos sobre las comunicaciones entre las neuronas, y cómo se adquiere las señales que se miden en la EEG. Esta sección se brinda, para que el lector esté en mayor contexto con la técnica que se usó para la adquisición de la base de datos. Pero, en caso de no ser de su interés, puede pasar de largo. No interfiere en la comprensión del desarrollo central.

En la **segunda**: se expone cuál fue el experimento que se utilizó para la adquisición de la base de datos.

En la **tercera**: se muestra las principales características de la red neuronal BiLSTM y se muestra qué resultados se obtuvieron con el trabajo.

En la **cuarta**: se exponen las conclusiones del trabajo.

Finalmente, se presenta las **perspectivas** del trabajo. En donde se proponen acciones que se podrían implementar en el futuro, para mejorar un trabajo del mismo índole.

1. Electroencefalografía (EEG)

Como se dijo al inicio, la electroencefalografía es una técnica cuyo propósito es *registrar* la actividad cerebral. Con **actividad cerebral** se refiere a los patrones de actividad eléctrica que se generan en el cerebro.

A continuación se explica brevemente, qué es y cómo se da la actividad cerebral.

La siguiente figura muestra dos neuronas que se pueden comunicar:

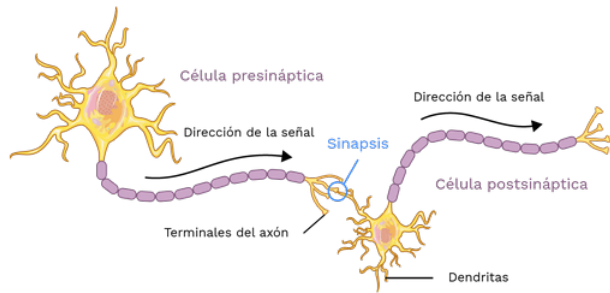


Figura 1: Imagen tomada de [2]

Supongase que la neurona **postsináptica** no ha experimentado ninguna señal por parte de la neurona **presináptica** (ni de ninguna otra). Para dicho caso, la neurona postsináptica, está rodeada de cargas positivas. En la mayoría de los casos estas cargas se tratan de iones de sodio (Na^+) y calcio (Ca^{2+}). Mientras que, en el interior de la célula hay carga negativas, debido a la presencia de proteínas y moléculas orgánicas cargadas negativamente, como los aniones orgánicos y los iones de proteínas. Por lo tanto, la **membrana** de la neurona, que es la estructura que delimita la célula del exterior, está a cierto potencial. Ese potencial oscila entre los -70mV (milivoltios). Dicho potencial se conoce como **potencial de reposo**.

Si ahora, la célula presináptica envía una señal¹ a la célula postsináptica, los receptores ubicados en las dendritas de estas últimas se abren, logrando así la entrada de iones positivos al interior de la célula. También, existen neurotransmisores que al llegar a una zona de las dendritas, activa otras zonas, logrando el mismo efecto.

Para este punto, el potencial en el que se encuentra la membrana de las dendritas, ya no es el potencial de reposo. Si la diferencia de potencial de la membrana en esa zona es cercano a cero (existe la misma cantidad de cargas negativas en el interior que en el exterior), ocurrió una **despolarización**. Cuando se da la despolarización en las dendritas, el **soma** de la neurona, desencadena la apertura de receptores a lo largo del **axón**, haciendo que de forma consecutiva vayan entrando cargas positivas al axón.

A medida que entran las cargas positivas, el potencial neto en la neurona va cambiando. Se dice que

se están generando un **potencial de acción**. Ese potencial de acción se propaga a lo largo del axón de la neurona generando una corriente eléctrica. El **impulso nervioso** se refiere a la transmisión de esta señal eléctrica a lo largo del axón de la neurona. Cuando el impulso nervioso llega al final del axón, llamado botón **terminal** o **presináptico**, acciona la salida de vesículas. Dichas vesículas contienen neurotransmisores. Estos neurotransmisores, se vacían en el espacio sináptico. Reiniciando el proceso anteriormente explicado, pero en otra neurona [3],[4].

Los electrodos de EEG son pequeños sensores de metal que se colocan en la cabeza del paciente (en zonas específicas y estandarizadas mundialmente). Estos electrodos están diseñados para captar las fluctuaciones en el voltaje eléctrico entre diferentes puntos de la cabeza. Cada electrodo del EEG está conectado a un amplificador que amplifica la señal eléctrica captada. Estos amplificadores registran las fluctuaciones de voltaje y las convierten en señales eléctricas más grandes y medibles. Las señales amplificadas son luego transmitidas a un equipo de registro del EEG, que registra y muestra la actividad eléctrica cerebral en forma de trazos gráficos o ondas cerebrales en tiempo real. Estos trazos gráficos representan la actividad eléctrica sincronizada de muchas neuronas en el cerebro en un determinado momento [0]. Cada electrodo del EEG captura la actividad eléctrica de un área específica del cerebro. Al colocar múltiples electrodos en diferentes posiciones del cráneo, se puede obtener una imagen global de la actividad eléctrica cerebral y analizar patrones de ondas cerebrales en diferentes regiones.

Entre las aplicaciones que tiene la EEG, están: 1) El diagnóstico de trastornos neurológicos (como la epilepsia); 2) Diagnóstico de enfermedades cerebrales (tumores cerebrales, lesiones cerebrales); 3) Evaluación de trastornos del sueño; 4) Evaluación de trastornos psiquiátricos (depresión, esquizofrenia, TDAH) [5].

La aplicación del EEG más afín a este trabajo es: la investigación científica en neurociencia. La EEG puede ser utilizada para investigar la cognición, la percepción, la atención, el aprendizaje y la memoria, entre otros procesos mentales. Además, el EEG se utiliza en estudios de **neurofeedback**, donde se busca entrenar a las personas para que modifiquen su actividad cerebral y mejoren su funcionamiento cognitivo o emocional.

¹ Las señales entre neuronas son **neurotransmisores** que salen de los **terminales del axón**, pasan por el **espacio sináptico** y llegan a las dendritas de la célula postsináptica.

2. Base de datos

La base de datos con la que se trabajó fue construida por Haohan Wang y otros en 2013 [1].

Anteriormente [6], se encontró que unas de las principales causas ante las cuáles las personas permanecían renuentes a tomar cursos en línea, era la falta de retroalimentación instantánea. En el aprendizaje de forma presencial, los maestros (más o menos) pueden intuir cuando un alumno está teniendo problemas en la retención de la información, gracias a su lenguaje corporal. Signos como rascarse la cabeza, poner caras de desaprobación y fruncir el ceño, son indicadores de que el alumno está teniendo complicaciones. Normalmente, al advertir esto, los docentes incitan a preguntar. La conversación dinámica esclarece las dudas de los alumnos. Este acercamiento, claramente no se da en las clases en línea.

En su trabajo [1], Wang y su equipo querían responder dos preguntas:

- ¿Puede el EEG detectar la confusión?
- ¿Puede el EEG detectar la confusión mejor que los observadores humanos?

A fin de resolver los interrogantes, idearon el siguiente experimento:

Seleccionaron una **población** de 10 estudiantes universitarios.

Tomaron 20 vídeos propios de clases en línea (ofrecidos por Coursera o edX). De los 20 vídeos: 10 pertenecían a la categoría **fácil**, y los otros 10, a la categoría **difícil**. La primera categoría englobaba temas que (se esperaba), cualquier alumno universitario manejase, como la introducción al álgebra. La segunda categoría, en cambio, trataba temas más complejos del rango de la mecánica cuántica, y las células madres.

Cada vídeo se extrajo del medio de la lección (así, el estudiante no tenía ningún tipo de introducción y/o conclusión), y tenía una duración de tan solo 2 minutos.

Cada estudiante tuvo 10 secciones, 5 secciones para cada categoría. En cada sección, se le pedía a los alumnos que relajaran su mente durante 30 segundos. Además, se les pedía que aprendieran lo máximo posible de los vídeos.

Al terminar, los estudiantes debían puntuar cuán confundidos habían estado durante el vídeo. La escala correspondía de 1-7. Siendo 1 el nivel más bajo de confusión, y 7 el nivel más alto. Además, mientras los estudiantes veían los vídeos, tres personas

observaban su lenguaje corporal. A partir de ello, puntuaban la confusión del estudiante.

Para medir la actividad cerebral de los estudiantes, se utilizó el encefalógrafo **Mindset** inalámbrico (de la marca **Neurosky**). El Mindset, es de un solo canal, lo que significa que solo puede medir la actividad cerebral de una ubicación en específico. A saberse, la ubicación Fp1. Un electrodo en esa ubicación mide la actividad cerebral del lóbulo frontal. El estudio se hace en esa zona, ya que el EEG del lóbulo frontal puede ser utilizado para estudiar la atención, la concentración, el control de impulsos y la resolución de problemas. Ver figura 5.



Figura 2: Esta imagen muestra el encefalógrafo utilizado, y la posición de la ubicación de los electrodos. Imagen tomada de [7].

Lo que mide el Mindset es la diferencia de potencial entre el electrodo que descansa en la frente, y los otros dos electrodos que se posicionan en las orejas (uno es de referencia y el otro es la tierra). Para recopilar los datos, ellos utilizaron la API de Neurosky.

Wang y su equipo mostraron que si bien la confusión es un estado subjetivo, existen ciertas señales de EEG que tienen correlación con la confusión. Además, exhibieron que las señales de EEG 'detectan' (de forma indirecta) la confusión de forma comparativa con las observaciones humanas.

3. BiLSTM para predecir la confusión

- ¿Qué son las redes neuronales LSTM bidireccional (BiLSTM)?

Las redes bidireccionales LSTM son un tipo de arquitectura de redes neuronales recurrentes que se utilizan para procesar y modelar secuencias de datos en dos direcciones: hacia adelante y hacia atrás. A diferencia de las redes LSTM unidireccionales,

que solo consideran la información pasada o futura en un solo flujo de tiempo, las redes LSTM bidireccionales pueden capturar la información contextual en ambos sentidos, lo que les permite capturar relaciones complejas y mejorar el rendimiento en diversas tareas de procesamiento de secuencias [8], [9].

- ¿Cómo funcionan las BiLSTM [10]?

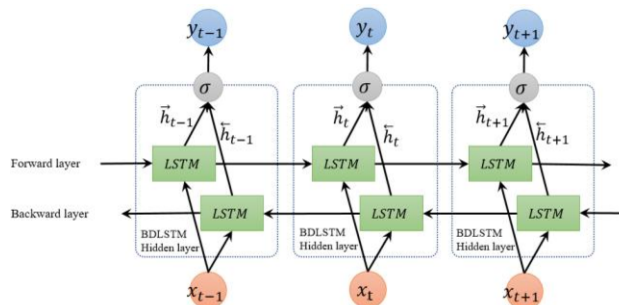


Figura 3: Esta imagen muestra la arquitectura básica de una red BiLSTM.
Imagen tomada de [9].

Todas las redes BiLSTM, siguen los siguiente paso básicos:

- La secuencia de entrada se 'dirige hacia' dos partes. Una parte es la que procesa la entrada en dirección hacia adelante y la otra, que la procesa en dirección hacia atrás.
- La primera parte alimenta a una serie de unidades LSTM en orden secuencial, desde el primer elemento hasta el último. Cada unidad LSTM en el lado hacia adelante recibe una entrada en el tiempo actual y la salida de la unidad anterior en el tiempo anterior. Esto permite que la red capture la información pasada y la procese de manera secuencial.
- La segunda parte se invierte, es decir, el último elemento se convierte en el primero, el penúltimo en el segundo, y así sucesivamente. Esta secuencia invertida alimenta a otra serie de unidades LSTM, pero esta vez en orden inverso, desde el último elemento hasta el primero. Cada unidad LSTM en el lado hacia atrás recibe una entrada en el tiempo actual y la salida de la unidad posterior en el tiempo siguiente. Esto permite que la red capture la información futura y la procese en orden inverso.
- Una vez que las secuencias de salida generadas por las unidades LSTM hacia adelante y

hacia atrás están disponibles, se pueden combinar de diversas formas, dependiendo de la tarea y los objetivos del modelo. Algunas opciones comunes incluyen concatenar las secuencias de salida para obtener una secuencia más larga que contenga información de ambas direcciones, sumar o promediar las salidas para obtener una representación combinada, o incluso utilizarlas como entradas separadas en una capa posterior de la red.

- ¿Por qué se eligió este tipo de redes para estudiar la base de datos?

Trabajar señales de EEG utilizando redes BiLSTM, brinda algunas ventajas [11]:

- Las redes neuronales del tipo LSTM son apropiadas para el tratamiento de secuencias y de series de tiempo, y justamente, las señales de EEG son secuencias temporales (de impulsos nerviosos).
- Al procesar las señales de EEG en ambas direcciones (hacia adelante y hacia atrás), las redes LSTM bidireccionales tienen la capacidad de capturar tanto la información pasada como la futura. Esto puede ser beneficioso para el análisis de señales de EEG, ya que se pueden tener en cuenta el contexto completo de las señales.
- En la gran mayoría de las veces, las EEG tienen relaciones temporales complejas. Por su parte, las BiLSTM son redes robustas (al capturar información contextual en ambas direcciones), así que usar ese tipo de redes hace que se potencie la precisión en las predicciones en las bases de datos EEG.

La siguiente imagen ilustra la arquitectura de la red neuronal que se construyó:

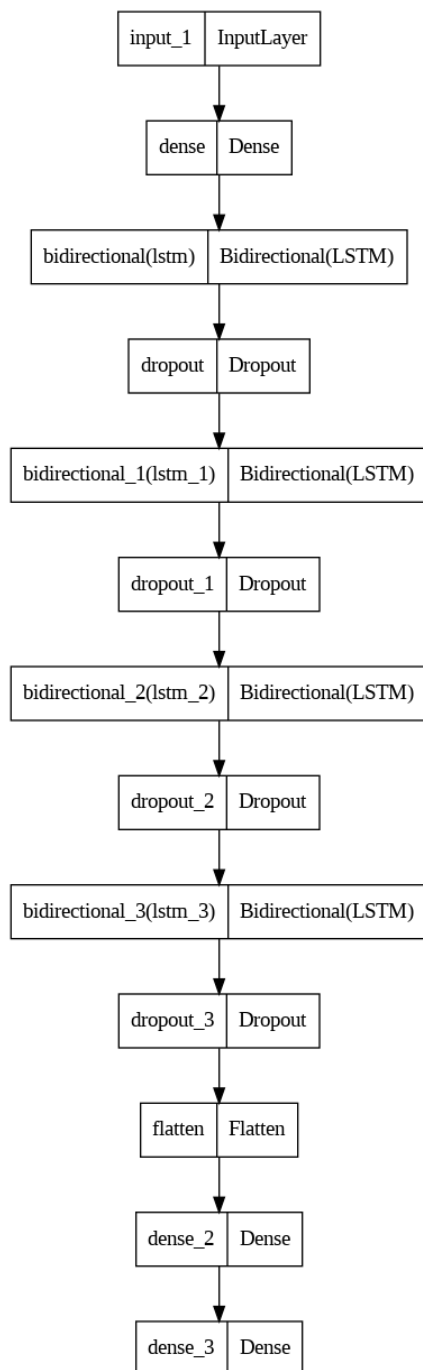


Figura 4: Arquitectura: En esta figura se puede ver la arquitectura del modelo, donde se uso **dropout** como método de regularización.

Como es habitual, al entrenar un modelo, es importante analizar los datos para comprender su comportamiento y determinar el procedimiento más adecuado. En primer lugar, se examinó si existía un desequilibrio de clases. La variable objetivo, que indica si una persona está confundida o no, tenía 6662 observaciones en las que el sujeto de prueba

no estaba confundido y 6149 observaciones en las que sí lo estaba. No se observó una desproporción significativa, por lo tanto, no fue necesario realizar un balanceo de clases previo al entrenamiento. Sin embargo, algunas características como la nacionalidad y el género presentaban desequilibrios considerablemente altos. Dado el enfoque del trabajo, estas variables no se eliminaron, pero tampoco se balancearon.

A continuación, se analizó la correlación entre las características. Esto es importante porque las variables altamente correlacionadas pueden afectar negativamente al modelo, ya que las variables colineales aportan poca información nueva y pueden conducir a problemas de overfitting, además de requerir una mayor capacidad computacional.

En el caso de las señales cerebrales registradas, se observó una alta correlación, lo cual es comprensible dado que se trata de una red neuronal en la que las señales emitidas por unas neuronas afectan a las vecinas. También se notó que aparentemente las señales registradas no estaban correlacionadas con la meditación y la concentración de los sujetos. Dado el alto nivel de correlación entre las características y la necesidad de que el modelo capture la información secuencial de las señales cerebrales, resulta necesario utilizar un método con una capacidad de memoria elevada. En este caso, las redes LSTM bidireccionales son una buena opción.

Otro procedimiento habitual al entrenar una red neuronal es dividir el conjunto de datos en conjuntos de entrenamiento y validación (85 % y 15 %, respectivamente). Se entrenó el modelo con el conjunto de entrenamiento y se utilizó el conjunto de validación para verificar que el modelo minimice la función de costo, que en este caso fue la entropía cruzada binaria. Se empleó la métrica de precisión (accuracy) y una tasa de aprendizaje de 0.001.

Como se puede observar en la figura 4, el modelo consta de varias capas y técnicas de regularización para abordar el problema del overfitting y mejorar el rendimiento. Después de la capa de entrada, se utilizó una capa densa con 64 nodos, seguida por una capa de 256 nodos de LSTM bidireccionales. Dado que algunas de las variables presentan alta correlación, se emplearon métodos de regularización para evitar el overfitting. En este caso, se utilizó una capa de Dropout con una tasa de 0,3, lo que significa que se desactivó aleatoriamente el 30 % de los nodos durante el entrenamiento. A continuación, se agregaron dos capas de LSTM bidireccionales con

512 nodos cada una, seguidas por una capa de 256 nodos, y se aplicaron capas adicionales de Dropout respectivamente. Luego, los datos se aplanaron y se conectaron con una capa densa de 128 nodos, utilizando la función de activación ReLU para mitigar el problema del desvanecimiento de gradientes. Finalmente, los datos pasaron por una única neurona con la función de activación sigmoide, ya que el objetivo es realizar una clasificación binaria entre 'confundido' y 'no confundido'. Esta configuración permite que el modelo aprenda y realice predicciones basadas en las características de entrada para clasificar adecuadamente los sujetos.

En la siguiente imagen se muestran las señales neuronales de dos personas, en donde la primera era clasificada como no confundida y la segunda como confundida.

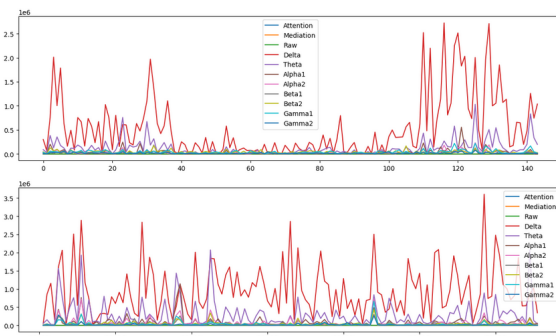


Figura 5: **Gráfico comparativo:** El gráfico superior corresponde a la respuesta de EEG para un estudiante que no estaba confundido; el segundo, para uno que sí.

Ese gráfico es uno en donde se relaciona amplitud de la señal contra frecuencia. Una persona que está confunda presenta una mayor amplitud para señales Delta, así como para señales Theta.

Finalmente, se construyó un modelo sustituto global con el objetivo de obtener información sobre la importancia de las características y su influencia en la confusión de las personas (ver figura 6). Como era de esperar, la atención desempeña un papel importante en la reducción de la confusión. Esto implica que cuando una persona está más atenta y enfocada, es menos propensa a experimentar confusión. Además, como se observó en el gráfico de las actividades cerebrales 5, las series Delta y Theta resultan ser muy relevantes para determinar si un individuo está confundido o no.

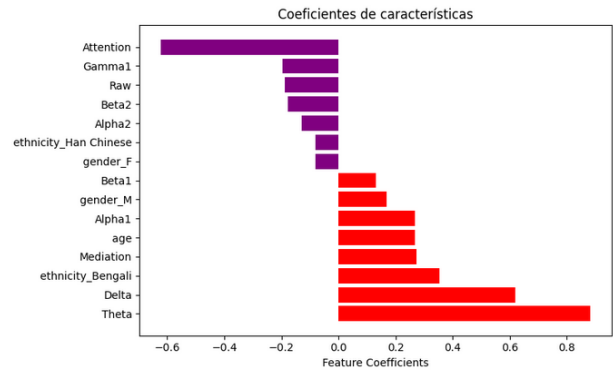


Figura 6: Modelo sustituto global: En este gráfico se puede ver la relevancia de las características para predecir la confusión.

4. Conclusiones

- Se encontró que la atención desempeña un papel importante en la reducción de la confusión: una persona enfocada es menos propensa a experimentar confusión. Además, se pudo corroborar de manera gráfica, y por medio del un modelo sustituto global, que las series Delta y Theta resultan ser muy relevantes para determinar la confusión.
- La serie Delta, se asocia con el sueño profundo y la relajación profunda. Su presencia en el contexto de confusión puede sugerir una falta de concentración y un estado mental más relajado o desconectado. Por otro lado, la serie Theta se asocia con la relajación, la meditación y el estado de ensoñación. Su presencia en el contexto de confusión puede indicar una mente más dispersa y menos enfocada en la tarea en cuestión.
- Estos hallazgos respaldan la idea de que el estado de atención y las características de las ondas cerebrales, especialmente las series Delta y Theta, desempeñan un papel crucial en la confusión de las personas. Comprender estos patrones puede ser útil para desarrollar estrategias de intervención y mejorar la capacidad de concentración y claridad mental en situaciones donde la confusión es indeseable.

Perspectivas

Para obtener mejores resultados, el principal ajuste sería una base de datos más grande. Una base de datos construida con gran control experimental, y

que no solo se oriente a clases en línea abriría muchas puertas para estudiar la confusión. Otra opción sería una red neuronal más robusta, que pudiera adecuarse a secuencias tan complejas como la de las EEG. Empero, dicha situación requeriría un mayor rendimiento computacional.

Explorar este tipo de tópicos es realmente beneficioso para la humanidad. Por su parte la neurociencia contribuye garrafalmente al bienestar de salud de las personas, cómo a su plenitud (neurofeedback). Por su lado, la computación tiene mucho que ofrecer en esa tarea, los modelos y el conocimiento que se adquiere con dicha herramienta resulta en un sin fin de ventajas. Los investigadores pueden ahorrarse

mucho dinero y tiempo gracias a esto. Además de evitar poner en riesgo la salud de las personas, al utilizar (quizá) técnicas invasivas.

Cabe destacar que al estudiar la cognición y el cerebro con métodos computacionales, en especial con Machine Learning, resulta en una 'paradójica' retroalimentación: muchas de las técnicas que se utilizan en el aprendizaje automático están basadas en las interpretaciones y conocimientos que los humanos poseen de su cerebro (las redes neuronales, por ejemplo). Nuevos descubrimientos en neurociencias, inevitablemente repercutirán en la computación, y en efecto, nuevas técnicas computacionales podrían surgir.

Referencias

- [0] F. Ramos-Argüelles, G. Morales, S. Egozcue, R.M. Pabón, M.T. Alonso. (2009). *Basic techniques of electroencephalography: principles and clinical applications*. Anales Sis San Navarra vol.32 supl.3 Pamplona 2009.
https://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1137-66272009000600006
- [1] Haohan Wang, Yiwei Li, Xiaobo Hu, Yucong Yang, Zhu Meng, Kai Min Chang. (2013, July). *Using EEG to improve massive open online courses feedback interaction*. CEUR Workshop Proceedings, 1009. <https://www.cs.cmu.edu/~kkchang/paper/WangEtAl.2013.AIED.EEG-MOOC.pdf>
- [2] Synapsis. Labster Theory. <https://theory.labster.com/synapses-es/>
- [3] Biólogo de Cabecera. (2020, diciembre). *ENTIENDE Que es la SINAPSIS NEURONAL y el IMPULSO NERVIOSO [FÁCIL] Impulso Nervioso y Sinapsis*. Youtube. https://youtu.be/___pi00Cv69I
- [4] El impulso nervioso. BIOLOGÍA - GEOLOGÍA. https://biologia-geologia.com/BG3/73_el_impulso_nervioso.html
- [5] Electroencefalografía (EEG). Brainsigns. <https://www.brainsigns.com/es/science/s2/technologies/eeg>
- [6] Thompson, G., *How Can Correspondence-Based Distance Education be Improved?: A Survey of Attitudes of Students Who Are Not Well Disposed toward Correspondence Study*. The Journal of Distance Education, 1990. 5(1): p. 53-65.
- [7] Passthoughts? Brainwave-Based Passwords a Reality. abcnews. <https://abcnews.go.com/blogs/technology/2013/04/passthoughts-brainwave-based-passwords-a-reality>
- [8] Differences Between Bidirectional and Unidirectional LSTM. Baeldung. <https://www.baeldung.com/cs/bidirectional-vs-unidirectional-lstm>
- [9] Bidirectional LSTM (BiLSTM) Model. Gabormelli. http://www.gabormelli.com/RKB/Bidirectional_LSTM_%28BiLSTM%29_Model
- [10] Complete Guide To Bidirectional LSTM (With Python Codes). Analytics India Mag. <https://analyticsindiamag.com/complete-guide-to-bidirectional-lstm-with-python-codes/>
- [11] Sotaquirá, Miguel. (2021, Septiembre). Ventajas y limitaciones de las Redes Neuronales Recurrentes y LSTM. Codificando Bits. <https://www.codificandobits.com/curso/fundamentos-deep-learning-python/redes-recurrentes-13-ventajas-limitaciones/>