



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
FACULTAD DE ESTUDIOS SUPERIORES ZARAGOZA

CARRERA DE PSICOLOGÍA

**Las densidades de las capas mieloarquitectónicas
usadas como parámetro de inicialización para un
Perceptrón Multicapa**

TESIS

QUE PARA OBTENER EL TÍTULO DE:
LICENCIADO EN PSICOLOGÍA

Presenta: MAURICIO ALEJANDRO VENTURA RODRÍGUEZ

Directora de tesis:

Dra. Lilia Mestas Hernández

Asesores:

Dr. Francisco Abelardo Robles Aguirre

Dr. Eduardo Alejandro Escotto Córdova

Sinodales:

Dr. Robles Aguirre Bernardo Adrian

Mtro. Torres Agustin Ruben



CDMX

2004



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

ÍNDICE

Resumen.....	4
Introducción.....	5
Capítulo 1: Desarrollo neurogenético y estructura neuronal.....	7
1.1 Neuronas.....	7
1.2 Origen del sistema nervioso central.....	8
1.3 Sinapsis.....	14
1.4 Histogénesis y Mieloarquitectura.....	18
Capítulo 2: Teoría y topografía cerebral.....	24
2.1 Topología y topografía cerebral.....	24
2.2 Teoría del cerebro.....	27
2.3 Topología y topografía.....	29
Capítulo 3: Teoría de la información y la relación con la estructura de comunicación de las neuronas.....	32
3.1 Entropía e información.....	32
3.1.1 Cognición y teoría de la información.....	39
3.2 Codificación de la información.....	45
3.2.1 Velocidad sináptica, frecuencias y codificación.....	49
3.3 Negentropía.....	55
Capítulo 4: Cibernética, control y comunicación.....	58
4.4 Cibernética.....	58
4.5 Perceptrón.....	61
4.6 Redes neuronales profundas.....	66
Capítulo 5: Metodología.....	71
5.1 Planteamiento del problema.....	71
5.2 Procedimiento.....	74
Capítulo 6: Resultados.....	80
Capítulo 7: Discusión y Conclusiones.....	91
Referencias.....	92

Resumen

La relación entre la mieloarquitectura cerebral y las redes neuronales artificiales ha sido un área de interés, pero no ampliamente explorada. En este contexto, esta investigación se centró en usar las densidades de las capas mieloarquitectónicas como parámetro de inicialización en un Perceptrón Multicapa.

Se implementaron dos versiones de un Perceptrón Multicapa de seis capas ocultas. En el primero, la capa de entrada y salida se inicializaron con un valor estándar de 0.001 y en el segundo se usó el valor de 1, mientras que las capas intermedias en ambos casos se basaron en valores de densidades mieloarquitectónicas. La arquitectura del primer perceptrón multicapa consiste en capas ocultas de 100 nodos cada una, además de utilizar la función ReLU de activación para evitar saturación de gradientes, con una capa de salida empleando la función softmax y entropía cruzada como función de retropropagación. La segunda versión usó 500 nodos en cada capa oculta e incorporó el optimizador Adam, modificaciones en el inicializador mediante la multiplicación de una matriz de pesos He Uniforme por la densidad de mielina, y un proceso de regularización que combina las normas L1 y L2, adaptándolas según la densidad estimada de mielina de cada capa.

Los hallazgos indican que las densidades mieloarquitectónicas tienen una mayor eficiencia de la red comparado con una inicialización aleatoria. Además, se identificó una correlación entre la media de estos valores y la eficiencia de la red neuronal. También se encontró que el área visual primaria de Brodmann destacó en eficiencia y precisión de aprendizaje. También a través de un análisis de entropía y negentropía de las matrices de pesos, se obtuvieron nuevas perspectivas y enfoques sobre la organización y codificación de la información en la red, respaldando los efectos de los valores de inicialización basados en la mieloarquitectura.

Palabras clave: Perceptrón multicapa, densidades mieloarquitectónicas, valores de inicialización, red neuronal profunda, entropía, negentropía.

Introducción

La intersección entre neurociencia y tecnología ha ofrecido perspectivas revolucionarias en el estudio de la mente humana. Esta tesis explora cómo las densidades de las capas mieloarquitectónicas, que se refieren a la organización y distribución de la mielina en el cerebro, pueden servir como parámetro de inicialización para un Perceptrón Multicapa, una red neuronal artificial compuesta por múltiples capas interconectadas. A través de la unión de la neurociencia, la teoría de la información y la cibernética, se busca arrojar luz sobre la eficacia de las redes neuronales artificiales al incorporar detalles biológicamente relevantes en sus estructuras.

El *Capítulo uno* detalla el desarrollo neurogenético y la estructura neuronal. Se profundiza en las neuronas, el origen del sistema nervioso central, la sinapsis y la histogénesis. Se subraya la relevancia de la mielinización y de las células afines en el funcionamiento cerebral.

El *Capítulo dos* aborda la teoría y la topografía cerebral. Se presenta una visión general de cómo se procesa la información en el cerebro y cómo la topología y la topografía cerebral influyen en dicho procesamiento.

El *Capítulo tres* introduce la teoría de la información en el contexto de la neurociencia. Desde la entropía hasta la negentropía, se exploran los mecanismos subyacentes en la codificación de la información neuronal y cómo estos conceptos pueden aplicarse en el ámbito de la inteligencia artificial.

El *Capítulo cuatro* examina el mundo de la cibernética y las redes neuronales. Se brinda un recorrido desde los orígenes de la cibernética hasta la estructura y funcionamiento de las redes neuronales artificiales, incluyendo el Perceptrón y las redes neuronales profundas.

El *Capítulo cinco* desglosa la metodología utilizada en esta investigación, detallando el planteamiento del problema, las variables consideradas y el procedimiento seguido.

El *Capítulo seis* presenta los hallazgos obtenidos al experimentar con el perceptrón multicapa y sus valores de inicialización basados en las densidades mieloarquitectónicas. Se discuten los patrones observados y las correlaciones identificadas.

El *Capítulo siete* concluye con una discusión de los resultados obtenidos, las limitaciones encontradas y las perspectivas futuras en esta intersección entre neurociencia y cibernética. Se resalta la importancia de esta investigación para el campo de las redes neuronales artificiales.

Capítulo 1: Desarrollo neurogenético y estructura neuronal

1.1 Neuronas

El estudio de las neuronas, células fundamentales para la comunicación rápida y específica dentro del cuerpo, ha sido una tarea intrincada que ha evolucionado con el tiempo (Kristan Jr., 2016). A pesar de la existencia de células que presentan características neuronales, como las células epiteliales de ciertas medusas y bacterias, la definición contemporánea se centra en la capacidad de una neurona de transmitir información en respuesta a estímulos (Nicholls et al., 2001). Esta comunicación neuronal, en contraposición a la comunicación hormonal, que se basa en la liberación de moléculas que viajan a través del torrente sanguíneo y actúan sobre células diana a distancia, tiene la ventaja de ser mucho más rápida, alcanzando velocidades que superan los 100 metros por segundo (Maynard & Downes, 2019).

La diversidad neuronal es extensa y a menudo simplificada en la literatura. Cada neurona, dependiendo de su localización y morfología, posee propiedades únicas de procesamiento de información. Esta variabilidad resulta en un sistema complejo, donde las redes neuronales no pueden ser simplemente reducidas a componentes individuales. Estas redes son intrincadas, con propiedades jerárquicas y densidades variables, y juegan roles esenciales en la modulación de la actividad neuronal. Esto lo podemos ver con los receptores neuronales, como los corpúsculos de Krause y Ruffini, ya que resaltan la adaptabilidad y especificidad de las neuronas para captar estímulos del ambiente (Ruffini, 1906; Krause, 1860). La diversidad de estos receptores y la variabilidad en los sistemas nerviosos a lo largo de diferentes especies recalcan la necesidad de entender la filogénesis, ontogénesis y sinaptogénesis para apreciar plenamente el papel de las neuronas. No obstante, la comprensión profunda de la estructura y funcionamiento de las neuronas provee parámetros cruciales para la inicialización y estructuración de redes neuronales artificiales, como el Perceptrón Multicapa. Al imitar la arquitectura neuronal en algoritmos computacionales, se busca capturar la eficiencia y precisión del sistema nervioso en soluciones tecnológicas.

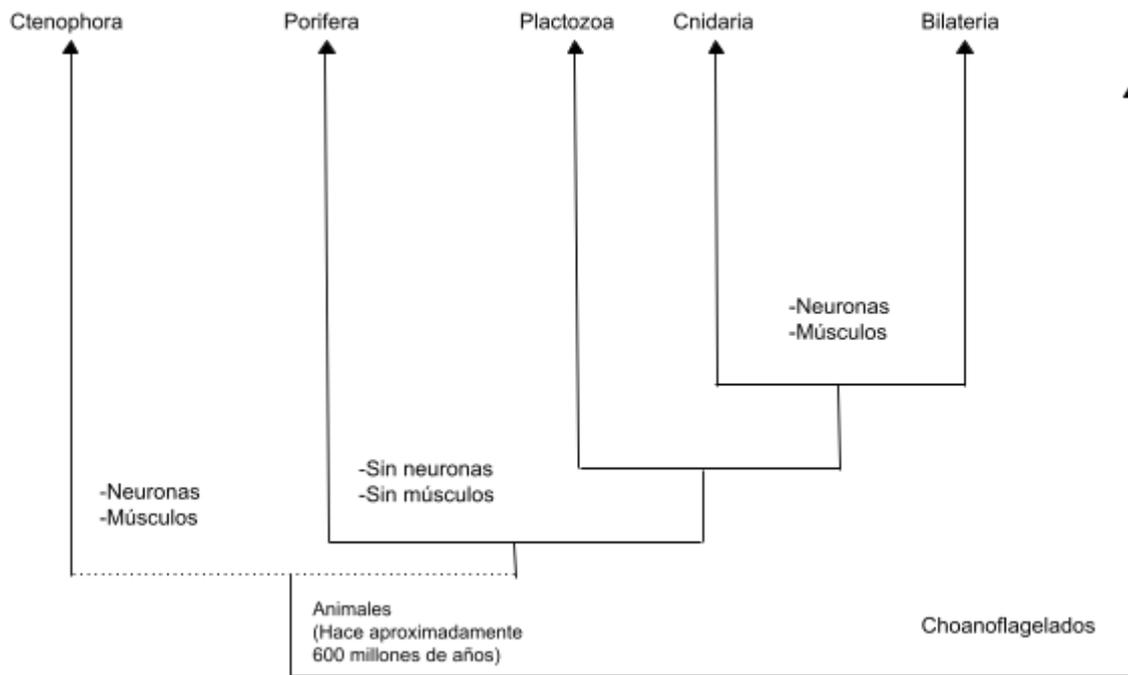
1.2 Origen del sistema nervioso central

El origen y evolución del sistema nervioso en seres humanos y otros animales ha sido un tema de debate y estudio desde hace mucho tiempo. La tarea de rastrear la evolución de nuestro sistema nervioso se complica debido a fenómenos como la evolución divergente y convergente. Estos procesos nos dejan con la incertidumbre de si desarrollamos de manera independiente mecanismos similares de comunicación celular nerviosa o si hubo una división evolutiva con desarrollo posterior de características distintas.

Comprender el origen de las neuronas en el reino *Animalia*, es crucial para elucidar la evolución y diversificación del sistema nervioso central permitiendo así establecer parámetros más precisos para modelos neuronales artificiales. El origen de las neuronas en el reino *Animalia* se encuentra inmerso en debates científicos, en gran parte debido a la multiplicidad que se observa en diferentes filos taxonómicos. Los cinco principales filos que representan las grandes divisiones del reino *Animalia* son *Ctenophora*, *Porifera*, *Placozoa*, *Cnidaria* y *Bilateria*; siendo este último el filo al cual pertenecen los cordados, incluidos los seres humanos (Moroz et al., 2014). Estos grupos evolucionaron hace más de 4 mil millones de años, mucho antes de que se conservaran fósiles animales identificables. No obstante, se han observado particularidades notables en la distribución de ciertos tipos de neuronas entre estos filos. Por ejemplo, las neuronas motoras están presentes en los filos *Ctenophora*, *Cnidaria* y *Bilateria*, pero notablemente ausentes en *Porifera* y *Placozoa*. Estas discrepancias evolutivas, combinadas con los desafíos de diferenciar entre *homología* (características compartidas por ascendencia común) y *homoplasia* (similitudes que no se derivan de un ancestro común), complican las reconstrucciones filogenéticas. Así, el esfuerzo de trazar el desarrollo y la diversificación de las neuronas en estos linajes es un reto que la ciencia está empezando a desentrañar con más precisión (Moroz et al., 2014; Kristan Jr, 2016).

Un punto de enfoque ha sido la diferencia en la disposición del eje corporal dorsoventral entre cordados (como los humanos) e insectos. A pesar de las aparentes diferencias morfológicas, la evidencia filogenética sugiere que ambos grupos podrían compartir un ancestro común (Arendt & Nübler-Jung, 1994; Denes et al., 2007).

Figura 1.
Relaciones entre los principales clados animales



Nota. Visualización de las relaciones entre los principales clados animales con coanoflagelados hermanos de todos los Metazoos.

En el reino Animalia, los humanos y otros animales del filo Bilateria se caracterizan por una simetría bilateral. Se han propuesto varias teorías sobre la evolución del sistema nervioso. Anton Dohrn en 1875 sugirió que los cordados heredaron su sistema nervioso central de un ancestro común, planteando una inversión en su disposición (Dohrn, A., 1875; Denes et al., 2007), teoría que fue actualizada por Nübler-Jung y Arendt en 1994. Por su parte, Bateson en 1886 comparó a los enteropneustos con los vertebrados, hipótesis también reconsiderada por Nübler-Jung y Arendt (Holland, 2016). Sin embargo, estudios recientes desafían la idea de un origen único del sistema nervioso, sugiriendo que este podría no ser monofilético y que el ancestro de todos los animales podría haber tenido un sistema nervioso que se perdió o evolucionó de manera diferente (Liebeskind et al., 2016; Moroz et al., 2014; Moroz, 2015; Moroz & Kohn, 2016; Jacobs et al., 2007).

El origen exacto del sistema nervioso es incierto, pero se conoce que los *Urbilateria* destinaron tejidos para formar un sistema nervioso central, evolucionando a las neuronas

actuales (Hartenstein & Stollewerk, 2015). La formación de la sinapsis es crucial, con componentes identificados incluso en organismos sin sistema nervioso como las esponjas y placozoos, lo que indica una base genética temprana para sistemas nerviosos más complejos, como por ejemplo el caso de la esponja *Amphimedon queenslandica* contiene elementos genéticos, y diversos compuestos de la densidad postsináptica (la PSD es una proteína especializada, unida a la membrana postsináptica) esencial para el funcionamiento de sinapsis en el filo Bilateria (Richards, et al., 2008; Sakarya, et al., 2007; Srivastava et al., 2008; Srivastava et al., 2010; Elliott & Leys, 2010; Burkhardt et al., 2014; Burkhardt, 2015; Arendt et al., 2016). En vertebrados, hay variaciones en el sistema nervioso. Los *Anamniotas*, por ejemplo, tienen órganos mecanorreceptores y detectores de campos eléctricos (Northcutt, 1984). Además, rutas neuronales como las espino-reticulares son comunes en vertebrados, mientras que otras varían entre especies, proporcionando pistas sobre la evolución y diferencias filogenéticas (Northcutt, 1984; Ronan & Northcutt, 1990).

La evolución de las células neuronales refleja adaptaciones cruciales para la supervivencia, con cada filo desarrollando mecanismos de comunicación neuronal según sus necesidades (Burkhardt et al., 2014; Burkhardt, 2015). Las teorías sobre la evolución del sistema nervioso ofrecen perspectivas valiosas, incluyendo la encefalización, que postula una evolución hacia una mayor capacidad de procesamiento de información y toma de decisiones (Sarnat & Netsky, 1981; Northcutt, 1984). La alometría estudia las relaciones no lineales entre tamaño corporal, anatomía y comportamiento, transformando estas relaciones en lineales a través de logaritmos, útil para comparar cerebros entre especies de distintos tamaños (Cuvier, 1825; Pilbeam & Gould, 1974; Martin, 1983). En mamíferos, se sugiere que el aumento del tamaño cerebral podría haber ocurrido de manera independiente en diferentes linajes, conocido como homoplasia (Hofman, 1982; Northcutt, 1984).

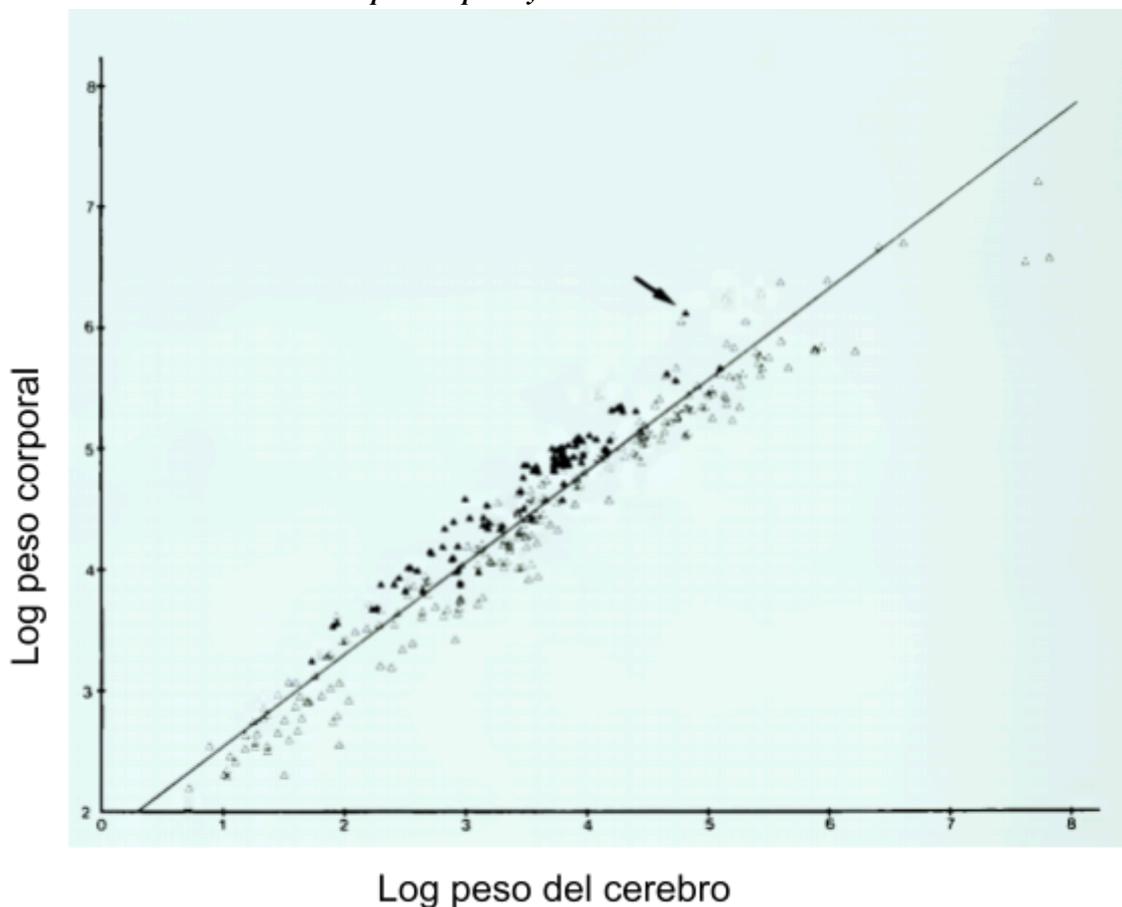
Existen otras dos hipótesis sobre cómo se han formado y adaptado las conexiones neuronales:

Hipótesis de la invasión: Propone que los axones (extensiones neuronales) desarrollan ramificaciones laterales, las cuales pueden formar nuevas conexiones en áreas no ocupadas previamente. Sin embargo, la evidencia sugiere que en la mayoría de los sistemas neuronales, las conexiones tienden a seguir rutas ya establecidas y

solo se pierden caminos mucho después debido a especializaciones (Herrick, 1948; Bishop, 1959).

Hipótesis de parcelación: Esta teoría sostiene que la evolución de los sistemas neuronales implica la pérdida selectiva de conexiones, lo que permite especializar funciones y potencialmente diversificar las células neuronales. En esta visión, las organizaciones neuronales primitivas serían más generales, y la especialización conllevaría la pérdida de ciertas conexiones (Ebbesson, 1984).

Figura 2.
Relaciones alométricas entre peso corporal y cerebral



Nota. Gráfica logarítmica del peso del cerebro (en miligramos) contra el peso corporal (en gramos) para una muestra de 309 especies mamíferas placentarias. Triángulos blancos = no primates; triángulos negros = primates; La flecha indica a los Homo Sapiens, siendo la especie más cercana un cetáceo (delfín) y no un primate. Modificado de Relative brain size and basal metabolic rate in terrestrial vertebrates. En Nature, vol. 293 no. 5827, (pp. 57-60). por Martin, R. 1981

Un gran problema con la hipótesis de parcelación es que presupone que los primeros vertebrados tenían un número limitado de tipos de células neuronales, pero que, al mismo tiempo, estas pocas células tenían que reflejar todos los caminos neuronales que vemos en los animales modernos. Tomando el cerebelo como ejemplo, sugiere que los primeros mamíferos tenían un cerebelo compuesto por una o dos tipos de células, en contraposición al cerebelo moderno con al menos ocho tipos diferentes (Northcutt, 1984).

Los seres humanos tienen una especialización neuronal que parece estar orientada al desarrollo de habilidades, tanto sociales como tecnológicas, como se evidencia en el desarrollo de la teoría de la mente y el uso distintivo del lenguaje (Bettle & Rosati, 2021). A diferencia de otros primates, los humanos muestran un grado superior de percepción de ilusiones ópticas debido a su procesamiento de información sensorial "holística" (Parrish et al., 2015). Al estudiar la evolución del sistema nervioso en homínidos, se ha observado que simplemente medir la capacidad craneal no es un buen indicador del comportamiento o las capacidades cognitivas. Lo realmente relevante es la organización y distribución neuronal (Holloway Jr, 1966; Tobias, 1971). La paleoneurobiología aborda esta cuestión estudiando estructuras craneales y conexiones corticales. Además, utiliza cocientes de encefalización para entender las diferencias y similitudes neurogénicas entre especies, incluyendo mamíferos más allá de los primates.

Aunque la morfología craneal ha sido útil para estudiar cambios comportamentales en muchas especies, tiene limitaciones, especialmente en primates, ya que no refleja adecuadamente funciones corticales complejas como memoria, percepción y lenguaje (Holloway Jr, 1966). La comparación directa de la neuroanatomía entre humanos y sus ancestros es imposible debido a la falta de cerebros preservados, y las comparaciones con otras especies son problemáticas debido a diferentes presiones evolutivas. La reorganización neuronal es crucial para entender la relación entre estructuras cerebrales y comportamientos y funciones corticales avanzadas. Los casos de *microcefalia vera* ilustran la importancia de esta reorganización. A pesar de tener cerebros con un número similar de neuronas funcionales al de gorilas o chimpancés, las personas con microcefalia pueden mostrar comportamientos exclusivos de humanos, que requieren de un procesamiento complejo de información (Holloway Jr, 1966).

Las neuroglías, además de las neuronas, juegan un papel crucial en el procesamiento de información en el sistema nervioso central. En animales más grandes como las ballenas, aunque hay una reducción en la densidad neuronal, el número de neuroglías aumenta proporcionalmente al tamaño del animal (Hawkins & Olszewski, 1957). En cuanto al cerebro humano, la clave de nuestras habilidades y comportamientos complejos no radica simplemente en más neuronas, sino en la reorganización y interacción de estas estructuras neuronales. Por ende, las funciones corticales superiores no dependen directamente del tamaño o densidad del cerebro, como se ha propuesto en investigaciones anteriores (Holloway Jr, 1966).

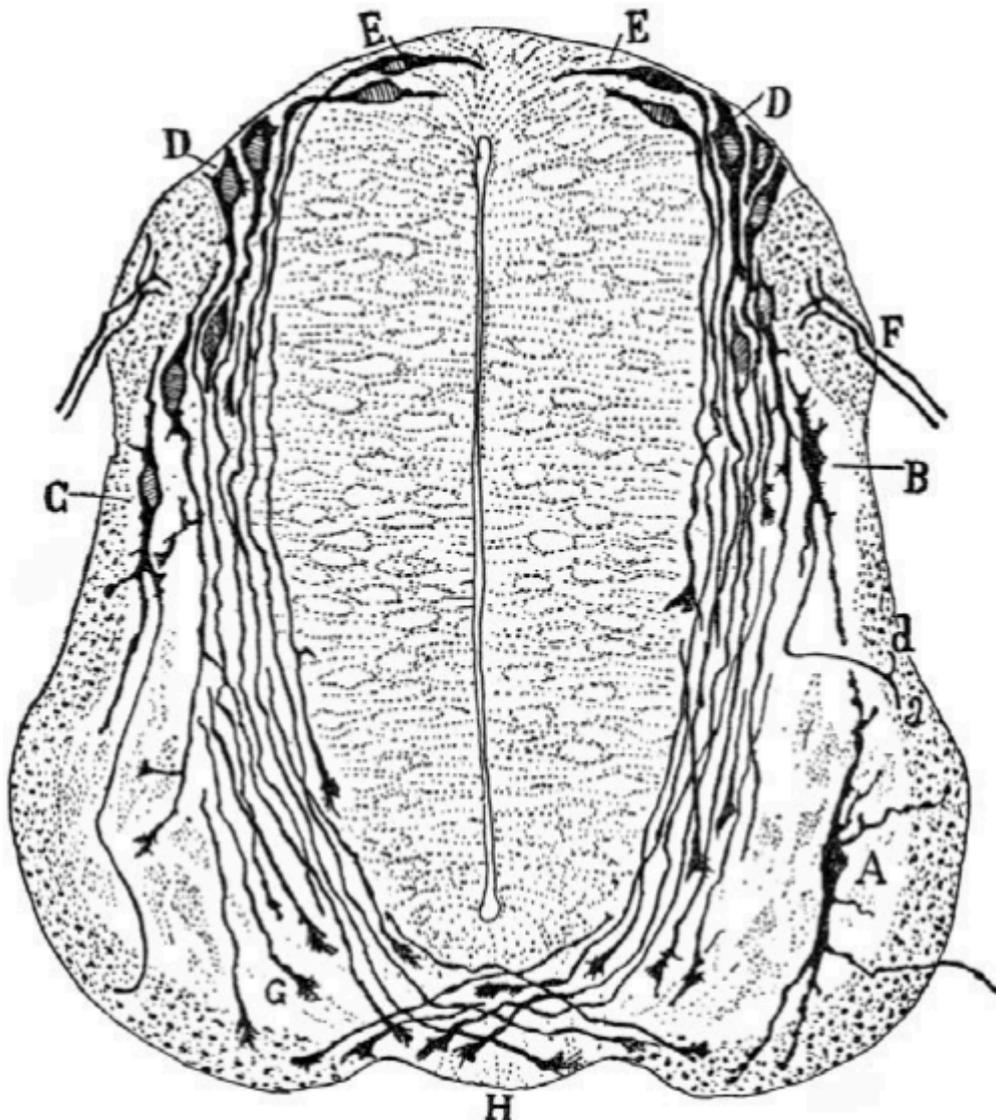
1.3 Sinapsis

Sherrington introdujo el término "sinapsis" en 1952, describiendo la interacción neuronal en el sistema nervioso, quizás influenciado por los estudios de Pavlov sobre glándulas digestivas (Granit, 1982). Destacó que la comunicación neuronal se da a través de propiedades conductivas tanto dentro de la célula como entre células, con un espacio intermedio para evitar la mezcla de señales. A pesar de que Pavlov rechazó inicialmente su idea de inhibición sináptica, Sir Charles Scott Sherrington y otros contribuyeron posteriormente al entendimiento de cómo la estimulación de nervios aferentes afecta los reflejos (Liddell & Sherrington, 1925; Eccles & Sherrington, 1931; Pearce, 2004). Por ejemplo sabemos que las funciones integradoras de los movimientos posturales, se encuentran esparcidos a lo largo del sistema nervioso central, dependientes de la tarea y el contexto en el cual se encuentran (Stuart, 2005). Los axones de vertebrados pueden ser mielinizados o amielínicos, con los mielinizados transmitiendo impulsos más rápidamente gracias a la mielina de células como los oligodendrocitos y las células de Schwann (Pfrieger & Barres, 1997; Frostick et al., 1998; Pfrieger, 2009). En la mayoría de las sinapsis, la transmisión es química, involucrando la liberación de neurotransmisores que desencadenan respuestas excitatorias o inhibitorias (Patton & Burgess, 2005). La acetilcolina fue uno de los primeros neurotransmisores descubiertos, implicado en la transmisión motora, aunque las primeras investigaciones se vieron afectadas por sustancias como el éter (Loewi, 1921; Brown et al., 1936; Dale et al., 1936). Las sinapsis rápidas se distinguen por la proximidad de sus receptores y el uso de neurotransmisores que actúan en canales iónicos, facilitando una comunicación neuronal eficiente y precisa (Sakmann et al., 1983; Patton & Burgess, 2005). Estas sinapsis permiten un control fino de la señalización gracias a la liberación y recepción de neurotransmisores, y la estructura sináptica influye en la codificación de la información (Zucker et al., 2014). La regulación de las conexiones sinápticas y la sinaptogénesis son procesos clave en el desarrollo neuronal, con factores tanto regresivos como progresivos que adaptan las conexiones a las demandas del entorno (Levi-Montalcini & Levi, 1944; Hamburger & Levi-Montalcini, 1949; Cowan et al., 1984; Chitnis & Kuwada, 1990). Además, el agrupamiento neuronal indica una especialización funcional, y las proteínas como la BMP-7 son relevantes para el desarrollo de dendritas (Lein et al., 1995). La navegación del cono

axonal depende de su habilidad para discernir señales moleculares y físicas, lo cual permite superar obstáculos y evitar moléculas inhibitorias (Letourneau, 1975; Patterson, 1988).

Figura 3.

El primer dibujo publicado de conos de crecimiento.



Nota. En esta sección transversal de la médula espinal de un embrión de pollo, en el día 4 de incubación, con el método de Golgi, revela la estructura de las neuronas individuales, desde los cuerpos celulares A-E hasta los conos de crecimiento G. Muchos de los conos de crecimiento se acercan o cruzan la línea media ventral H. Modificado de *A quelle époque apparaissent les expansions des cellules nerveuses de la moelle épinière du poulet* [¿Cuándo aparecen las expansiones de células nerviosas en la médula espinal del pollo?]. En *Anatomischer Anzeiger*, vol. 5, (pp. 609-613). por Ramón y Cajal, S. , 1890.

Los filopodios son cruciales para la movilidad y la dirección del crecimiento axonal, y su función puede ser modulada a través de la polimerización y despolimerización de microfilamentos (Gundersen & Barrett, 1980; Chien et al., 1993; Chien, 2005).

El crecimiento axonal consta del tubo axonal y del cono axonal, con sus componentes internos llamados axoplasma y cinetoplasma, respectivamente. El axoplasma se centra en el transporte, dominado por microtúbulos, mientras que el cinetoplasma prioriza la movilidad, siendo los filamentos de actina su principal componente (Mitchison & Kirschner, 1988). Los axones emplean señaladores de guía de crecimiento con diferentes efectos, encontrados en ciertas tablas. Estas señales pueden prolongar o terminar el contacto con células, dependiendo de su tipo. Las señales secretadas se clasifican en tres: moléculas pequeñas, componentes principales de la matriz extracelular (MEC), y proteínas que interactúan con la MEC sin formar parte de ella (Chien, 2005). Sin embargo, aún hay incertidumbres sobre cómo estas moléculas guían al cono axonal (Chien, 2005).

Tabla 1.
Actividad de las señales guía del axón

	Superficie celular	Secretadas
Positivo	<i>Adhesivo</i> Cadherina NCAM	<i>Permisivo</i> Laminina Fibronectina
		<i>Atractivo</i> Netrina Semaforina BDNF
Negativo	<i>Repulsivo</i> Efrina Semaforina Nogo MAG	<i>Inhibitorio</i> CSPG <i>Repulsivo</i> Netrina Slit Semaforina
	Modulador	HSPG Laminina Slit

Nota. Estos son ejemplos específicos de este tipo de clases de moléculas, estas pudiendo ser de tipo superficial de la célula o secretadas. Modificado de *Guidance of Axons and Dendrites*. En *Developmental Neurobiology* (pp. 241-267). por Chien, C. B., 2005,

El crecimiento axonal no es un proceso estático. Las neuronas tienen dinámicas complejas que les permiten no solo transmitir señales eléctricas sino también cambiar activamente su posición, adaptándose y comunicándose con su entorno, lo que potencia su eficacia frente a un modelo estático. Esto quiere decir que las redes neuronales dependen de un balance entre las señales excitatorias e inhibitorias, de las cuales sabemos las proteínas transmembrana como la *neuroligina en sus tres distintas isoformas* funcionan como promotoras de la diferenciación postsináptica, controlando la función y balance de las sinapsis excitatorias e inhibitorias en neuronas hipocámpales a través de las interacciones de su receptor axonal, la α -neurexina, y de estos ejemplos existen muchos de las formas regulatorias de las conexiones neuronales y su delicado balance (Chih & Scheiffele, 2005).

1.4 Histogénesis y Mieloarquitectura

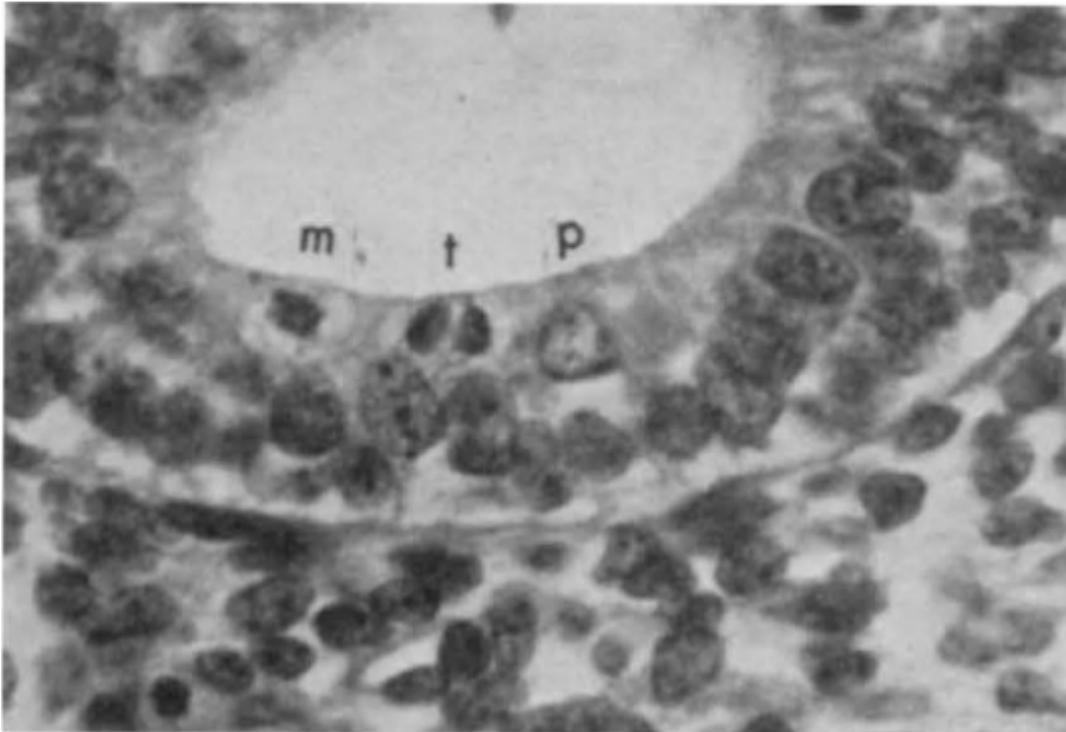
El proceso meticuloso de formación del sistema nervioso en mamíferos vertebrados destaca la importancia de la estructura y topografía neuronal, brindando pistas sobre su diseño y función (Fujita, 1960). Las neuronas, mediante su arquitectura, pueden converger o divergir señales sinápticas, esenciales para el procesamiento de información y regulación de funciones vitales.

El inicio de la histogénesis del sistema nervioso proviene de células madre multipotentes que adoptan roles específicos, guiadas por marcadores moleculares y caminos en desarrollo. Estas células se originan de una placa en el embrión durante la gastrulación, con la temporización crucial para determinar características como tamaño y forma del cerebro (Vetter & Dorsky, 2005; Hatakeyama et al., 2004). Las células del neuroectodermo evolucionan para volverse neurales. A medida que se dividen asimétricamente, pierden su habilidad de auto-renovarse y su potencial, controlando su proliferación. Finalmente, las células progenitoras se diferencian, rompen el ciclo celular estándar y expresan fenotipos específicos guiados por marcadores moleculares ambientales (Vetter & Dorsky, 2005). Es esencial que este proceso ocurra de manera ordenada y puntual para asegurar el desarrollo adecuado del sistema nervioso.

Dentro del tubo neural, la capa germinal genera células para la capa del manto, presentando un patrón mitótico distintivo (Fujita, 1960). Este orden es evidente en otros mamíferos, como el desarrollo neuronal en gatos, donde se observa la co-génesis de células en días específicos de gestación (Luskin & Shatz, 1985). También hay procesos de organogénesis que no dejan rastro en adultos, como sucede con las células gliales radiales en mamíferos (Luskin & Shatz, 1985). Investigar estas células transitorias en adultos resulta un desafío, ya que han culminado su desarrollo. Las células madre neurales poseen rutas migratorias específicas. Estos caminos pueden rastrearse utilizando tintes o empleando células donadoras de otras especies, aunque el tinte tiende a diluirse con el tiempo, complicando su seguimiento. Está claro que las redes neuronales se originan de una migración neuronal específica durante el desarrollo, guiada por marcadores moleculares que determinan su estructura y función final. Esta migración se divide en dos categorías: radial y tangencial, ambas presentes en la

migración de los oligodendrocitos. Específicamente, los predecesores de los oligodendrocitos, que provienen de dominios ventriculares específicos, se dispersan extensamente a lo largo del sistema nervioso central y periférico, jugando un papel crucial en la mielinización (Lachapelle et al., 1994).

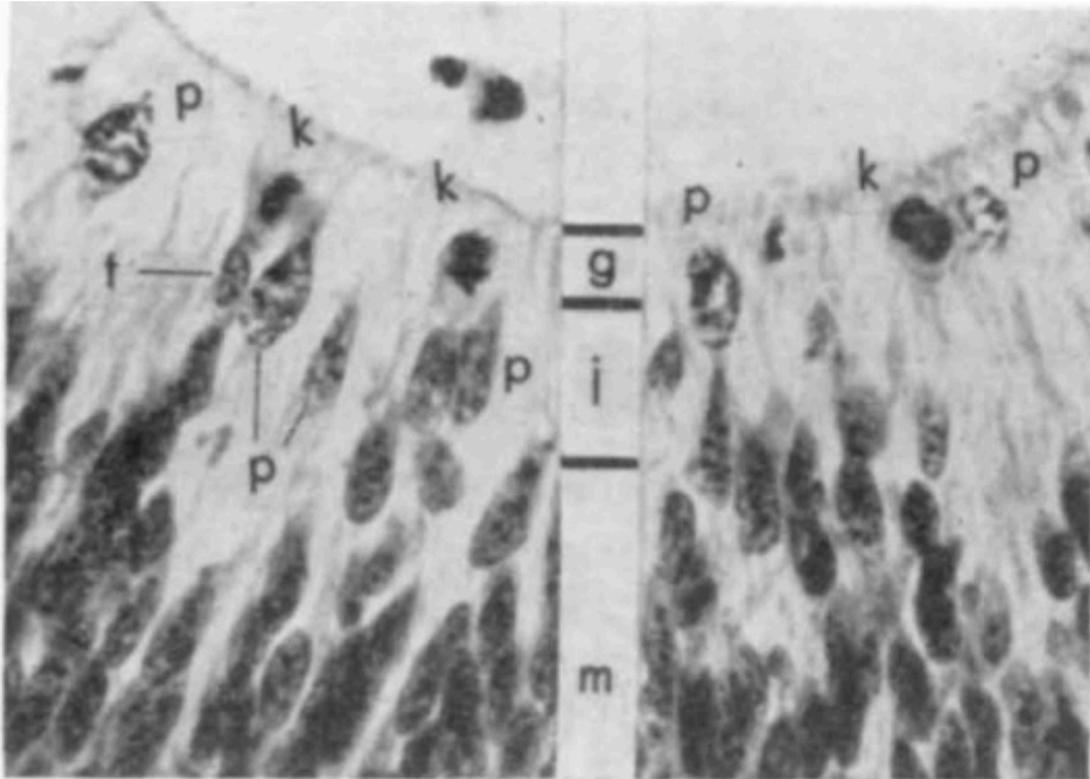
Figura 4.
Figuras mitóticas en el intestino



Nota. Figuras mitóticas en el intestino de un embrión humano siendo *p*, núcleo de profase; *m*, núcleo en metafase; *t*, núcleo telofase. Modificado de *Mitotic pattern and histogenesis of the central nervous system*. Nature, vol. 185 no. 4714, (pp. 702-703). por Fujita, S., 1960

Los oligodendrocitos se originan durante la embriogénesis en el tubo neuronal y en fases postnatales tempranas desde regiones germinales periventriculares. Estas células, derivadas de células progenitoras restringidas a la glía, son esenciales para la formación de mielina (Warf et al., 1991; Vallstedt, 2004; Menn et al., 2006). A pesar de su origen temprano, en cerebros adultos, los oligodendrocitos pueden regenerarse desde células madre en la zona subventricular para reparar mielina, especialmente tras daños como la desmielinización en ratones (Menn et al., 2006) y en humanos con afecciones como la epilepsia (Armstrong et al., 1992).

Figura 5.
La pared del tubo neural del embrión humano



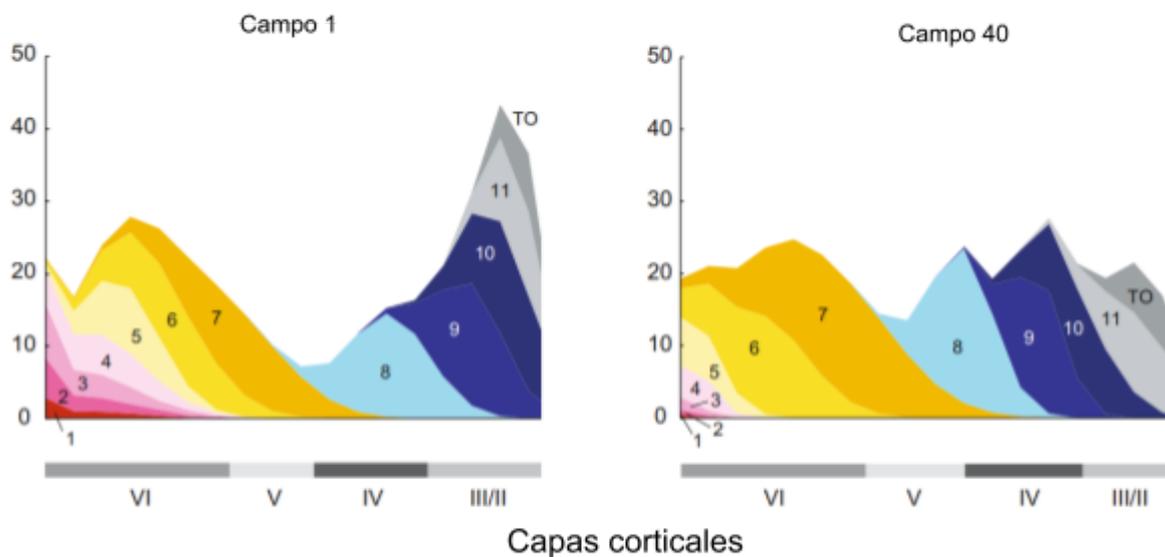
Nota. La pared del tubo neural del embrión humano donde *g*, epitelio germinal; *i*, zona intermedia; *m*: capa de manto; *k*, la llamada célula germinal que muestra núcleo mitótico; *p*, núcleo de profase; *t*, núcleo de telofase tardía. Modificado de *Mitotic pattern and histogenesis of the central nervous system*. Nature, vol. 185 no. 4714, (pp. 702-703). por Fujita, S., 1960

La actividad eléctrica cerebral está vinculada con la mielinización, influenciando la mieloarquitectura y las redes electrofisiológicas (Hunt et al., 2016). La presencia de mielina está correlacionada con la regulación de la plasticidad, aumento en la velocidad de conducción axonal y la reducción de la formación de sinapsis (McGee & Strittmatter, 2003; Glasser et al., 2014). Además, la maduración cortical y la mielinización en humanos están relacionadas, con patrones distintivos de mielinización en infantes (Deoni et al., 2015). Aunque la mielinización es prominente en la materia blanca, también ocurre en la materia gris con variaciones en su estructura (Nieuwenhuys, 2013).

El córtex cerebral inicia con células neuroepiteliales que proliferan y se dividen asimétricamente, generando neuronas y glía (Angevine & Sidman, 1961). La secuencia y posición de las neuronas en el córtex son dictadas por su tiempo de nacimiento, formando capas distintas, como el pre-plato y el plato cortical, cada uno con funciones específicas y

conexiones. Esta estructura laminar es fundamental para la funcionalidad del neocórtex y su alteración puede conducir a trastornos como la epilepsia (Vetter & Dorsky, 2005). Por lo que el Epitelio Ventricular Pseudoestratificado (EVP) resulta en una población diversa de células en el cerebro maduro, con un proceso de laminación que se despliega en múltiples ciclos, como se ha observado en ratones (Nowakowski & Hayes, 2005).

Figura 6.
Distribución laminar de neuronas durante Neurogénesis



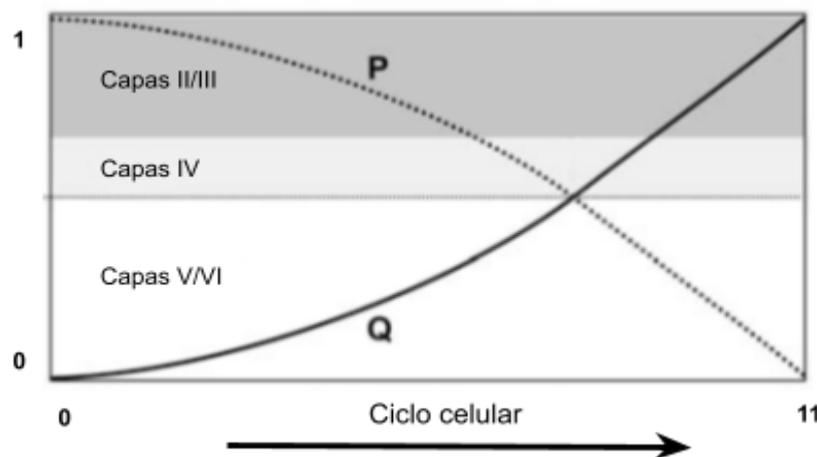
Nota. Gráfico de la distribución laminar de los 11 ciclos celulares de la Neurogénesis en un ratón. Donde el campo 1 está situado dorsomedialmente, y el campo 40 rostrolateralmente. Donde cada capa recibe neuronas que se producen durante más de un ciclo celular mostrando un claro cambio sistemático en las láminas de residencia con cada ciclo celular secuencial. Modificado de *Cell proliferation in the developing mammalian brain*. En *Developmental Neurobiology* (pp. 21-39). por Nowakowski & Hayes., 2005

El EVP regula la producción de neuronas mediante distintos tipos de divisiones mitóticas. Estas son:

- División celular terminal simétrica: genera dos células que se convierten en neuronas fuera del EVP.
- División celular simétrica no terminal: produce dos células que permanecen en el EVP, continuando su proliferación.
- División celular asimétrica: origina una célula que se mantiene proliferando en el EVP y otra que migra fuera de él.

En relación con las divisiones mitóticas, las células neuronales hijas se categorizan en dos tipos: las "Q", que son quiescentes y han salido del ciclo celular, y las "P", que siguen en el ciclo y son proliferativas. Independientemente de si deciden ser "Q" o "P", la suma total de ambas poblaciones debe permanecer constante ($Q + P = 1$). Se espera que al comienzo y al final del intervalo neurogénico, el total de estas células sea cero. La zona subventricular fue inicialmente descrita en 1905 por Schaper y Cohen. Sin embargo, dada la limitada proliferación y movilidad de sus células, se cree que esta zona podría ser una adaptación evolutiva reciente.

Figura 7.
Producción de neuronas Q y P a lo largo de la neurogénesis

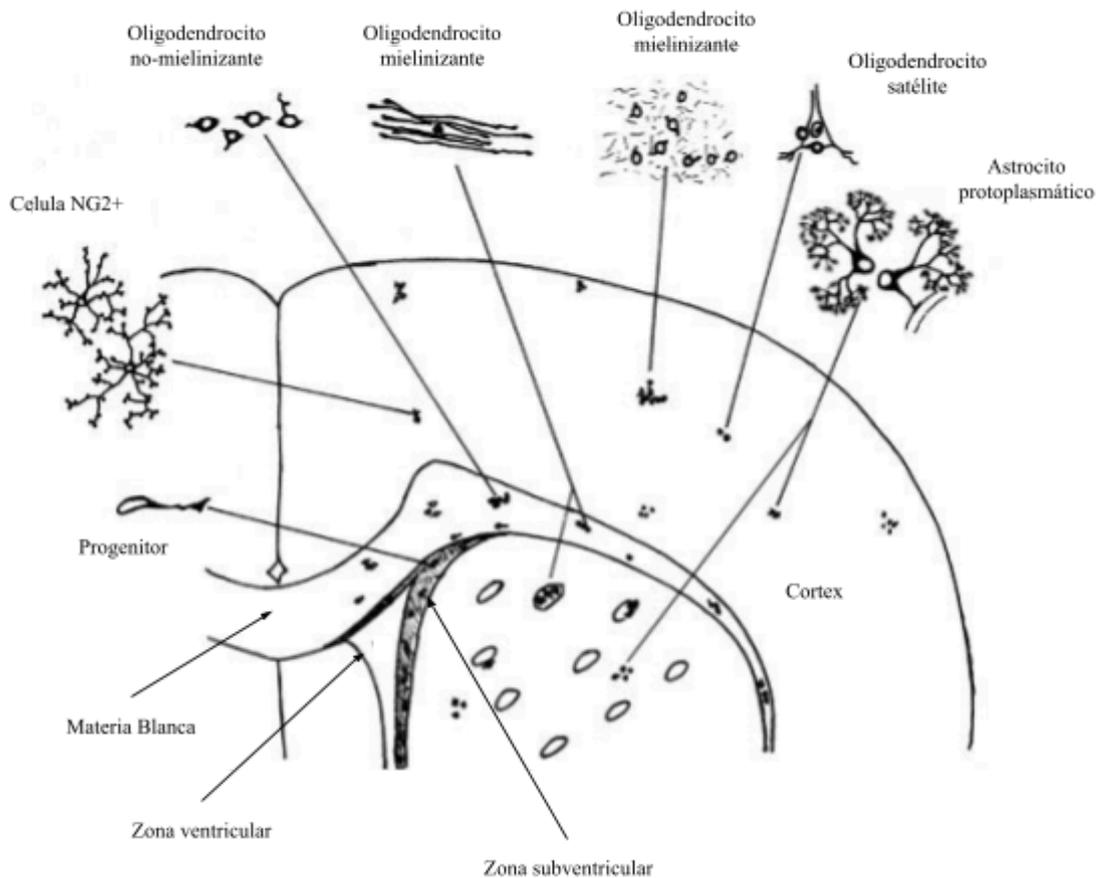


Nota. Gráfica del crecimiento y decrecimiento poblacional de las células Q y P a través de la neurogénesis en la formación de las capas corticales. Modificado de Cell proliferation in the developing mammalian brain. En *Developmental Neurobiology* (pp. 21-39). por Nowakowski & Hayes.

La zona subventricular contiene células pluripotentes capaces de producir oligodendrocitos, especialmente después de daños. Esta zona no es uniforme en todo el sistema nervioso central; en cambio, existen varias subdivisiones, cada una con células específicas: Zona subventricular anterior, dorsolateral, de eminencias ganglionares, neocortical y septal (Brazel et al., 2003). Estas divisiones indican especializaciones según las necesidades de las áreas vecinas. Aunque no se encuentra en todo el córtex, la zona subventricular está íntimamente ligada con la zona ventricular (Rakic & Sidman, 1968; Bystron et al., 2008). En función de la región y el momento, esta zona genera distintas neuronas y células gliales. Por ejemplo, la zona subventricular neocortical, que

coincide temporalmente con la producción neuronal para el neocórtex, es probable que también produzca neuronas y glías para esta área (Takahashi et al., 1995).

Figura 8.
Los descendientes de la zona subventricular perinatal dorso lateral.



Nota. Distribución de las células descendientes del área subventricular perinatal dorso lateral. Modificado de Brazel. En *Progress in neurobiology*, vol. 69 no. 1 (pp. 49-69). por Brazel, Romanko, Rothstein & Levison.

Las distintas poblaciones proliferativas tienen funciones esenciales en el desarrollo cerebral, encargándose de producir el número adecuado y tipo correcto de células (como neuronas y glías, junto con sus subtipos) para cada región. Sin esta precisa regulación, podrían surgir anomalías cerebrales.

Capítulo 2: Teoría y topografía cerebral

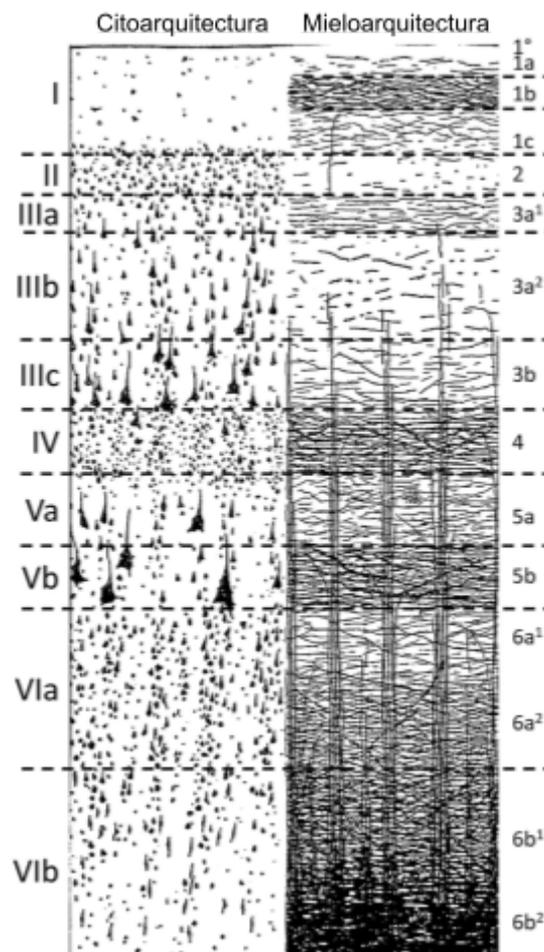
2.1 Topología y topografía cerebral

Las estructuras neuronales son críticas para las funciones corticales, y su daño puede tener efectos variados y a veces sorprendentes en la evolución futura de las funciones neuronales. La topología y la topografía son esenciales en el estudio de estas estructuras, influyendo en el diseño de modelos de redes neuronales como el Neocognitron, que refleja las jerarquías del procesamiento visual en los gatos (Fukushima, 1988). Estos modelos y la observación de estructuras reales sugieren una relación bidireccional entre las redes neuronales y el sistema nervioso. Científicos históricos han aportado conocimientos esenciales tales como Ramón y Cajal quien exploró la complejidad de las conexiones neuronales, mientras que otros como Huxley, Hodgkin, y McCulloch proporcionaron modelos matemáticos de estas conexiones. La teoría de códigos neuronales, argumentada por Curto en 2017, indica que las respuestas a estímulos pueden ser más críticas que las estructuras internas del cerebro, aunque la realidad sugiere una complejidad aún mayor. Hubel y Weisel (1959) descubrieron que existían neuronas con una orientación a la cual responden de forma específica, y después O'Keefe y Dostrovsky (1971) descubrieron la existencia de neuronas en el hipocampo que responden a la locación del ser vivo, por medio de las cuales se podía inferir a partir de la activación neuronal que se encontraba de ciertos grupos de neuronas en el hipocampo, demostrando la complejidad de las interacciones con las cuales se codifica información (Brown, et al, 1998).

En la historia de la neurociencia, Oskar y Cécile Vogt desarrollaron los estudios citoarquitectónicos y mieloarquitectónicos, con Korbinian Brodmann especializándose en la citoarquitectura cerebral, definiendo áreas específicas en el córtex humano a través de la tinción de Nissl. Brodmann estuvo aprendiendo de los Vogt durante el periodo de 1901 a 1909, en el cual se dedicó al estudio de las estructuras celulares del córtex de diferentes tipos de mamíferos, entre ellos algunos tipos de primates inferiores, mostrando su gran especialización en el estudio de la citoarquitectura. Estos estudios culminaron con la publicación de una primera monografía donde describen las diferentes áreas del cerebro humano (Brodmann, 1906). La totalidad de la funcionalidad del cerebro se puede diferenciar

a partir de las estructuras topológicas y topográficas neuronales, ya sea dividiendo por sectores tal como lo ha hecho Brodmann con 43 áreas en su monografía publicada en 1909, de las cuales terminarían siendo divididas y subdivididas para dar con las actuales 52 áreas de las cuales el área número 15 solo es encontrada en los monos de guinea ya que una estructura similar u homóloga no encontró para ese tiempo en el ser humano, y describiendo las funciones aún en la actualidad, o de manera más generalizada (Brodmann, 1909). Esta representación fue finalmente vista por última vez en su libro publicado en 1914, *Physiologie des gehirns*, con las últimas actualizaciones y cambios que hizo a su modelo (Brodmann, 1914), en el cual se puede observar una diferenciación de las estructuras, en este caso la citoarquitectura cerebral, haciendo uso del soma como una herramienta de diferenciación de un área a otra.

Figura 9.
Citoarquitectura y mieloarquitectura.



Nota. Representación gráfica de la citoarquitectura y la mieloarquitectura, con sus respectivas divisiones de capas corticales. Modificado de *Cortical layers: Cyto-, myelo-, receptor- and synaptic architecture in human cortical areas*. (pp. 716-741). por Palomero-Gallagher, N., & Zilles, K., 2019

A su vez, Von Economo y Koskinas (1925) detallaron áreas citoarquitectónicas incluso antes que Brodmann, aunque sus descripciones eran similares. A pesar de los desafíos técnicos de su tiempo, identificaron estructuras ligadas a funciones específicas. Sin embargo, estudios de lesiones cerebrales post-guerras mundiales cuestionaron el localizacionismo estricto que estas áreas sugerían.

Los Vogt tomaron un camino diferente, mapearon extensas redes de conexiones mielinizadas en el neocórtex usando la técnica de Weigert, clasificadas en orientaciones tangenciales y radiales, proporcionando una comprensión más rica de la organización cerebral (Vogt & Vogt, 1903). El mapeo citoarquitectónico busca categorizar el cerebro topográficamente, pero presenta limitaciones al tratar de visualizar aspectos como la mielina (Nieuwenhuys, 2013). Las distinciones citoarquitectónicas, como las áreas 44 y 45 de Brodmann, varían entre individuos. Sin embargo, hay evidencia de una lateralidad dominante relacionada con el lenguaje en el área 44, la cual muestra diferencias significativas en la densidad celular entre sujetos (Amunts et al., 1999).

Con la tecnología de imagenología moderna, reconocemos diferencias clave en la mieloarquitectura y la organización espacial de las fibras mielinizadas en el córtex (Shams et al., 2019). Los avances en neuroimagen sugieren que hay más áreas citoarquitectónicas que las 43 descritas por Brodmann, estimando hasta 200 áreas (Zilles & Amunts, 2010; Amunts & Zilles, 2015). Además, mientras que la topografía neuronal en el isocórtex típicamente presenta seis capas, ciertas áreas como el córtex entorrinal tienen más capas, y otras, como el hipocampo, tienen menos añadiendo aún más grados de complejidad a la distribución de la mieloarquitectura (Amunts & Zilles, 2015).

2.2 Teoría del cerebro

La teoría del cerebro explora cómo las estructuras cerebrales facilitan funciones complejas, extendiendo los conceptos de computación para modelar esta complejidad. La neurociencia computacional aplica análisis y simulación matemáticos para refinar los modelos cerebrales, basándose en el modelado neuronal previo y en la teoría de control biológico. En inteligencia artificial, se estudia la programación de comportamiento inteligente en computadoras, mientras que la robótica se enfoca en la gestión flexible de máquinas con capacidad de interactuar con su entorno. El cómputo neuronal, inspirado en la teoría cerebral, crea máquinas basadas en las redes neuronales, que se aplican tanto a sistemas biológicos como a sus equivalentes artificiales. Esta disciplina avanza tanto en IA como en robótica, e impulsa tecnologías de análisis de imagen y control de procesos. Las redes neuronales adaptables, capaces de autoorganizarse o modificarse sin programación explícita, son clave en este campo. Sin embargo, abordar problemas complejos a menudo implica integrar múltiples redes especializadas, lo que requiere un análisis y diseño detallado de estas redes interconectadas.

Por esto es que el estudio del cerebro es crucial para entender el procesamiento de la información. Ashby en 1954 vio el cerebro como un sistema cerrado influenciado por el entorno, y la computación ha buscado emular su procesamiento en máquinas. En la reunión de Trieste de 1984, se exploraron estas ideas (Palm & Aertsen, 2012).

Observar las neuronas, unidades fundamentales del cerebro, revela una red compleja que se adapta con la experiencia (Searle, 1980). La IA enfrenta el debate de si las máquinas pueden desarrollar procesos cognitivos comparables a los seres vivos, con Searle distinguiendo entre IA "fuerte" y "débil". Algunos, como McCarthy en 1979, sugieren que incluso máquinas simples pueden tener "creencias", pero atribuir cualidades mentales a máquinas es controvertido. Comprender la neurofisiología podría llevarnos a crear una "proto-mente" en máquinas. Los estudios de Hodgkin y Huxley en 1952 sobre la conductancia iónica abrieron la puerta a comprender cómo partes de la membrana celular podrían actuar como elementos

computacionales. Las neuronas artificiales de hoy, aunque simples, podrían evolucionar para incorporar estas complejidades y la neuromodulación.

Arbib en 1975 vio conexiones entre la teoría de autómatas y la teoría cerebral, y Von Neumann (2017) ya había reconocido los desafíos de entender el cerebro en 1951. Arbib abogó por atribuir "creencias" y "metas" a máquinas y programas para un entendimiento más eficiente, a pesar de los desafíos de definir estas atribuciones, por otro lado Von Neumann explicaba las problemáticas al intentar comprender y describir un sistema biológico tan complejo, y de la cual sabemos poco aún. Siendo la primera parte del problema la estructura y funcionamiento de estas unidades elementales, y la segunda en la axiomatización de estos elementos y su funcionalidad. La neurociencia computacional busca comprender cómo el cerebro procesa la información, pero se enfrenta a la dificultad de enlazar teorías cognitivas con neurofisiología debido a la abundancia de datos y la dificultad de correlacionarlos con mecanismos computacionales subyacentes (Kriegeskorte, 2018). Un ejemplo ilustrativo sería aplicar teorías cognitivas sobre módulos de funciones superiores para iniciar el mapeo cerebral usando técnicas de imágenes funcionales como electroencefalogramas, tomografías por emisión de positrones y resonancias magnéticas (Fodor, 1985). Sin embargo, este enfoque no está libre de sesgos cognitivos y, aunque proporciona un mapa de la activación cerebral, no aclara los mecanismos computacionales del cerebro. Además, persiste la incertidumbre sobre la existencia de módulos más diminutos. Esto era algo que había ya considerado Von Foerster (1968, 2002), el nos habla de la necesidad de tener un sistema dinámico de almacenaje de información, siendo los cambios constantes del mismo sistema, parte del mismo almacenaje de la información. Ya que la experiencia generada por la red neuronal crea cambios poco a poco, los cuales terminan apilandose, siendo esto una forma rústica de meta-almacenaje de información, sin tener la necesidad de involucrar otro sistema de almacenaje. La neurociencia computacional, en la actualidad, ha adoptado diferentes enfoques, desde el "top-bottom" al "bottom-up", descomponiendo procesos cognitivos complejos en sus componentes computacionales, y ensamblando interacciones neuronales simples hasta llegar a obtener funciones computacionales complejas (Dayan & Abbott, 2001; Ernst & Banks, 2002).

2.3 Topología y topografía

La topología, una rama de las matemáticas, estudia las propiedades de los cuerpos geométricos que no cambian ante deformaciones continuas, centrándose en funciones continuas y espacios topológicos. Al aplicar este concepto al sistema nervioso, se refiere a las características consistentes en el funcionamiento y procesamiento de información cerebral, independientemente de las variaciones individuales. El cerebro se organiza en diversas estructuras, y es común intentar asignar funciones específicas a áreas concretas. De hecho, ha sido atractivo y persistente en la neurociencia vincular funciones corticales a áreas específicas del encéfalo. Aunque a menudo observamos una relación entre daños en áreas cerebrales y déficits funcionales, la respuesta y recuperación ante lesiones varía entre individuos. Esta variabilidad se extiende a dominancias hemisféricas, donde, por ejemplo, la localización del lenguaje no siempre es en el hemisferio izquierdo. Un estudio destacado logró reubicar la dominancia del lenguaje en un paciente mediante estimulación transcraneal, previo a la eliminación de su hemisferio izquierdo (Barcia et al., 2012).

El sistema nervioso organiza la información en distintos niveles y rutas de procesamiento, como las vías neuronales, corticales y subcorticales. Esta complejidad multicapa, que abarca desde la percepción sensorial inicial hasta la activación cerebral global asociada al pensamiento y comportamiento, refuerza la idea de que el cerebro opera en su totalidad o mayoría constantemente. El procesamiento visual, aunque complejo, posee ciertas características topológicas en la forma en que las neuronas comunican la información. Un modelo que captura algunas de estas características es el Neocognitron, una red neuronal multicapa autoorganizada que supera a su predecesor, el Cognitron, al identificar patrones independientemente de su posición. Esta red, que inicialmente procesa información visual como sumas de matrices, luego se refina mediante retroalimentación, culminando en el área occipital del cerebro del gato (Fukushima, 1988). Estas características de procesamiento jerárquico y retroalimentación han sido adoptadas en las modernas redes neuronales de aprendizaje profundo, optimizando el procesamiento de información multidimensional, como en la tarea de reconocimiento de imágenes. Por ejemplo, podemos describir y entender las propiedades de un Perceptrón Multicapa usando la topología y la teoría de grafos. Para esto

debemos considerarlo como un grafo dirigido acíclico, donde los pesos en las aristas serán los pesos sinápticos en las conexiones neuronales. Tomando en cuenta esto, la salida de cada neurona en el PM se calcula como una función no lineal de la suma ponderada de sus entradas. Si denotamos la salida de la neurona i como y_i , la entrada a la neurona j como x_j , y la función de activación como f , entonces tenemos que

$$y_i = f \left(\sum_j w_{ij} x_j \right), \quad (2.3.1)$$

Desde el punto de vista topológico, cada capa del PM podría ser vista como una transformación del espacio de entrada en un nuevo espacio. Por lo tanto si denotamos la transformación realizada por la capa k como T_k , entonces la salida de la capa k se puede escribir como

$$y^k = T_k \left(y^{(k-1)} \right), \quad (2.3.2)$$

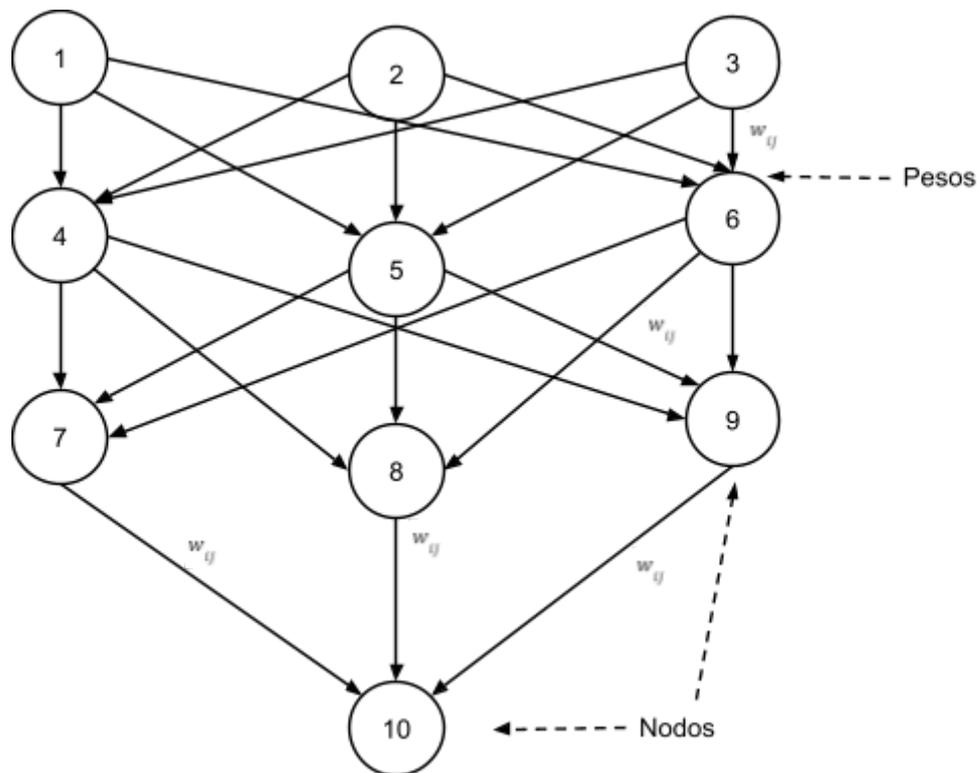
Donde $y^{(k-1)}$ es la salida de la capa $k - 1$, y T_k es una función que depende de los pesos sinápticos y la función de activación en la capa k . Vemos que también se puede entender la capacidad de un MLP para aproximar funciones en términos topológicos. Tomando en cuenta el teorema de aproximación universal (Hornik et al, 1989), el cual nos dice que un PM con una sola capa oculta puede aproximar cualquier función continua en un subconjunto compacto del espacio de entrada. Esto se puede expresar de manera compacta como

$$\forall f \in C(X), \exists g \in G : \|f - g\| < \epsilon, \quad (2.3.3)$$

Donde $C(X)$ es el conjunto de todas las funciones continuas en el espacio de entrada X , G es el conjunto de todas las funciones que pueden ser representadas por el PM, $\| \cdot \|$ es una norma (o diferencia en cierto sentido) existente entre f y g que es menor a un

número pequeño arbitrario ϵ . Esto quiere decir que para cualquier función continua f , podemos encontrar una función g representada por el PM que está ϵ - cerca de f .

Figura 10.
Grafo dirigido acíclico de un Perceptron Multicapa.



Nota. En este grafo representa a un PM con 10 nodos, los nodos representan las neuronas y las aristas las conexiones entre las neuronas. Los pesos en las aristas denotados por w_{ij} representan los pesos sinápticos en las conexiones neuronales. Aquí los nodos 1, 2, 3 representan las neuronas en la capa de entrada, los nodos 4, 5, 6, 7, 8, y 9 son las neuronas de las dos capas ocultas y el nodo 10 representa la neurona en la capa de salida.

Capítulo 3: Teoría de la información y la relación con la estructura de comunicación de las neuronas

3.1 Entropía e información

La teoría de la información, introducida por Claude Shannon, cambió la forma en que entendemos la información. Esta teoría usa la entropía discreta como la principal unidad de información. La entropía discreta representa las características actuales de un sistema, permitiéndonos comprender aspectos ocultos entre complicadas ecuaciones. Aunque el término "entropía" a menudo se asocia erróneamente con el desorden, su verdadera definición es más compleja y multifacética (Lebowitz, 1993). La comprensión original de la entropía proviene de la termodinámica, específicamente del análisis del proceso de Carnot realizado por Clausius en 1854. Aquí, la entropía se relaciona con la relación entre el trabajo y la temperatura en un sistema

$$dS = \delta Q_{rev}/T, \tag{3.1.1}$$

Donde S es la entropía, Q el trabajo ejercido dentro del sistema y T la temperatura. En la estadística mecánica, una rama que se centra en el comportamiento microscópico de los sistemas, la entropía se describe como

$$S = k \ln \Omega, \tag{3.1.2}$$

Donde S es la entropía total, k es la constante de Boltzmann (La constante de Boltzmann es el promedio de energía cinética encontrada en las partículas que conforman un gas) y Ω es el número de configuraciones microscópicas. la "información" se define en términos de reducción de la incertidumbre. Una manera técnica de verlo es a través de la entropía, un concepto clave en esta teoría. La entropía, denotada comúnmente como $H(x)$ para una fuente de información x , mide la cantidad de incertidumbre o sorpresa asociada con los posibles resultados de x . Matemáticamente, se expresa como

$$H(x) = \sum_{i=1}^n p_i(x_i)I(p_i), \quad (3.1.3)$$

Donde el valor de H es la *entropía o información* de Shannon, y p_i es la probabilidad de que ocurra el evento x_i (Proakis, Salehi, Zhou, & Li, 1994). En esta formulación es donde podemos ver que la teoría de Shannon se enfoca más en cómo medir y maximizar la transmisión de información a través de canales ruidosos, más que en las características específicas de las señales en sí, y podemos ver que estas diferentes formulaciones de entropía, aunque se originan en campos distintos, comparten similitudes estructurales y conceptuales, y todas intentan describir el estado actual de un sistema (Frigg y Werndl, 2011). La probabilidad es esencial para entender la entropía y la información. Mientras que su naturaleza exacta ha sido debatida (Jeffreys, 1998), en la teoría de la información, la probabilidad está vinculada con la incertidumbre y la información (Robert et al., 2009). Según Shannon (1949), más información lleva a menos incertidumbre y viceversa, y la entropía de Shannon cuantifica la información mínima necesaria para describir un sistema. Gracias a esta perspectiva, Shannon propuso una forma eficiente de codificar y transmitir mensajes. Esta idea llevó a la teoría de la codificación, que aborda cómo se puede codificar la información y los requerimientos para modelar o emular sistemas complejos en computadoras.

Los seres vivos, tienen centros de comunicación y control que intercambian información sobre el sistema. Esta comunicación, por ejemplo en el sistema nervioso, no está libre de interferencias o "ruido". Las comunicaciones son efectivas siempre que el ruido no supere la información enviada y el canal de comunicación tenga la capacidad suficiente. Además, hay factores externos, como el ambiente, o internos, como lesiones, que pueden afectar esta comunicación (Luria, 2011). Las relaciones en áreas específicas del cerebro, como la corteza occipital, temporal y parietal, enfrentan desafíos similares a los sistemas de comunicación artificiales, como la pérdida de energía a medida que la información viaja (Hodgkin, 1947).

Un aspecto interesante es la semejanza entre la comunicación neuronal y la comunicación intercontinental. Ambos sistemas tienen una resistencia eléctrica, una distancia que recorrer y un mensaje codificado. Una relación matemática para describir esto es

$$\lambda = \sqrt{r_m / r_t} \quad (3.1.4)$$

Siendo λ la constante de longitud, la cual habla de la distancia máxima que recorrerá el voltaje, r_m la resistencia del receptor en este caso la membrana, r_t es la resistencia del medio, esto es, del citosol que es el fluido intracelular (Falk & Fatt, 1964; Hodgkin & Rushton, 1946). Además, las neuronas utilizan umbrales eléctricos para filtrar señales débiles. Estos umbrales, junto con otras variables, determinan la eficacia de la comunicación neuronal. Otro aspecto crucial es la velocidad de cambio del potencial de la membrana

$$\tau = r_m c_m \quad (3.1.5)$$

Donde τ es la constante de tiempo, r_m es la resistencia de la membrana y c_m la capacitancia de la membrana (Nicholls et al., 2001). Integrando estas relaciones, obtenemos la ecuación cable, un modelo que describe la comunicación neuronal

$$\frac{1}{r_l} \frac{\partial V}{\partial t} = c_m \frac{\partial V}{\partial t} + \frac{V}{r_m} \quad (3.1.6)$$

Y si se retoman las dos ecuaciones anteriores que se describieron para el funcionamiento de las neuronas y sus neuritas y les incorporan, se obtiene finalmente la ecuación cable en su forma más conocida (Hodgkin y Rushton, 1946)

$$\lambda^2 \frac{\partial^2 V}{\partial x^2} = \tau \frac{\partial V}{\partial t} + V, \quad (3.1.4)$$

Sin embargo, es importante señalar que estas ecuaciones son generalizaciones y pueden variar según las características específicas de las neuronas. Aunque hay similitudes entre la comunicación neuronal y los sistemas artificiales, existen diferencias fundamentales, y las simplificaciones matemáticas en algunos casos omiten factores que podrían ser considerados menores. Comprender los límites de la capacidad de transmisión de información es esencial para reconocer la eficiencia de los canales de comunicación entre neuronas y, en última instancia, para la eficacia global de la comunicación en el sistema nervioso.

Retomando, para la teoría de la información de Shannon, el concepto de información, definido en términos de reducción de incertidumbre y cuantificado por la entropía, es más fundamental que el mero acto de enviar señales porque representa el contenido y el significado que se comunica. También sabemos que la comunicación eficiente entre neuronas es esencial para el funcionamiento óptimo del sistema nervioso, pero no lo es todo. Para comprender cuánta información puede ser transmitida de forma confiable entre neurona a neurona, es crucial examinar la capacidad máxima del canal de comunicación. Esta capacidad, denotada por C , se calcula utilizando la siguiente ecuación

$$\begin{aligned}
 C &= \max_{p(x)} I(X; Y) \\
 &= \max_{p(x)} [H(Y) - H(Y|X)],
 \end{aligned}
 \tag{3.1.7}$$

Siendo C la capacidad del canal de comunicación, y $I(x, y)$ es la información que comparten las dos variables. En otras palabras este teorema nos dice cuál es la máxima tasa de información C que puede ser enviada sobre un canal ruidoso con una cierta capacidad sin que la información sea distorsionada por el ruido en el límite cuando el tamaño del mensaje tiende al infinito (Cover y Thomas, 2012). La información mutua $I(X; Y)$ nos dice cuánto aprendemos sobre una cosa al conocer otra, siempre que ambas estén relacionadas (Cover y Thomas, 2012). $H(Y)$ es la entropía de salida, que es la cantidad de información de la variable Y , mientras que $H(Y|X)$ es la entropía condicional, que nos dice cuánto no sabemos de Y cuando ya conocemos X . Esta información mutua funciona igual en ambos sentidos, ya sea que aprendamos sobre X al ver Y o viceversa, y esto es muy importante para que la

comunicación funcione bien. Si X y Y son independientes, entonces $I(X; Y)$ es cero, pero si están totalmente conectadas, la información mutua es la mayor posible. Entender la comunicación también significa considerar cómo las interferencias externas pueden cambiar los mensajes que se envían.

Hartley, en 1928, creó una fórmula simple para entender cuánta información puede llevar un mensaje. La fórmula es

$$M = 1 + \frac{A}{\Delta V}, \quad (3.1.8)$$

Donde M es el número máximo de diferentes señales que podemos identificar claramente, A es el rango de señal en voltios, y ΔV es cuán bien el receptor puede distinguir la señal. Así que si subimos la fuerza de la señal, podemos enviar más información. Pero en la vida real, las interferencias pueden cambiar el mensaje que se envía. Hartley también nos dio otra fórmula

$$R = f_p \log_2(M), \quad (3.1.9)$$

Que nos dice qué tan rápido podemos enviar mensajes de forma clara, donde R es la velocidad de los mensajes y f_p es la frecuencia de los pulsos. Esto nos da la velocidad máxima para enviar mensajes sin que el ruido los afecte mucho. Shannon se dio cuenta de que el ruido, esas cosas no deseadas que se mezclan con lo que queremos comunicar, suele repartirse de una forma común llamada distribución Gaussiana (Stone, 2015). Para tener en cuenta este ruido, cambió la fórmula que ya tenía para que ahora incluyera el efecto del ruido en la comunicación

$$I(x, y) = H(y) - H(\eta), \quad (3.1.10)$$

$I(x, y)$ es la información útil que realmente obtienes, $H(y)$ es toda la información que se envía, y $H(\eta)$ es la interferencia o ruido que se mezcla con la información que se desea

enviar. En otras palabras, si restas la confusión de todo lo que se envió, obtienes la información que realmente puedes usar. Considerando esto podemos describir al ruido como

$$H(\eta) = \left(\frac{1}{2}\right)\log 2\pi e v_{\eta}. \quad (3.1.11)$$

Donde $2\pi e v_{\eta}$ describe las características gaussianas del ruido. Lo que esta ecuación hace es darte un número que te dice cuánta “*confusión*” está añadiendo el ruido al mensaje que quieres transmitir. Y para la información de salida, donde la suma del valor de salida $y = x + \eta$, con una varianza del valor $v_y = v_x + v_{\eta}$ se tiene que

$$H(y) = \left(\frac{1}{2}\right)\log \left(1 + \frac{v_x}{v_{\eta}}\right), \quad (3.1.12)$$

Donde el término y es el resultado de sumar la señal original x , el ruido η , y la expresión $\frac{v_x}{v_{\eta}}$ compara la variabilidad de la señal original con la del ruido. La entropía de la información aquí representada por $H(y)$ mide cuánta incertidumbre o “*sorpresa*” hay en la señal de salida que recibimos después de que ha sido alterada por el ruido. Por lo tanto si combinamos las ecuaciones tenemos que

$$I(x, y) = \frac{1}{2}\log \left(1 + \frac{v_x}{v_{\eta}}\right), \quad (3.1.13)$$

Y podemos ver que esta ecuación es una combinación de las dos anteriores, y representa la cantidad de *información útil*, o *información mutua*, $I(x, y)$ que se puede obtener de la señal de salida y cuando se conoce la señal de entrada x , incluso cuando hay ruido presente en el canal. Y es justo esta la ecuación de Shannon para un canal con propiedades Gaussianas. A partir de la fórmula anterior, Shannon pudo darnos la fórmula de la capacidad máxima para un canal de comunicación pero para mandar señales de telegramas (Pierce, 1980). Ya que tenemos que el poder de la señal es igual a la variabilidad de la misma medida en joules por

segundo, y el ruido en el canal es de naturaleza gaussiana e igual se mide en joules por segundo, podemos obtener la famosa fórmula de Shannon para la capacidad máxima de un canal con propiedades Gaussianas

$$C = \frac{1}{2} \log\left(1 + \frac{S}{N}\right), \quad (3.1.14)$$

Donde C es la capacidad del canal, S es la media de potencia de la señal, N es la media de la potencia del ruido, y $\frac{S}{N}$ representa la proporción de la señal a ruido. Esta ecuación se asemeja a la propuesta por Hartley, culminando en el teorema Shannon-Hartley:

$$C = B \log_2\left(1 + \frac{S}{N}\right), \quad (3.1.15)$$

Aquí se añade el valor de B , que es el ancho de banda en hertz. Entender estas formulaciones es crucial para comprender cómo se codifica y transmite la información, incluso a nivel neuronal.

Borst y Theunissen (1999) y Ince et al. (2010) realizaron estudios que demuestran la importancia de estas ecuaciones en la neurociencia. Finalmente, la relación entre energía y transmisión de mensajes es vital para entender el alto costo energético del cerebro humano. Se ha observado que el cerebro humano usa entre 10^5 y 10^8 veces más energía que el mínimo termodinámico requerido por bit transmitido, lo que refleja la complejidad y la importancia de los procesos cognitivos en nuestra especie (Howard, 1996; Lennie, 2003). Estos hallazgos resaltan por qué actividades cognitivamente demandantes pueden resultar agotadoras (Inzlicht et al., 2015; Inzlicht et al., 2018).

3.1.1 Cognición y teoría de la información

La capacidad de las neuronas para comunicarse es fundamental en la transmisión de información. Las herramientas matemáticas nos ofrecen la posibilidad de calcular rigurosamente esta capacidad, particularmente a través del uso de ecuaciones que se pueden aplicar a redes neuronales complejas. El punto esencial de este texto no es sólo cómo se codifica y transmite la información a nivel neuronal, sino cómo el cerebro procesa funciones cognitivas complejas. Un fenómeno interesante se observa en la toma de decisiones, donde la relación entre el tiempo y las decisiones se aproxima a la forma de $\log M$. Hick (1952) observó esta relación, aunque ajustó su modelo a $\log M+1$, sugiriendo que el tiempo de respuesta varía logarítmicamente con la cantidad de opciones equiprobables (Hyman, 1953). Esta característica logarítmica es similar a la forma en que se calcula la entropía de una variable equiprobable (Sayood, 2018), indicando que los principios subyacentes de las neuronas pueden extenderse a procesos cognitivos más complejos.

Este análisis se enriquece al considerar el trabajo de Alexandre, Oleg y Giovanni (2018), quienes ofrecen una perspectiva basada en la teoría de la información para entender cómo funciona la cognición, especialmente en situaciones que desafían nuestras expectativas diarias. Central para este entendimiento es la divergencia de Kullback-Leibler (Kullback & Leibler, 1951), una métrica que mide cuánto difiere una distribución de probabilidad de otra. Este concepto proviene de la teoría de la medición y está relacionado con el teorema Radon–Nikodym, que se ocupa de la representabilidad de medidas continuas (Nikodym, O., 1930; Blasco, O., 1996). El teorema dice que dada una medición positiva de Lebesgue μ en un espacio (X, Σ) , respecto a una medición λ de Lebesgue absolutamente continua con μ ($\lambda \ll \mu$, por lo que si $\lambda(A) = 0 \rightarrow \mu(A) = 0$), existe una función $f: X \rightarrow [0, \infty]$ tal que

$$\lambda(A) = \int_A f d\mu, \tag{3.1.1.1}$$

Donde f es la derivada Radon–Nikodym de λ con respecto a μ , la cual funciona como una función de densidad para la medición. Este teorema establece que, bajo ciertas condiciones,

es posible representar una medida en función de otra. Halmos y Savage (1949) encontraron aplicaciones para este teorema en la teoría de la probabilidad y la estadística. Su aplicación en este campo subraya la interconexión entre las estadísticas y sus propiedades probabilísticas, proporcionando una base para la teoría de la estadística suficiente.

Halmos y Savage adaptaron el teorema Radon–Nikodym para aplicarlo a la estadística, creando un puente entre la teoría de la medición y la estadística (Halmos, 2013). Esta adaptación fue retomada por Kullback y Leibler (1951) para explorar la suficiencia estadística. Se define una estadística como "suficiente" si no se necesita información adicional a la estadística para determinar la distribución de probabilidad de la población de la que se tomó una muestra. Esta perspectiva implica que la información, según la definición de Shannon-Wiener, no cambia mediante operaciones estadísticas cuando se utilizan estadísticas suficientes. Asumiendo en primera instancia que dados los espacios de probabilidad (X, S, μ_1) , $i = 1, 2$, tal que existe una equidad $\mu_1 \equiv \mu_2^1$ y considerando a λ como la medida de la probabilidad que cumple con la equivalencia $\lambda \equiv \{\mu_1, \mu_2\}$ (esto nos quiere decir que λ puede ser tanto μ_1 , o μ_2 , o también $\frac{1}{2}(\mu_1 + \mu_2)$, etc.). Recordando al teorema Radon–Nikodym, se tiene que debe existir una función $f_i(x)$, $i = 1, 2$, única a la medición de conjuntos de medida cero en λ , medible en λ con $0 < f_i(x) < \infty[\lambda]$, $i = 1, 2$ tal que

$$\mu_1(E) = \int_E f(x) d\lambda(x)$$

$$i = 1, 2, \quad (3.1.1.2)$$

para todos los valores de $E \in S$. Si H_i , $i = 1, 2$, es la hipótesis de que se ha seleccionado x de una población donde su probabilidad de medición es μ_1 , $i = 1, 2$, se define a

$$\log \frac{f_1(x)}{f_2(x)}, \quad (3.1.1.3)$$

como la información en x para la discriminación entre H_1 y H_2 . Y la media de esta información, una generalización de la formulación de Shannon, es:

$$I_{1:2}(E) = \frac{1}{\mu_1(E)} \int_E d\mu_1(x) \log \frac{f_1(x)}{f_2(x)} = \frac{1}{\mu_1(E)} \int_E f_1(x) \log \frac{f_1(x)}{f_2(x)} d\lambda(x)$$

para $\mu_1(E) > 0$,

$$= 0$$

para $\mu_2(E) = 0$.

(3.1.1.4)

La cual es la generalización de la información de Shannon, ya que como él lo hizo, definió a la información en un espacio finito y discreto. Entonces tenemos que el cálculo de la divergencia entre dos distribuciones de probabilidad, denotado por $J(1, 2)$, es dado por

$$J(1, 2) = J_{12}(X) = \int (f_1(x) - f_2(x)) \log \frac{f_1(x)}{f_2(x)} d\lambda(x),$$

(3.1.1.5)

Esta divergencia, llamada divergencia Kullback-Leibler, representa la cantidad de información que diferencia una distribución de probabilidad de otra. Es esencial resaltar que esta formulación generaliza y extiende la comprensión de la información y la suficiencia estadística al conectarse con conceptos fundamentales de la teoría de la medición. La divergencia entre dos distribuciones de probabilidad es esencial para entender cuánta información se necesita para actualizar un modelo por otro. Esta divergencia se denomina "entropía relativa" y nos da una medida de la diferencia de información entre dos distribuciones de probabilidad. Partiendo del ejemplo anterior es posible definir la divergencia de dos distribuciones de probabilidades X y Y con densidades $p(x)$ y $q(x)$ respectivamente. como

$$D(X||Y) = - \sum \log \frac{Pr_y}{Pr_x} Pr_x$$

(3.1.1.6)

Y considerando que existe una asociación con el espacio de muestra $S = \{o_1, \dots, o_x, \dots, o_n\}$

$$D(X||Y) = \sum_{j=1}^n p_j \log\left(\frac{p_j}{q_j}\right) = E_p \log\left(\frac{p_j}{q_j}\right), \quad (3.1.1.7)$$

Donde E_p es el valor esperado de p_j y $\log(p_j/q_j)$ como la información para la discriminación entre p y q . Suponiendo que Y está uniformemente distribuida ($q_j = 1/n$):

$$D(X||Y) = \log(n) - H(X), \quad (3.1.1.8)$$

A esta divergencia es a lo que se le llama *entropía relativa*, y es posible reescribir esta misma ecuación en términos de su contenido de información

$$D(X||Y) = - \sum_{j=1}^n p_j (I(y) - I(x)), \quad (3.1.1.9)$$

Esto significa que la entropía relativa mide la diferencia promedio de información por resultado entre X y Y . El costo de los procesos cognitivos, aunque a primera vista no parezca directamente relacionado con la teoría de la información, se puede entender como una refinación de los modelos internos basada en nueva información. Esta idea, basada en la divergencia Kullback-Leibler, sugiere que la cognición tiene costos asociados, ya sea en términos de energía, metabolismo o capacidad de almacenamiento de la información. Estos costos están relacionados con la discrepancia entre lo que ya sabemos (nuestra representación interna) y lo nuevo que observamos o aprendemos (Zenon, Solopchuk, & Pezzulo, 2018).

El costo cognitivo se conceptualiza en términos de la divergencia de información entre estos dos modelos: el interno y las observaciones del mundo. Matemáticamente, el costo se define como

$$C = I(x; z) + I(z; y) \quad (3.1.1.10)$$

Donde $I(x; z)$ es la *información mutua* entre x y z que habla de la reducción de información o entropía en z después de las observaciones en x , mientras que $I(z; y)$ representa la reducción de entropía o información de una respuesta dada una representación interna sobre los datos de entrada z (Hick, 1952; Hyman, 1953). Es importante destacar que x representa la información sensorial externa y z simboliza cómo interpretamos internamente esa información. Estas representaciones internas o interpretaciones, como sugieren diferentes modelos de cognición, son esenciales para nuestra comprensión y respuesta al mundo (Park, & Pillow, 2017; Genewein et al., 2015; Tishby et al., 2000). Un ejemplo práctico es la prueba de Stroop, donde dos fuentes de información (nombre del color y el color real del texto) pueden ser contradictorias, lo que exige un esfuerzo cognitivo adicional para dar una respuesta correcta. En resumen, el costo cognitivo se refiere a cómo nuestro cerebro procesa, compara y ajusta información basándose en modelos internos y entradas sensoriales.

La cognición, y en particular la tarea de leer, conlleva un proceso automático y un costo asociado que se ve influenciado por el contexto. En esencia, cuando leemos, tenemos una respuesta automática esperada, $p(y|z)$, que se actualiza según el contexto $p(y|zT)$. Este proceso puede ser representado matemáticamente

$$p(y|z) = \sum_T p(y|z, T)p(T), \quad (3.1.1.11)$$

El costo total de información para este proceso viene dado por

$$C = I(x;y) + I(z;y) + I(T;y|z), \quad (3.1.1.12)$$

Siendo $I(x; y)$ el costo perceptual, $I(z; y)$ es el costo del proceso automático de respuesta, y $I(T; y|z)$ siendo el costo del proceso dependiente del contexto. Cambiar la acción, como en la prueba de Stroop donde se pasa de leer a decir el color, introduce costos adicionales que concuerdan con investigaciones previas (Rubin, & Meiran, 2005). Es interesante notar que la

relación señal-ruido juega un papel crucial en este costo. Cuando esta proporción disminuye, observamos un decremento en la eficiencia de las tareas y un aumento en el esfuerzo cognitivo (Manohar et al., 2015; Sarampalis et al., 2009; Ohlenforst et al., 2017). Estos costos cognitivos pueden ser entendidos desde un punto de vista termodinámico y biológico, considerando las interacciones neuronales y cómo estas consumen energía.

Pasando a la relación entre teoría de la información y procesamiento cerebral, la ley de Hick, a pesar de su relevancia, no parece describir totalmente la realidad. Aunque relaciona acertadamente cantidad de elecciones con tiempo de decisión, la estructura lineal podría no ser del todo adecuada. En situaciones con múltiples decisiones, una distribución sigmoide se ajusta mejor, confirmando el papel de la entropía en la predicción de procesos cognitivos (Pavão et al., 2016). Estos costos energéticos en la cognición, descritos por Vigotski como "gasto de energía psíquica", no se limitan solo al cerebro. Los efectos del pensamiento se extienden al sistema nervioso periférico, mostrando que nuestras respuestas cognitivas tienen una manifestación física, como mover los labios al leer o calcular mentalmente. En resumen, Vigotski (2001) sugiere que el pensamiento puede verse, hasta cierto punto, como una conducta.

Queda claro como la divergencia Kullback-Leibler, permite aproximar una medida de la eficiencia de la transmisión de información del cómo el cerebro procesa información con costos cognitivos. En neurociencia computacional, la divergencia KL se utiliza para estimar estos costos al cuantificar la discrepancia entre las distribuciones de probabilidad de los estímulos percibidos y las respuestas neuronales. Así, facilita la comprensión de cómo el cerebro optimiza el procesamiento de información minimizando el "sorpresa" o desviación de las expectativas, lo que refleja un paralelo con cómo un Perceptrón Multicapa ajusta sus pesos para minimizar la diferencia entre las predicciones y los datos reales, subrayando la relevancia de los principios de la teoría de la información tanto en la inteligencia artificial como en la cognición humana

3.2 Codificación de la información

La comunicación entre neuronas es vital para la adaptación y respuesta al ambiente. Estas células, con su vasta diversidad en forma y función, transmiten señales electroquímicas llamadas picos de activación. Sin embargo, su capacidad de transmisión está limitada por el canal de comunicación. Este límite ha llevado a la especulación sobre un lenguaje o código común que las neuronas emplean para asegurar señales coherentes y efectivas (Berger, & Levy, 2010). Las neuronas, al comunicarse, transmiten mensajes a varias contrapartes simultáneamente, sugiriendo la existencia de una codificación especializada (Wang, & Poor, 2004). En términos de teoría de la información, este fenómeno se asimila al acceso múltiple por división de frecuencia o FDMA (Frequency Division Multiple Access). Este proceso se alinea con la "hipótesis del código eficiente" derivada de las propuestas de Barlow (1961). Más allá de la comunicación interneuronal, es esencial reconocer la codificación en respuestas automáticas o reflejas del cuerpo. Estas respuestas pueden ser influenciadas por diversas áreas del sistema nervioso e incluso ser activadas voluntariamente, como se observa en estructuras como el cerebelo (Walter & Khodakhah, 2006).

La esencia de la comunicación neuronal radica en el mensaje y su codificación. Kolmogorov (1963) propuso una definición de complejidad, referida como la complejidad Solomonoff–Kolmogorov–Chaitin, que se equipara con la entropía H de Shannon. Esta relación sugiere que lo que procesa el cerebro humano es directamente proporcional a la cantidad de información del estímulo original. Así, la complejidad de Kolmogorov se presenta como una herramienta valiosa para comprender la información y se puede relacionar con la complejidad computacional y descriptiva de un sistema. Se puede decir entonces que el flujo de información entre neuronas está constreñido por la capacidad del canal de comunicación. Esta restricción, identificada por Shannon (1948), indica que las neuronas deben codificar eficientemente las señales para transmitir mensajes coherentes en el menor tiempo posible. Es esencial entender este proceso para comprender mejor la adaptabilidad y respuesta del sistema nervioso ante diferentes estímulos.

La codificación se refiere al conjunto de reglas que transforman o traducen la información de una forma a otra. Irvine y Harle (2001) identifican cuatro tipos principales de codificación de información:

Comprensión de la información: La codificación eficiente considera la compresión de información. En términos neurobiológicos, esta compresión es crucial dada la capacidad limitada del canal de comunicación, como neuritas, axones y dendritas. Estos elementos también poseen sistemas de retroalimentación, esenciales para la transmisión eficaz de la información.

Control de errores: La codificación para control de errores implica añadir redundancia al sistema para verificar la integridad de la información. Gracias a la teoría de la información de Shannon, sabemos que un mensaje es confiable si el ruido es menor que el contenido del mensaje. Ejemplos de esta codificación están presentes en las redes neuronales biológicas, que usan retroalimentación. Estos sistemas en el cerebro humano mejoran, por ejemplo, la precisión de las respuestas de las neuronas corticales visuales para identificar y contrastar imágenes (Bullier et al., 2001).

Codificación criptográfica: Como su nombre indica, busca ocultar el mensaje original, transformando símbolos (como palabras) en una forma que solo el receptor deseado pueda entender. Existen múltiples variantes de codificación criptográfica, cada una con un propósito específico.

Codificación o códigos en línea: Relacionado con la transmisión digital, este tipo de codificación convierte la información en una línea de datos que se envía a través de canales de comunicación como variaciones de voltaje o corriente eléctrica.

La codificación es un proceso esencial que transforma o traduce la información para su transmisión. Uno de sus componentes clave es el sistema de retroalimentación, que actúa como un puente entre varios tipos de codificación, permitiendo la corrección de errores y mejorando la eficiencia de la transmisión. El canal de comunicación entre neuronas, por ejemplo, puede modelarse como un DMC (canal discreto sin memoria). Cuando se integra un

sistema de retroalimentación sin ruido en este canal, el emisor tiene la capacidad de corregir errores basados en la respuesta del receptor. Esta retroalimentación reduce efectivamente los errores del receptor respecto al mensaje original (El Gamal & Kim, 2011). Un ejemplo prominente de este enfoque es el esquema de codificación Schalkwijk–Kailath, en el cual la probabilidad de error disminuye de manera doblemente exponencial cuando se utiliza retroalimentación en comparación con sistemas que no la tienen.

Dentro de un canal Gaussiano con retroalimentación y sin ruido, donde se tiene en cuenta el ruido $Z \sim N(0, 1)$, se puede describir una transmisión de señales que comienza en el tiempo $t = 0$. En este sistema, el codificador ajusta su mensaje basándose en el ruido percibido, creando un bucle de retroalimentación. Este proceso, en esencia, implica incorporar el ruido detectado Z_0 y corregirlo usando el sistema de retroalimentación, resultando en una transmisión de señales más eficiente. Lo fascinante es que este marco teórico de telecomunicaciones, desarrollado en el último siglo, desde el telégrafo hasta la tecnología celular actual, podría proporcionar intuiciones sobre cómo el cerebro codifica la información. Al analizar y comprender estos sistemas de comunicación avanzados, podríamos estar un paso más cerca de descifrar los misterios de la comunicación neural.

Considerando que la comunicación en redes neuronales no es un proceso simple de un punto a otro, sino una intrincada danza entre miles de millones de nodos, buscando transmitir mensajes de manera eficiente y precisa. La complejidad de esta comunicación difiere notablemente de la comunicación a través de un solo canal, y algunas leyes que aplican en sistemas más simples pueden no ser directamente aplicables a redes neuronales.

En la búsqueda por comprender cómo fluye la información en estas redes, se ha reconocido que la capacidad máxima de un canal no es necesariamente la que dicta el flujo máximo de información. Un estudio por Ford y Fulkerson en 1956, originalmente dirigido al tráfico y optimización del mismo, encontró propiedades aplicables a todo tipo de redes, incluidas las neuronales. Shannon, Elías y Feinstein propusieron una ecuación, en 1956, para medir la información en redes. Esta ecuación resalta la capacidad del "corte" entre distintas partes de la red (Elias et al., 1956). Esencialmente, demostraron que, al agregar nodos intermediarios

(retransmisores), la probabilidad de errores no aumenta, independientemente de la distancia entre nodos comunicantes.

Una hipótesis temprana sobre cómo el cerebro codifica la información fue propuesta por Barlow en 1961. Sugirió tres funciones principales para los retransmisores de señales en nuestro sistema nervioso: funcionar como una contraseña específica para la transmisión de información, actuar como sistemas de retroalimentación que ajustan la señal hasta que sea correcta y reducir la redundancia en la transmisión de señales. Sin embargo, los experimentos de Barlow mostraron que el cerebro no siempre transmite señales de la manera teóricamente más eficiente.

Esta aparente ineficiencia se debe a la increíble complejidad del sistema nervioso, que debe manejar múltiples señales simultáneas, integrar diferentes tipos de información y producir respuestas conductuales. Además, a pesar de su complejidad, el cerebro puede usar circuitos relativamente simples para procesar y transmitir información. Un ejemplo es el caso de ciertos crustáceos que, con redes neuronales comparativamente simples, pueden tomar decisiones conductuales óptimas (Crossley, Staras, & Kemenes, 2016). Además, algunas células, como las células Purkinje, parecen realizar gran parte de su procesamiento interno, aprovechando propiedades básicas de los canales de calcio y sodio y transmitiendo señales de manera casi lineal, a pesar de su intrincada arquitectura y conexiones (Schonewille et al., 2006). El cerebelo desempeña un papel vital en la coordinación de la información sensorial y cortical, especialmente en lo que respecta a las señales de movimiento. Esta coordinación es posibilitada por la estructura y conexiones de las células del cerebelo, como lo describió Ramon y Cajal (1911). Al investigar cómo se codifica la información en el cerebelo, Walter y Khodakhah (2006) estudiaron la actividad de las células Purkinje al estimular células granuladas vecinas con glutamato. Observaron que la respuesta de las células Purkinje era directamente proporcional a la fuerza de la estimulación, sugiriendo un mecanismo de codificación lineal en esta región del cerebro. Sin embargo, esta codificación lineal es solo una de las formas en que el cerebro procesa la información. Debido a la complejidad y diversidad del sistema nervioso, existen diversas formas de codificación adaptadas a las necesidades evolutivas y funciones de cada organismo, desde la regulación de actividades automáticas hasta el procesamiento de experiencias conscientes.

3.2.1 Velocidad sináptica, frecuencias y codificación

El cerebro comunica y procesa información a través de complejas interacciones electroquímicas entre neuronas y neuritas. La velocidad y naturaleza de estas comunicaciones nos da una perspectiva sobre la cantidad de información que el cerebro puede procesar en un tiempo determinado. Al evaluar la velocidad de transmisión y la densidad neuronal, podemos estimar cuánta información se comunica entre las neuronas. Para analizar estas conexiones sinápticas y determinar si hay variaciones entre ellas, se han empleado métodos ópticos. Estos estudian cuán rápidamente los receptores sinápticos regresan a un estado que les permite enviar o recibir otra señal. Esto nos ofrece una perspectiva directa sobre la velocidad de procesamiento de un grupo específico de neuronas.

Skaggs, McNaughton y Gothard (1993) propusieron un enfoque teórico-práctico para calcular cuánta información se comunica en un intervalo de tiempo. En un experimento, colocaron a una rata en un cilindro con pellets en diferentes lugares, intentando comprender la codificación de la información del hipocampo de la rata. Se propuso un modelo donde la información se considera en función del tiempo, representada por la ecuación:

$$I = \int_x \lambda(x) \log_2 \left(\frac{\lambda(x)}{\lambda} \right) p(x) dx, \quad (3.2.1.1)$$

Donde I es la información que se transmite en bits por segundo, x es la localización espacial, $p(x)$ es la densidad de probabilidad de que la rata este en la ubicación x , $\lambda(x)$ es la media de disparos o picos eléctricos registrados cuando se encuentra en la locación x , y $\lambda = \int_x \lambda(x)p(x)$ es la media de los picos registrados en general. El resultado mostró que la activación celular era más alta cuando la rata estaba cerca de una pella. Se encontró que la posición y orientación de la cabeza de la rata daban una mayor cantidad de información que solo su ubicación espacial. Además, cuando se combinaba la información de ambas variables, el valor resultante no era la suma directa de ambas, sugiriendo una correlación en la codificación y envío de información.

El cálculo de la información transmitida puede generalizarse con la ecuación

$$X = \sum_s P(s) \left(\frac{r_s}{\langle r \rangle} \right) \log_2 \left(\frac{r_s}{\langle r \rangle} \right), \quad (3.2.1.2)$$

Donde $\langle r \rangle$ es la tasa media entre estímulos generales en varias pruebas, y r_s es la media de disparos observados (Rolls & Treves, 2011). Esta ecuación permite calcular la cantidad de información transmitida en bits.

Los procesos electroquímicos que permiten la comunicación entre neuronas y neuritas son fundamentales para entender la capacidad del cerebro humano para procesar y transmitir información. Esta transmisión no es uniforme y su velocidad varía según diversos factores.

El estudio de la velocidad a la que se envían estos mensajes nos brinda valiosos datos sobre la capacidad máxima de procesamiento cerebral en un periodo determinado. Para comprender este proceso, la neurociencia ha utilizado técnicas como la regresión y métodos ópticos. Estos ayudan a determinar la rapidez con la que los receptores sinápticos regresan a su estado normal, permitiendo recibir o enviar otra señal. Por ende, estos métodos nos proporcionan una visión directa de la capacidad de procesamiento de ciertos grupos de neuronas (Skaggs et al., 1993).

Skaggs, McNaughton y Gothard (1993) desarrollaron un experimento para cuantificar la cantidad de información transmitida en un intervalo de tiempo. Situaron a una rata en un cilindro con alimentos en diferentes lugares, observando cómo el hipocampo codificaba esa información. Descubrieron que la cantidad de información se relaciona con variables como la posición de los alimentos respecto a la rata, su ubicación y orientación. Sorprendentemente, la suma de estas variables no daba un valor total, sino que mostraba alguna correlación en la codificación y transmisión de la información.

A nivel bioelectroquímico, existen diferentes canales de comunicación, asociados a distintos tipos de neurotransmisores, como aminoácidos, gasotransmisores, monoaminas, y otros (como el acetilcolina y anandamida).

Es esencial también distinguir la actividad neuronal por sus oscilaciones. Estas oscilaciones, como las ondas *delta*, *theta*, *alpha*, *beta* y *gamma*, varían en frecuencia y se asocian con distintos estados y funciones cerebrales, como el sueño profundo, el estrés emocional o la relajación (Sawant & Jalali, 2010). Estas ondas pueden verse como canales de comunicación, que evitan la sobrecarga de información sensorial. Es relevante no solo la activación de estas ondas, sino también su desactivación. Recientes investigaciones han destacado la importancia de estas oscilaciones en la codificación de la información. No se trata solo del procesamiento individual de una neurona, sino del patrón global de codificación cerebral (Schyns, Thut & Gross, 2011). Por ejemplo, diferentes ondas cerebrales trabajan en conjunto para tareas como el reconocimiento facial o la memoria (Hanslmayr, Staudigl, Fellner, 2012).

La alta conectividad entre regiones cerebrales significa que hay una fuerte interacción entre los sistemas de memoria en la mayoría de las situaciones. Una teoría reciente, PIMMS (sistemas interactivos predictivos de memoria múltiple), propone que el cerebro intenta predecir su entorno y que las memorias se distribuyen por todo el cerebro, optimizando la predicción y minimizando los errores (Henson, & Gagnepain, 2010). Esta visión permite una comprensión más rica de cómo y dónde se almacenan las memorias y por qué parecen dispersas por todo el cerebro. Aunque áreas como el tallo cerebral son vitales para la memoria a largo plazo, es fascinante observar que incluso cuando estas áreas están comprometidas, algunas funciones, como aprender nuevas habilidades, permanecen intactas.

El cerebro es una entidad compleja que procesa información a través de descargas neuronales sincronizadas. Estas descargas generan fluctuaciones rítmicas, que son registradas por el electroencefalograma (EEG). En esencia, el EEG nos muestra una especie de "promedio" de la actividad cerebral en ciertas áreas del cerebro. Así, podemos interpretar estas fluctuaciones como la forma en que el cerebro procesa la información.

Sin embargo, analizar estas señales no es sencillo. Una de las dificultades principales es que, al intentar medir la entropía espectral a través del EEG, esta no se define en función del tiempo. Eso restringe su estudio. Un método para superar parcialmente esto es usar transformaciones Fourier, pero tiene sus propios problemas, principalmente relacionados con la precisión de las mediciones. En términos sencillos, si nos enfocamos mucho en un breve periodo de tiempo, perdemos precisión en la frecuencia de la información, y viceversa (Rosso, 2007). Una alternativa es la entropía Wavelet, que nos brinda una mejor separación del ruido y una comprensión más detallada de la actividad cerebral. Desde una perspectiva más amplia, se puede ver al cerebro y su red neuronal de diversas maneras: como una unidad, como una red compleja o como la masa total del cerebro en acción. Lo fascinante es cómo grandes grupos de neuronas pueden trabajar juntos o separados, y aún así lograr procesar información en tiempos extremadamente rápidos, como 20 milésimas de segundo. Y curiosamente, cuanto más neuronas estén disponibles para procesar, más rápido y eficiente se vuelve el procesamiento, al menos hasta cierto punto (Rolls & Treves, 2011).

El comportamiento de las neuronas también varía según las áreas del cerebro. Por ejemplo, las neuronas en el hipocampo procesan información espacial, mientras que otras áreas, como el pre-subículo (área 27 de Brodmann), están involucradas en procesar la dirección de la cabeza en primates (Rolls & Treves, 1998; Robertson et al., 1999). Esta variedad se debe en parte a la diversidad de células neuronales y sus características únicas. Además, mientras que algunas neuronas podrían estar más involucradas en almacenar recuerdos, otras juegan un papel más importante en procesar estímulos sensoriales (Rolls & Treves, 1990; Rolls, 2008).

El cerebro humano es una estructura compleja que procesa información a través de redes neuronales. Estas redes interactúan y responden a estímulos de diversas maneras, dependiendo de las características y localizaciones de las neuronas involucradas. Al analizar estas respuestas y patrones de activación, podemos comenzar a entender cómo el cerebro codifica y procesa la información sensorial. Las características topológicas y topográficas nos ayudan a entender esta codificación. Mientras la topología se enfoca en el procesamiento de información sensorial, la topografía observa las rutas específicas que sigue la información. Estas dos características influyen cómo se procesa la información. Por ejemplo, la codificación de la información auditiva es sensible a la velocidad de los picos de activación

neuronal, y su representación en el cerebro depende de este ritmo (Decharms & Merzenich, 1996). Por otro lado, la codificación visual no parece depender tanto de estos patrones de activación.

Un aspecto crucial en la codificación neuronal es la oscilación de las señales. Aunque no aumenta la cantidad de información, sí puede influir en la velocidad de procesamiento. Rolls y Treves (2011) identificaron varios principios relacionados con la oscilación:

1. Las oscilaciones pueden aumentar la cantidad de picos de activación, mejorando la velocidad de procesamiento en varias redes neuronales.
2. Estas oscilaciones también pueden sincronizar picos de activación, conduciendo a una generación más rápida de potenciales de acción. Esto se evidencia al aumentar las oscilaciones gamma y se ve afectado al inhibir ciertas conductancias (Buehlmann & Deco, 2008; Gauck & Jaeger, 2003).
3. La coherencia cortical puede influir en cómo se transmite la información entre diferentes poblaciones neuronales. Esto se basa en la hipótesis CTC (communication-through-coherence). Además, una función adicional de esta coherencia podría ser reiniciar la actividad neuronal, lo que permitiría múltiples intentos para interpretar información sensorial ambigua en un corto período (Rolls & Treves, 1998; Jezek et al., 2011).
4. Recuperar una memoria autoasociativa podría suceder dentro de un ciclo theta, independiente de las memorias en ciclos siguientes.

Algunas investigaciones indican la capacidad de ciertas células del cerebelo para sincronizar la actividad neuronal, y estas propiedades oscilatorias nacen de interacciones neuronales (Vos et al., 1999; Gauck & Jaeger, 2003). Las interacciones y caminos de la información en forma de picos de activación reflejan la complejidad del cerebro y la importancia de considerar tanto activaciones individuales como interacciones en entornos neuronales variados.

Existen por ejemplo otras hipótesis de cómo se podría codificar la información en el cerebro, tal es el caso de la teoría holonómica, de la cual algunos aspectos de la teoría, como la idea de

que la memoria y las percepciones funcionan de una manera no localizada y distribuida, han encontrado respaldo en investigaciones sobre cómo el cerebro procesa la información.

Esta se basa en campos de potencial locales en las fibras de las redes neuronales, utiliza transformaciones Fourier o wavelets que operan en coordenadas espacio-temporales, proceso conocido como "Holografía Cuántica" (Marcer, P., & Schempp, 1998). Denis Gabor resaltó la relevancia de las interferencias sobre las ondas en dichas transformaciones, crucial tanto para la transmisión óptima de señales eléctricas como para la interpretación de fenómenos cuánticos (Gabor, D., 1946). En el ámbito de la decodificación visual, Pribram y Carlton (1986) propusieron que los movimientos nistagmoides y la actividad oculomotora son claves para traducir patrones visuales en "Grupos de simetría", coordinados por el córtex motor adyacente al córtex visual primario.

La transformada de Fourier corta (STFT) añade una dimensión de análisis al integrar el tiempo y el espacio en su fórmula. Según Switkes, Bradley, y De Valois (1988), las neuronas del córtex visual responden no sólo a líneas rectas sino a patrones más complejos, favoreciendo la percepción de imágenes granuladas. Estas características de las transformaciones Fourier son fundamentales para distinguir frecuencias individuales y bandas de frecuencia en electroencefalogramas y es la propuesta que dan de cómo puede estar codificando el cerebro estas señales visuales.

Con todo esto en cuenta queda claro que desde la perspectiva de la teoría de la información, las funciones corticales humanas pueden modelarse hasta cierto punto mediante redes neuronales artificiales, lo que ayuda a investigar fenómenos como la conciencia y la memoria. Estos modelos también explican cómo optimizar la comunicación neuronal reduciendo la información procesada y los recursos utilizados, clave para la eficiencia energética en los seres vivos. Sin embargo, la teoría de la información enfrenta limitaciones en el control experimental y la complejidad de las funciones cognitivas superiores. Los estudios teóricos sobre el gasto energético para transmitir un bit de información estiman un coste de 2.805 zJ a temperatura ambiente de 293.15 K (Szilard, 1929; Landauer, 1961), aunque en la práctica, los costos metabólicos adicionales y la redundancia en las redes aumentan significativamente este valor.

3.3 Negentropía

La relación entre los seres vivos y la termodinámica siempre ha parecido paradójica. Mientras que la termodinámica establece que los sistemas evolucionan hacia un estado de máxima entropía o desorden, los organismos vivos desafían esta norma, manteniendo estructuras complejas y ordenadas durante su vida. Esta capacidad de los seres vivos para mantener el orden ha sido un enigma desde un punto de vista termodinámico. Los seres vivos requieren de una fuente de energía constante, ya sea química o radiactiva, para mantener sus funciones vitales. De alguna manera, utilizan esta energía para resistir el aumento de entropía en sus sistemas. Schrodinger (1944) sugirió en su libro "¿Qué es la vida?" que los seres vivos se "alimentan" de entropía negativa para mantener su estructura y función. La conexión entre la termodinámica y la teoría de la información se hizo evidente cuando Brillouin (1953) integró la descripción estadística de la entropía de Boltzmann con conceptos de información. Simplificando, propuso una relación donde la entropía final de un sistema es afectada por la información que se obtiene de él. Esta perspectiva sugiere que hay un límite en nuestra capacidad para medir sistemas sin afectar su entropía.

Un término importante surgido de estas discusiones es la "negentropía". A diferencia de la entropía, que describe el desorden, la negentropía se refiere al orden o la información en un sistema. Schrödinger (1944) postuló que los seres vivos se sustentan en negentropía, la cual puede ser vista como la energía utilizable para mantener la vida. Un ejemplo de cómo los seres vivos utilizan negentropía es el proceso de las bombas de sodio y potasio en las membranas celulares, que permiten a ciertos iones pasar y repeler a otros, algo esencial para el funcionamiento de las neuronas. La comprensión de los seres vivos desde la termodinámica requirió el desarrollo del concepto de sistemas abiertos, que intercambian energía y materia con su entorno (Ho, 1994). Es crucial distinguir entre energía libre, disponible para uso, y energía almacenada, reservada para tiempos futuros (Friston, 2010). Por otro lado, Wiener ofreció una conceptualización diferente de la información comparada con la de Shannon. Mientras que Shannon la veía como una medida de desorden, Wiener la entendía como orden o negentropía. Esta dualidad refleja la complejidad de la relación entre información, entropía y vida.

Los conceptos de entropía y negentropía en la teoría de la información pueden parecer complejos inicialmente. En el núcleo de estas ideas, encontramos la relación entre orden, desorden, y cómo se mide y se comprende la información. La negentropía no es simplemente el opuesto de la entropía. En lugar de verla como el valor negativo de la entropía, es más preciso entenderla como la "distancia" entre una distribución de probabilidad dada y la distribución gaussiana, que representa el estado de mayor desorden o entropía. Una distribución que tiene una estructura o orden se alejará de esta distribución gaussiana. Por lo tanto, una mayor negentropía indica más información y una mayor desviación del estado de máxima entropía posible.

Para clarificar, imaginemos que estamos evaluando la confiabilidad de un sistema basándonos en su estructura probabilística. En este contexto, la información total del sistema se compone tanto de la entropía (H) como de la negentropía (I). Si ambos valores son iguales, entonces el sistema está en su máximo estado de caos. Sin embargo, cualquier reducción en la entropía se reflejará como un aumento en la negentropía. Esta dualidad se asemeja a la idea de que un sistema puede estar en un estado operativo o no operativo.

Brillouin (1953) y Dulesov, Karandeev, & Krasnova (2017) sugieren que, al analizar sistemas, es posible diferenciar cualitativamente la información. Por ejemplo, si se examina un sistema técnico con estados operativos y no operativos, se pueden usar las fórmulas de Shannon para esta diferenciación cualitativa. Estas fórmulas muestran que parte de la información se relaciona con la entropía y otra con la negentropía. Esencialmente, la entropía y la negentropía son como las dos caras de una moneda, donde una representa el desorden y la otra el orden.

La negentropía y la entropía son conceptos fundamentales en la teoría de la información. A diferencia de la entropía, que mide el desorden o incertidumbre, la negentropía refleja el grado de orden o estructura en la información. No se trata simplemente del valor negativo de la entropía; en cambio, la negentropía se refiere a la distancia entre una distribución de probabilidad y la distribución gaussiana o normal, la cual simboliza el estado de máxima entropía posible (Dulesov, Karandeev, & Krasnova, 2017; Dulesov, Karandeev, & Dulesova, 2018).

Esta distancia es crucial porque cuanto mayor sea la negentropía, mayor será la información presente y, por lo tanto, mayor será la distancia con respecto al estado de máxima entropía. Esto es esencial para sistemas técnicos y seres vivos, ya que buscan mantener niveles elevados de negentropía para funcionar de manera óptima. El debate sobre la posibilidad de que ciertos fenómenos, como los movimientos Brownianos y cuánticos, violen la segunda ley de la termodinámica ha sido un tema de interés. A pesar de que estos movimientos parecían contradecir esta ley, investigaciones demostraron que no se produce una violación en el mundo cuántico (McClare, 1971; Ford & O'Connell, 2006).

Lo fascinante es cómo podemos relacionar estos conceptos con el cerebro humano. Se puede considerar al cerebro como un dispositivo termodinámico que procesa información. Esta perspectiva combina las teorías de la información y termodinámica, ofreciendo un enfoque novedoso para entender la mente humana (La Cerra, 2003; Collell & Fauquet, 2015). Estudios han demostrado equivalencias entre la forma en que funcionan ciertas células cerebrales y las redes neuronales artificiales, como las células Purkinje y el Perceptrón (Brunel et al., 2004). El cerebro busca minimizar la energía libre para maximizar la eficiencia y la información mutua entre los estados internos y los sensoriales (Ashby, 1991; Dayan, Hinton, Neal, & Zemel, 1995; Friston, Kilner, & Harrison, 2006).

Por lo que aunque la entropía y la negentropía están relacionadas con el concepto de información, no son lo mismo (Cackowski, 1982). Es fundamental reconocer esta diferencia, pues la información es contextual y requiere un procesamiento adecuado para ser útil. Abordar estos fenómenos desde una perspectiva multidimensional es esencial para comprender plenamente la complejidad de la comunicación y la información en seres vivos y máquinas (Corning & Kline, 1998; Corning, 2007).

Capítulo 4: Cibernética, control y comunicación

4.4 Cibernética

El término "cibernética" proviene del griego "*kybernetes*", que Platón usó en "La república" para referirse al acto de dirigir o gobernar (Vallée, 1990). En el siglo XIX, Ampère lo utilizó como sinónimo de la ciencia del gobierno. Sin embargo, fue Norbert Wiener quien, en 1948, redefinió cibernética como "El estudio del control y la comunicación en el animal y la máquina" (Wiener, 1948). Wiener se inspiró en la necesidad militar de mejorar sistemas de control y comunicación. A diferencia de otras disciplinas, la cibernética se centra en cómo sistemas naturales y artificiales establecen y persiguen sus objetivos. Durante la Segunda Guerra Mundial, Wiener y el ingeniero Julian H. Bigelow desarrollaron sistemas para predecir trayectorias de aviones enemigos, junto con el desarrollo de una teoría matemática de la comunicación (o información) por Claude Shannon (1948, 1953). Estos sistemas, llamados servomecanismos, usaban datos pasados para hacer predicciones. Sin embargo, Wiener notó que reducir la fricción causaba oscilaciones incontrolables en estos sistemas.

Al discutir con el médico Arturo Rosenblueth, Wiener encontró similitudes entre este comportamiento errático de las máquinas y ciertas enfermedades humanas, como cuando una persona con una lesión en el cerebelo derrama agua al intentar beber. Estas observaciones lo llevaron a concluir que cualquier acción controlada necesita un circuito cerrado de información que permita ajustar el comportamiento según los resultados anteriores. Wiener y Bigelow identificaron este principio como retroalimentación negativa, vital tanto en la maquinaria como en los organismos humanos para corregir acciones.

Para abordar estos conceptos, se realizaron las Conferencias de Macy entre 1944 y 1953, reuniendo a expertos como Wiener, John Von Neumann, y Claude Shannon (Pias, 2016). Esta serie de encuentros consolidó la cibernética como un campo interdisciplinario, expandiéndose a áreas como la psicología y la sociología. Paralelamente, la conceptualización moderna de la información, como algo medible, se originó de la reinterpretación que hizo Claude Shannon (1948) de la entropía en términos estadísticos. Shannon postuló que la entropía de un sistema representa la cantidad mínima de información necesaria para describirlo completamente. Esta

nueva perspectiva sobre información y entropía definió límites y capacidades en la transmisión de datos. Ya que para la época, el enfoque analítico tradicional era insuficiente para abordar estas cuestiones complejas y no lineales.

Las funciones de series de tiempo tienen la capacidad de expandirse en base a la magnitud de sus evaluaciones numéricas. En el contexto de juegos, una máquina asigna valores tentativos a piezas y movimientos y, basándose en ellos, toma decisiones. Sin embargo, el verdadero aprendizaje comienza cuando la máquina redefine estos valores tras cada juego, asignando un valor de 1 para juegos ganados, 0 para los perdidos y 0.5 para empates. Así, mejora su estrategia con el tiempo (Wiener, 1961). Este principio es esencial en el funcionamiento de las redes neuronales artificiales. La cibernética se centra en la naturaleza de los sistemas no lineales. Las leyes físicas son inmutables con el tiempo, lo que significa que los experimentos producirán los mismos resultados independientemente de cuándo se realicen. Por ejemplo, en circuitos eléctricos, las oscilaciones pueden reducirse a una combinación de dos funciones trigonométricas, permitiendo el análisis armónico de propiedades lineales. Pero cuando se combinan con otras funciones, esta propiedad no se mantiene.

Para predecir trayectorias, como la de una pelota en movimiento, una extrapolación lineal simple es efectiva. Sin embargo, si el observador también se mueve, la predicción se complica. En este contexto, surge la retroalimentación (feedback). Por ejemplo, un piloto ajusta su vuelo basándose en la respuesta de su avión. La retroalimentación también está presente en acciones simples como tomar un lápiz. Problemas en el sistema nervioso pueden interrumpir esta retroalimentación, como en enfermedades como el Parkinson o la epilepsia, donde se observan alteraciones en la señalización nerviosa y oscilaciones en la actividad neuronal.

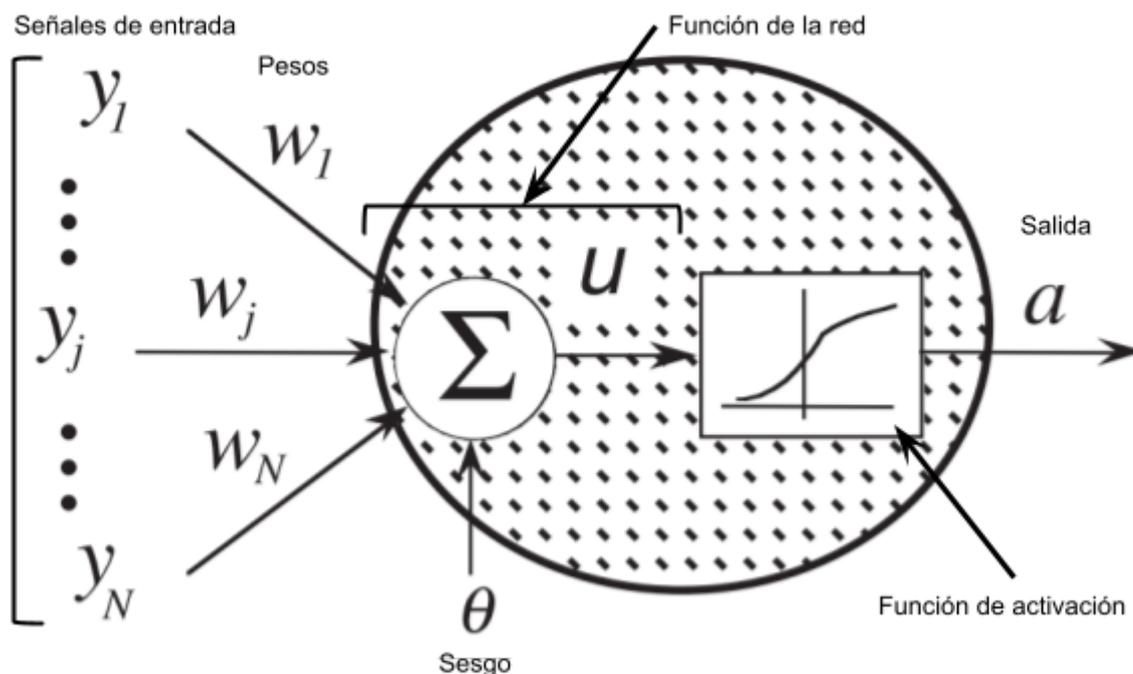
En epilepsia por ejemplo, la actividad anormal puede deberse al bloqueo de conductancias inhibitorias o a la activación de conductancias excitatorias. Intervenir en estos procesos, junto con entrenamiento en biofeedback, ha mostrado ser efectivo en controlar episodios epilépticos (Monderer et al., 2002; Staley, 2015). Además, daños en el cerebelo pueden resultar en temblores, subrayando el papel crucial de la retroalimentación en el sistema nervioso (Wiener, 1961). Esto permite inferir que la comunicación entre neuronas es

cuantificable y se basa en diferencias de potencial que son influenciadas por varios factores, como se mencionó en el Capítulo 1, y estas diferencias pueden ser comparadas con la conducción pasiva de calor a través de una membrana celular. Sabemos por ejemplo que en la epilepsia, las alteraciones en la mieloarquitectura pueden afectar la velocidad y la sincronización de la comunicación neuronal, conduciendo a las descargas excesivas asociadas con las convulsiones.

4.5 Perceptrón

Las redes neuronales son sistemas matemáticos que imitan la funcionalidad del cerebro para aprender y adaptarse en la resolución de problemas. Su estudio proviene del análisis de neuronas biológicas, como el realizado por Hodgkin y Huxley en 1952, quienes descubrieron cómo los iones fluyen a través de los axones (Hodgkin & Huxley, 1952a, 1952b; Rinzel, 1990). Esta investigación biológica, junto con los fundamentos de la computación digital establecidos por McCulloch y Pitts en 1943, inspirando el desarrollo de la computadora digital moderna, o el cerebro electrónico, como lo llamó John von Neumann (Von Neumann, 1993), y la conceptualización del "perceptrón" por Frank Rosenblatt en 1958, han sido esenciales para el desarrollo de las redes neuronales modernas.

Figura 11.
Modelo McCulloch-Pitts de una neurona.



Nota. Representación del modelo McCulloch-Pitts, donde y representa los valores de entrada de información, w los valores de pesos, a es el valor de salida, u es el valor relacionado a los valores de entrada. Dentro del nodo tenemos la función de la red, que termina como la información de entrada se combina dentro de este modelo de neurona. Modificado de *Introduction to Neural Networks for Signal Processing*. En *Handbook of neural network signal processing*. por Hu, Y. H., & Hwang, J. N., 2002

Actualmente, las redes neuronales artificiales han evolucionado hasta convertirse en herramientas no lineales autónomas, útiles en múltiples aplicaciones, especialmente en el procesamiento de señales. Comprender su estructura es crucial, diferenciando entre diseños establecidos y aquellos en desarrollo, para seleccionar la red adecuada que optimice rendimiento y costo. En su funcionamiento, las redes combinan entradas usando pesos sinápticos y sesgos que determinan el umbral de activación, a través de la función de red. Luego, la función activadora, como las sigmoideas o tangente hiperbólica, decide la activación de las neuronas. La selección de la estructura y técnica adecuada es clave para garantizar un rendimiento fiable. Frank Rosenblatt revolucionó la inteligencia artificial en la década de 1950 al concebir el perceptrón, un modelo simplificado de neurona biológica (Rosenblatt, 1957, 1958). Este modelo incluye receptores de entrada, comparables a las dendritas, y utiliza términos como "retina" para la capa de entrada y "área de asociación" para la salida, subrayando su analogía con la estructura cerebral. El perceptrón, reconocido como el modelo más simple de redes neuronales artificiales, puede considerarse hoy una red de neuronas tipo McCulloch-Pitts con funciones de activación no lineales y diferenciables, como las funciones sigmoideas y tangentes hiperbólicas, además de otras modernas como ReLU, Softmax y GELU. Estas neuronas están estructuradas en capas, incluyendo "capas ocultas", así llamadas porque su procesamiento interno no se observa desde fuera, y una capa de salida que presenta el resultado final. Sin embargo, el modelo original del perceptrón tenía limitaciones significativas en la capacidad de procesamiento lógico, como reconocieron Cowan y otros en 1990, evidenciando su incapacidad para ejecutar ciertas operaciones lógicas, problema que persistió incluso con modelos multicapa. Pero esta problemática fue prácticamente resuelta por Rosenblatt (1961), con el añadido de unidades ocultas con conexiones modificables, que añadían un dinamismo, el cual posibilita el cambio de los pesos sinápticos entre los nodos neuronales. Y la estructura básica del perceptrón, al menos para el caso que usa el modelo de McCulloch-Pitts la función de red sería

$$u = \sum_{j=1}^N w_j y_j + \theta, \quad (4.5.1)$$

En el contexto del perceptrón, la función de red es esencial para el funcionamiento del mismo, ya que permite la combinación de entradas, la incorporación de pesos, la preparación para la activación y el aprendizaje. La función de activación, está dependerá del tipo de red usada, pero usualmente se usa la función de activación sigmoideal que es

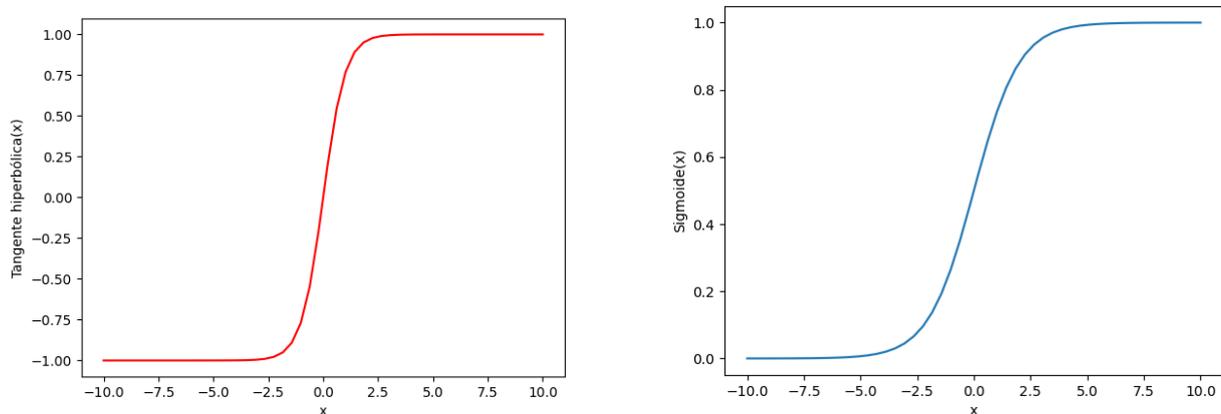
$$f(u) = \frac{1}{1+e^{-u}}, \quad (4.5.2)$$

La función de activación es aquella que introduce la no linealidad, tomar decisiones de activación, permitir la diferenciación y normalizar las salidas en un perceptrón. También existen otros parámetros de activación, tal como la tangente hiperbólica, que a simple vista se parece bastante a la función sigmoide

$$f(u) = \frac{2}{1 + e^{-2x}} - 1, \quad (4.5.3)$$

Y no solo se parecen, ambas enfrentan un problema de saturación de gradiente, ya que después de algunas épocas (epochs) de entrenamiento, la parte lineal de cada neurona tendrá valores que son demasiado grandes o pequeños, esto causa que la actualización de los valores de pesos sea muy lenta, y en algunos casos cause un error de *overflow* o desbordamiento (es un tipo de error que se produce cuando un programa recibe un número, valor o variable fuera del alcance de su capacidad de manejo).

Figura 12.
Funciones de activación



Nota. En la izquierda se encuentra la representación gráfica de la función tangente hiperbólica, y en la derecha la función sigmoide. Se puede visualizar mejor cómo es que se pueden agrupar los valores en los extremos superiores e inferiores, causando la saturación de gradientes.

Luego tenemos la inicialización de una red neuronal, la cual es esencial para evitar problemas de desvanecimiento o explosión de gradientes, romper la simetría, acelerar la convergencia y mejorar el rendimiento del modelo. El cómo ocurre la inicialización de una red neuronal cambia dependiendo de los valores de entrada, el tipo de no-linealidad y de la topología de la red usada. Por lo que podemos considerar los valores de pesos al inicio al menos como independientes de la tarea a la cual se va a dedicar la red neuronal. Tenemos por ejemplo los análisis de Glorot y Bengio (2010). Que para una función lineal de activación el valor óptimo de es

$$W^{(l)} = \sqrt{\frac{1}{N}}, \quad (4.5.4)$$

Donde N es el número de nodos que alimentan a esa capa. Y la razón de usar este tipo de inicialización es por que busca que de los pesos, la varianza de las activaciones sea la misma en todas las capas. Pero este tipo de inicialización no funciona tan bien cuando hacemos uso de la función de activación ReLU (la cual se describe más adelante), y esto se debe a que la media de la función ya no es cero. En otras palabras, la inicialización de Xavier asume una

derivada de 1 para la función de activación que es hasta cierto punto razonable para una tangente hiperbólica, añadiendo un factor de escala para poder corregir este problema de moverse la media lejos del cero de los valores esperados

$$W^{(l)} = \sqrt{\frac{2}{N}} . \quad (4.5.5)$$

4.6 Redes neuronales profundas

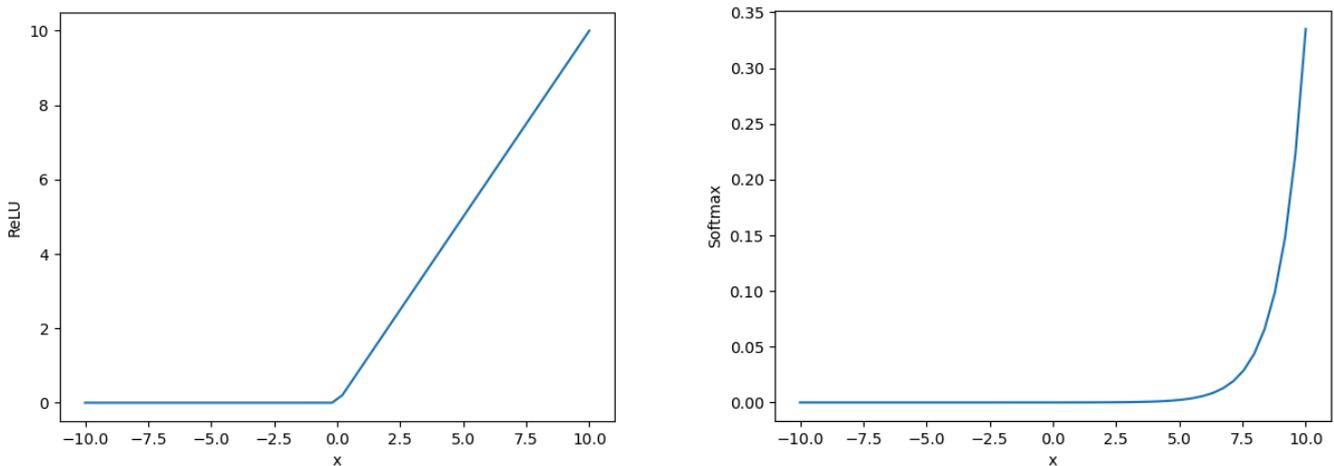
Las máquinas originalmente enfrentaban desafíos para procesar información en su forma original, requiriendo transformaciones para hacerla comprensible (LeCun, Bengio & Hinton, 2015). Las primeras redes neuronales, como el perceptrón, aunque prometedoras, tenían limitaciones en su capacidad de clasificación debido a restricciones como la cantidad de nodos y capas (Cowan, 1990; Rosenblatt, 1958). A pesar de que añadir más "capas escondidas" podría haber superado algunas de estas limitaciones, la tecnología de la época no lo permitía. En el contexto de las redes neuronales profundas, la selección de hiperparámetros, la regularización y la estructura adecuada de la red son aspectos complejos, a menudo considerados un arte dominado solo por expertos (LeCun, Bengio & Hinton, 2015). Esta complejidad a menudo resulta en la elección de parámetros subóptimos, y aunque existen metodologías para optimizarlos, encontrar los valores ideales puede ser una tarea extensa y, en ocasiones, inalcanzable (Smith, 2018; LeCun, Bengio & Hinton, 2015).

En cuanto a la inicialización de parámetros, las funciones de activación sigmoideas a menudo enfrentan problemas, principalmente debido a la saturación de gradientes que afecta la actualización de pesos en la red. Por eso, se prefiere la función de activación ReLU (Rectified Linear Unit), ya que es más eficiente computacionalmente y evita problemas presentes en otras funciones no lineales

$$R(u) = \max(0, x), \tag{4.6.1}$$

Que como puede verse, la función devuelve un valor de 0 si el valor de x (valor de entrada) es menor que 0, y devuelve un valor u si la x (valor de entrada) es mayor que 0.

Figura 13.
Funciones de activación



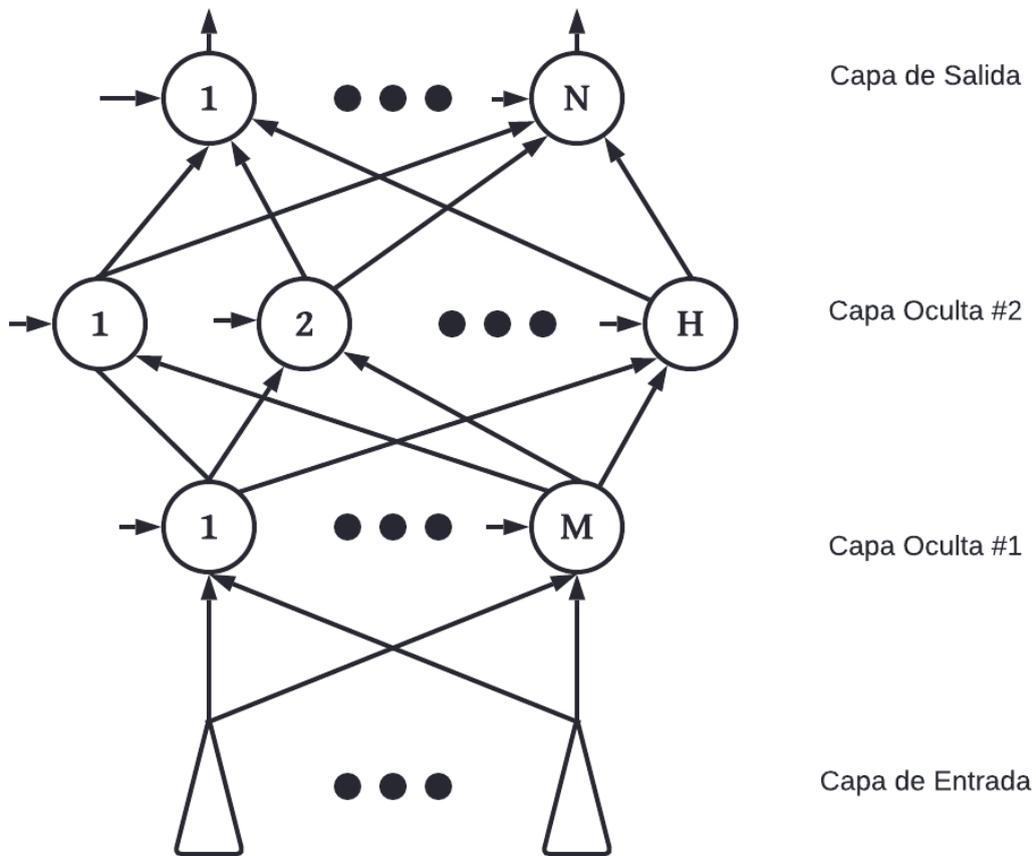
Nota. En la izquierda se encuentra la representación gráfica de la función ReLU, y en la derecha la función Softmax.

También está la función softmax, esta se utiliza a menudo como la última función de activación de una red neuronal para normalizar la salida de una red a una distribución de probabilidad sobre las clases de salida predichas

$$\hat{y}(x)_i = \frac{e^{z_k}}{\sum e^{z_j}}, \quad (4.6.2)$$

Esta función toma como valor de entrada un vector x de números reales, que los normaliza dentro de una distribución de probabilidad constando de una cantidad Z de probabilidades proporcionales a los exponentes de los valores de entrada. Esta elección en cambiar la función de activación se basa en el axioma de elección de Luce, que puede resumirse como que la probabilidad de seleccionar un elemento sobre otro de un conjunto de muchos elementos no se ve afectada por la presencia o ausencia de otros elementos en el conjunto (Luce, 2012).

Figura 14.
Perceptrón Multicapa



Nota. Diagrama de un MPL (*Multilayer Perceptron*) con 3 capas ocultas.

Para el perceptrón multicapa (MLP, *Multilayer Perceptron*), que es una variante del perceptrón, la función de la red $u(x)$ es la suma de los pesos de los valores de entrada

$$u(x) = w_0 + \sum_{i=1}^n w_i x_i, \quad (4.6.3)$$

y los valores de salida $y(x)$ se obtienen mediante la función de activación

$$y(x) = \begin{cases} 1 & u(x) \geq 0 \\ 0 & u(x) < 0 \end{cases}, \quad (4.6.4)$$

Los MLP (Perceptrones Multicapa) se estructuran en capas, donde cada capa interactúa solo con la siguiente. Las denominadas "capas ocultas" tienen este nombre porque sus salidas se envían directamente a la capa posterior, sin que el usuario pueda visualizarlas. En cuanto a la retropropagación, un concepto central en el aprendizaje de redes neuronales, podemos visualizarlo como un proceso de ajuste de pesos utilizando gradientes (Hu & Hwang, 2002). Imagina el gradiente descendente, g , como una dirección y el ritmo de aprendizaje como la distancia que recorres en esa dirección.

Una ecuación clave para ajustar los pesos es

$$\Delta W(t) = - \eta dE \Big| dW, \quad (4.6.5)$$

Donde la cantidad escalar E con respecto a los valores de pesos individuales puede ser computada, por lo que z es el valor de salida y d es el valor deseado. A partir de estos cálculos, obtenemos una fórmula que nos dice cuánto cambiar los pesos

$$\frac{\partial E}{\partial w_i} = - 2 \sum_{k=1}^k \delta(k) x_i(k), \quad (4.6.6)$$

Donde $\delta(k)$ es la señal de error $e(k) = d(k) - z(k)$, que es modulada por la derivada de la función $f'(u(k))$, representando así la cantidad correcta de corrección que se necesita aplicar a los pesos w_i dado el valor de entrada $x_i(k)$. Con esta fórmula, podemos actualizar los pesos de la red

$$w_i(t + 1) = w_i(t) + \eta \sum_{k=1}^k \delta(k) x_i(k). \quad (4.6.7)$$

Este proceso de actualización se repite múltiples veces, y cada ciclo completo de actualización se denomina "epoch" (Hu & Hwang, 2002).

La optimización es esencial para el proceso de actualización de los pesos en las redes neuronales. Una de las técnicas más destacadas es el optimizador Adam, introducido por Diederik Kingma y Jimmy Ba en 2014. Observaron que los métodos convencionales solían ser lentos en convergencia y propusieron Adam para abordar este problema. Básicamente, Adam mejora la velocidad y precisión de la optimización ajustando las tasas de aprendizaje y utilizando un concepto conocido como "momentum" para suavizar las actualizaciones (Kingma & Ba, 2014).

En esencia, Adam trabaja con una combinación de la media móvil de gradientes recientes y su magnitud. Estos dos componentes ayudan a determinar la tasa de aprendizaje adaptativa para cada parámetro de la red. Para hacerlo más intuitivo, piensa en Adam como un conductor de auto que ajusta su velocidad basándose tanto en su velocidad reciente como en la topografía del camino que se aproxima. El proceso comienza con valores iniciales de cero para estimaciones del "primer" y "segundo momento" de los gradientes. El "primer momento" puede entenderse como una media móvil del gradiente y el "segundo momento" como una media móvil de la magnitud del gradiente. Estos se actualizan en cada paso basados en los gradientes actuales y los valores previos, con ciertos factores de decaimiento que modulan la influencia de gradientes pasados.

Un aspecto crucial es corregir el sesgo en estas estimaciones. Dado que comienzan en cero, se les considera "sesgadas" y necesitan ajuste. Una vez corregidos estos valores, se usan para determinar cuánto ajustar cada peso en la red. Aquí, la tasa de aprendizaje, junto con algunos pequeños ajustes para la estabilidad numérica, juega un papel esencial.

Capítulo 5: Metodología

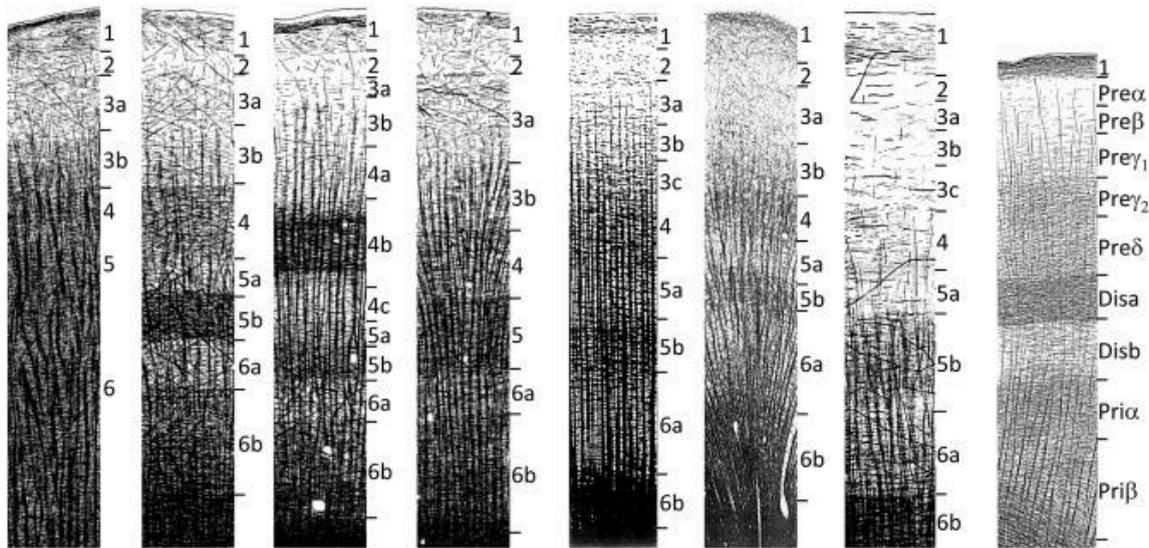
5.1 Planteamiento del problema

La relación entre los pesos en una red neuronal y su rendimiento es una cuestión compleja. Mientras que los pesos no son hiperparámetros en sí mismos, los hiperparámetros pueden influir en cómo se ajustan esos pesos. Ajustar la tasa de aprendizaje, por ejemplo, puede acelerar el ritmo al que los pesos cambian. Esto podría llevar a un aprendizaje más rápido, pero también corre el riesgo de omitir la solución óptima. Del mismo modo, la regularización puede influir en la magnitud y distribución de los pesos, afectando la capacidad de la red para generalizar (evitando el sobreajuste). Así que, aunque los pesos no son hiperparámetros per se, ciertamente son cruciales para el rendimiento de la red.

Hay un cierto misterio en torno a cómo funcionan exactamente los pesos. A pesar de la familiaridad de los expertos con sus redes neuronales, la verdadera naturaleza de cómo los pesos influyen en el resultado final es a menudo esquiva. A veces, la arquitectura de la red se prioriza sobre la inicialización de los pesos. En biología, la distribución de la mielina en el cerebro es específica y sigue patrones que reflejan la conectividad de los circuitos neuronales. Esta mielina tiene diferentes densidades en distintas zonas del cerebro. En analogía con las redes neuronales artificiales, los pesos pueden verse como representaciones de la fuerza de las conexiones entre diferentes partes de una red. Así, uno podría argumentar que hay una correspondencia entre áreas del cerebro con alta mielinización y conexiones fuertes en una red artificial.

Pero en este caso, se tomaron como referencia láminas de distintas áreas de un encéfalo, para poder modificar los valores de inicialización de los pesos de cada una de las capas de la red neuronal profunda con respecto a distribuciones de mielina extraídas de diferentes láminas. Esto para comprobar la posibilidad de la implementación de las densidades topográficas de la mieloarquitectura de diferentes áreas isocorticales y una de la allocorteza para poder evaluar el proceso de aprendizaje en estas distintas áreas usando una arquitectura de red neuronal artificial conocida y poder observar qué valores pueden acoplarse mejor a está misma.

Figura 15.
Cortes laminares de Mieloarquitectura.



Nota. Ilustraciones de los patrones laminares de la mieloarquitectura, en orden de izquierda a derecha; corteza motora primaria, corteza somatosensorial primaria, corteza visual primaria, corteza visual secundaria, corteza de asociación multimodal de la circunvolución frontal inferior, corteza de asociación multimodal de la circunvolución frontal inferior, corteza de asociación multimodal de la lóbulo parietal, corteza límbica de la circunvolución del cíngulo y corteza entorrinal. Modificado de Cortical layers: Cyto-, myelo-, receptor-and synaptic architecture in human cortical áreas. En *Neuroimage*, vol. 197, (pp. 716-741).

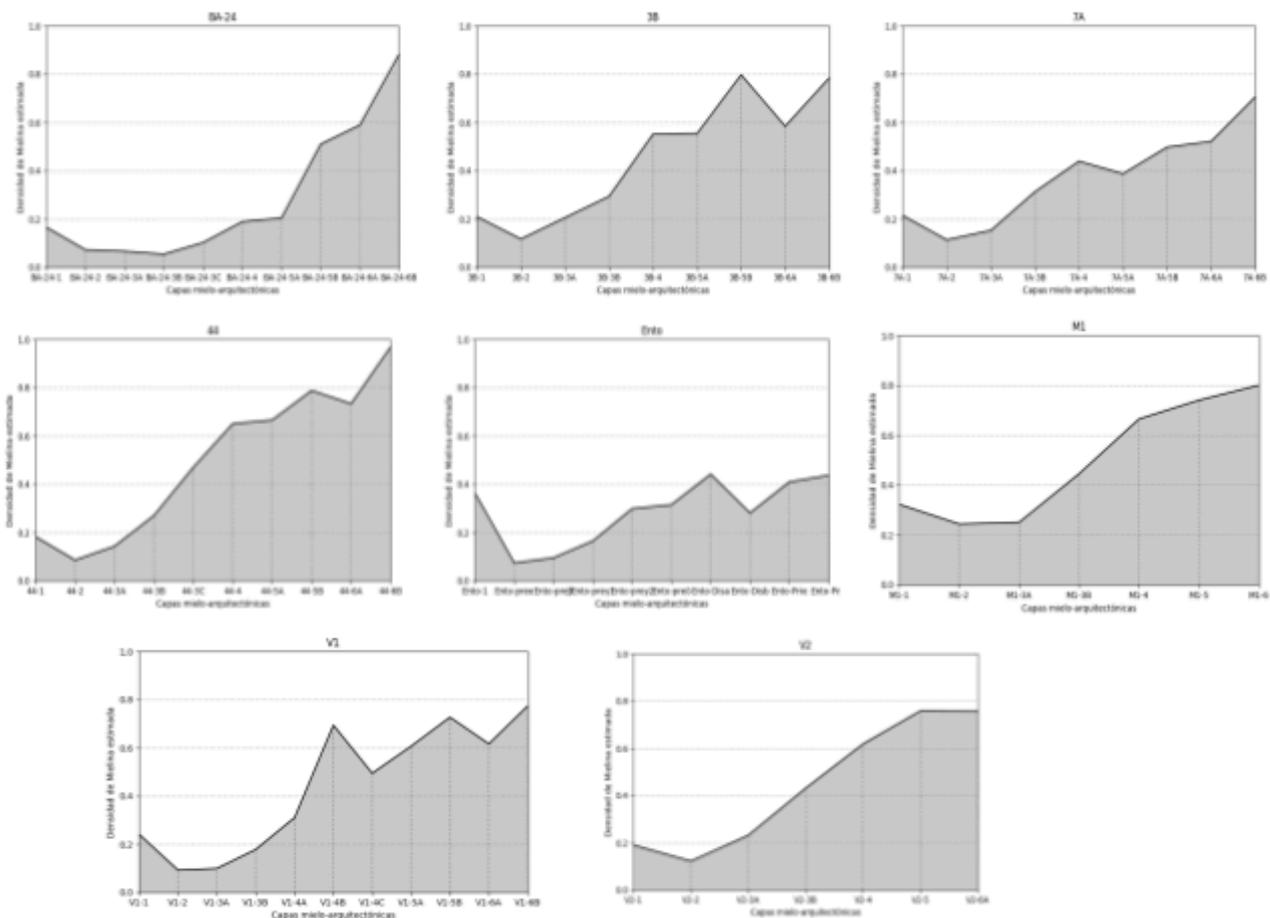
La integración de la neurofisiología con las redes neuronales artificiales podría ofrecer una nueva perspectiva sobre cómo abordar y utilizar hiperparámetros. No sólo podría mejorar el rendimiento y la velocidad de aprendizaje de estas redes, sino que también podría ofrecer a la neurofisiología una herramienta valiosa para realizar experimentos que arrojen luz sobre la topografía neuronal y su funcionalidad en las áreas corticales, proporcionando un método replicable y claro.

Un aspecto clave de este enfoque es la selección de áreas corticales específicas para estudiar. Estas áreas fueron escogidas porque desempeñan roles particulares y distintivos en el cerebro. Además, se incluyó una lámina de la corteza entorrinal, que es notablemente diferente al isocortex debido a su mayor número de capas. Esta inclusión ayudará a determinar si un mayor número de capas influye en la capacidad de aprendizaje de una red neuronal.

Para guiar este estudio, se utilizaron ilustraciones detalladas de patrones laminares que muestran la arquitectura de la mielina en áreas específicas del cerebro, incluyendo áreas

como la corteza motora primaria, somatosensorial primaria, visual primaria y secundaria, entre otras. Estas ilustraciones usadas son el trabajo de Sgonina y el matrimonio Vogt compuesto por Cécile Vogt-Mugnier y Oskar Vogt (Sgonina, 1938; Vogt, 1911; Vogt & Vogt, 1919), las cuales proporcionan una base sólida para explorar la relación entre la estructura neuronal y la funcionalidad de la red.

Figura 16.
Gráficas del estimado de densidad de fibras mielinizadas.



Nota. Gráficos de los valores estimados de mielina (valor porcentual estimado) usando un histograma de iluminación para diferenciar las áreas oscurecidas de las fibras mielinizadas de las áreas ausentes de las mismas de las áreas 24 (B24), 3B (BA3), 7A (BA7), 44 (BA44), Ento (área $\lambda 15d$), M1 (BA4), V1(BA17) y V2 (BA18).

La investigación se enfocó en analizar láminas corticales específicas, descomponiéndolas en sus capas individuales y estudiando sus histogramas de iluminación. Gracias al tintado de estas láminas, se logró un alto contraste y resolución, permitiendo identificar con precisión los detalles topográficos formados por fibras mielinizadas. La elección de estudiar la densidad de

mielina radica en sus propiedades inhibitorias sobre el crecimiento axonal (McGee & Strittmatter, 2003). En términos simples, una mayor densidad de mielina podría reducir las conexiones neuronales, pero también podría mejorar la conectividad al aumentar la velocidad de las conexiones, ya que se cree que los pesos sinápticos influyen en la probabilidad de conexiones nodales.

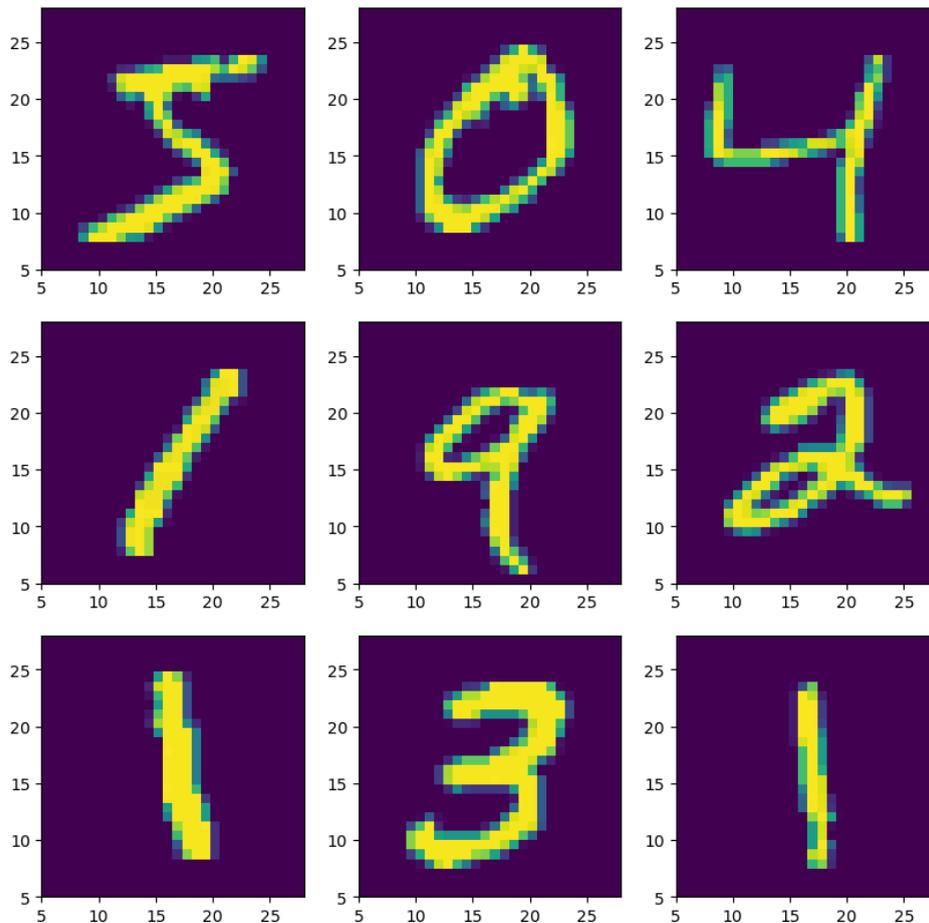
La metodología empleada se inspiró en investigaciones anteriores, específicamente en los trabajos de Braitenberg (1962) y Hopf (1954, 1968, 1968b, 1969). Estos investigadores utilizaron análisis fotométricos para examinar la mielinización en diversas áreas corticales.

La hipótesis central es que las variaciones en la densidad de mielina entre las distintas capas corticales podrían actuar como un factor que evite la simetría en el procesamiento de información en redes neuronales artificiales. Es decir, las diferentes densidades podrían favorecer un rendimiento más óptimo que valores generados aleatoriamente. Considerando que la mieloarquitectura es fundamental para el cerebro: facilita la conducción veloz de impulsos eléctricos mediante la conducción saltatoria, protege los axones, optimiza la energía necesaria para la transmisión de señales, coordina la actividad neuronal para funciones cognitivas complejas, y soporta la plasticidad neural que permite al cerebro adaptarse y aprender.

5.2 Procedimiento

Se utilizó un Perceptrón Multicapa ya que es la forma más simple de una red neuronal profunda. La arquitectura de esta red neuronal está compuesta de ocho capas, el cual fue inicializado integrando los valores de densidades aproximados de mielina obtenidos de los histogramas, exceptuando la capa de entrada y la de salida las cuales fueron inicializadas con el valor de 0.001, ya que al atenuar la influencia de las capas de entrada y salida, las capas ocultas y su inicialización basada en la densidad de la mielina serían más influyentes en el comportamiento inicial del modelo. Las primeras 7 capas estaban formadas por 100 nodos cada una, y la última con 10 para los valores de salida. Se usó la función ReLU de activación para mejorar no solo la eficiencia del procesamiento de información y evitar la saturación de gradientes, así como también reducir el tiempo de evaluación en las primeras siete capas, y en la capa de salida la función softmax y como función de retropropagación la entropía cruzada.

Figura 17.
Números de MNIST

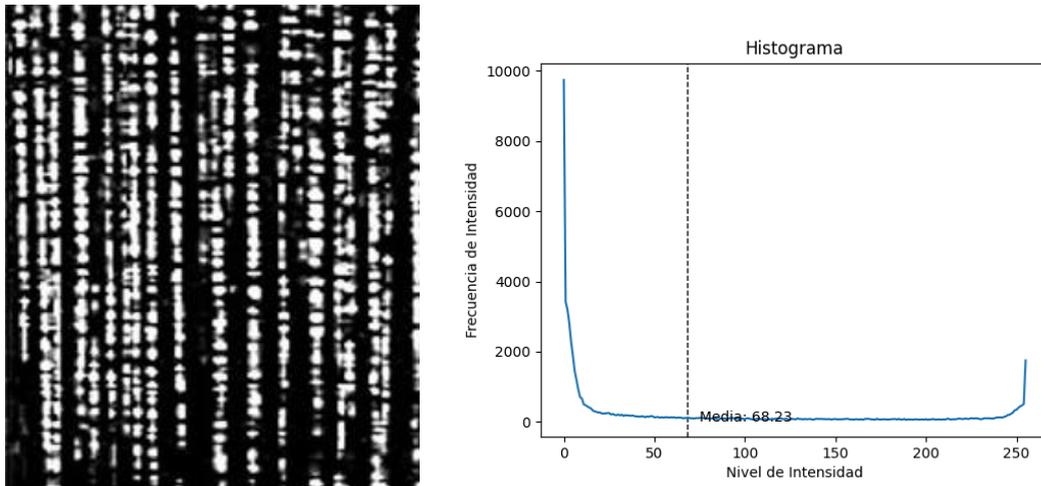


Nota. Muestra de dígitos de la base de datos MNIST con proporción de 28x28 pixeles.

A pesar de las diferencias en las densidades de neuronas reales entre las diferentes áreas corticales, se eligió este enfoque para eliminar posibles variaciones de la citoarquitectura y enfocarnos en el efecto de la densidad de mielina en las capacidades del PM.

La base de datos usada para entrenar a la red neuronal, dada la importancia histórica y facilidad de obtener, se usó la base de datos MNIST (fig. 18), compuesta de dígitos escritos a mano. Esta base de datos fue creada de la base de datos original NIST, normalizando estas imágenes a una proporción de 28x28 píxeles y con un suavizado espacial para minimizar la distorsión artificial cuando se representa una imagen de alta resolución a una menor.

Figura 18.
Histograma de una de las capas corticales.



Nota. Ejemplo de los histogramas hechos de los cortes de las láminas, tomando en cuenta la división de mieloarquitectura. A la izquierda está un corte del área de brodmann 44, del área de mieloarquitectura 6a, y a la derecha la evaluación del nivel de intensidad y la frecuencia de la iluminación del corte.

Las densidades de mielina de áreas de Brodmann y la área entorrinal $\lambda 15d$ se calcularon mediante un histograma de iluminación de láminas corticales, dividiendo las áreas mieloarquitectónicas y empleando los promedios para inicializar las capas neuronales correspondientes. Las densidades neocorticales y archicorticales se integraron de manera distinta en la red debido a su variación estructural. En la configuración inicial del perceptrón multicapa, se seleccionó un batch size de 1024 y un learning rate de $1e-4$ para 400 epochs con la base de datos MNIST. Dado que no se apreciaron mejoras significativas más allá de este punto, se estableció aquí el límite del entrenamiento.

Para acelerar el aprendizaje en pruebas posteriores, se ajustó la inicialización de las capas y se usó un learning rate de $1e-3$ con el optimizador Adam, incluyendo un regularizador que integraba densidades de mielina. Para controlar la variabilidad estocástica, se fijó la semilla en 42 para NumPy y TensorFlow, garantizando y facilitando la reproducibilidad del experimento (Tanksley & Wunsch II, 2019). Se incluyó una estimación de la entropía y negentropía para analizar los efectos estructurales del entrenamiento en las matrices de pesos, utilizando una descomposición de valores singulares para medir la entropía. Finalmente, se entrenó la red con un batch size de 128 durante 10 epochs.

Para evaluar cómo las alteraciones en los valores de inicialización influyen en el entrenamiento del Perceptrón Multicapa (PM), se estimó la entropía y la negentropía de las matrices de pesos. La entropía fue aproximada utilizando un método basado en la descomposición de valores singulares, específicamente, la entropía de la distribución de valores singulares que refleja la información captada por cada uno.

$$H = \sum_{i=1}^N \frac{1}{1-\alpha} \log_2 \left(\sum_{j=1}^k \left(\frac{\sigma_j^i}{\sum_{l=1}^k \sigma_l^i} \right)^\alpha \right), \quad (5.2.1)$$

Donde H es la entropía, l es el número de capas densas en el modelo, k es el número de valores singulares en cada capa i , y σ_{ij} es el valor singular j de la matriz de pesos para la capa i , y α es un parámetro que determina la sensibilidad a los valores singulares pequeños. Esta ecuación usada es una variante de la entropía de Rényi, que es una familia de mediciones de entropía introducidas por Alfred Rényi (1960, 1961), y que en este caso en particular el valor de α usado fue de 1.001. La fórmula empleada es una adaptación de la entropía de Rényi, una familia de mediciones de entropía propuestas por Alfred Rényi, utilizando un valor de α de 1.001. Esta adaptación es útil especialmente cuando los valores de inicialización son muy pequeños, evitando así resultados negativos en la entropía. Un

aumento en la entropía tras el entrenamiento indica un PM más complejo y viceversa; sin embargo, un valor más bajo podría significar mejor generalización y menos sobreajuste.

Para la negentropía, que estima el orden en la distribución de pesos, primero se normalizan los valores singulares y se calcula su entropía de Shannon. La negentropía se obtiene restando la entropía actual de la máxima posible, indicando la capacidad de la red de representar los datos de entrada de forma compacta. Un incremento en la negentropía durante el entrenamiento suele ser deseable, aunque no es un indicativo directo del rendimiento

$$I(M) = \log_2(n) + \sum_i \left(\frac{\sigma_i}{\sum_i \sigma_j} * \log_2 \left(\frac{\sigma_i}{\sum_i \sigma_j} + 10^{-10} \right) \right) \quad (5.2.2)$$

Donde σ_i son los valores singulares de la matriz, n es el número de valores singulares y la constante $10e-10$ se agrega para evitar el logaritmo de cero. En esta ecuación, la primera parte $\log_2(n)$ representa la entropía máxima, y la segunda parte representa la entropía de Shannon (con signo negativo), por lo que la suma de ambas da como resultado la negentropía que es una medida del orden o la previsibilidad de los pesos de la red.

Las modificaciones en el segundo PM, incluyendo el inicializador y el proceso de regularización, se ajustaron para evaluar el efecto de las densidades de mielina. Se usó una combinación de regularización L1 y L2, donde la primera minimiza los pesos y la segunda previene su crecimiento excesivo

$$R = l1 \cdot \sum |x * d| + l2 \cdot \sum (x \cdot d)^2 \quad (5.2.3)$$

Donde $l1$ y $l2$ son los parámetros de regularización para las normas L1 y L2, x representa los pesos en la capa, y d es una constante definida por la densidad estimada de mielina de esa capa que multiplica a los pesos. Donde la norma L1 añade una penalización proporcional al valor absoluto de los pesos, y la norma L2 añade una penalización proporcional al cuadrado de los pesos. Esto quiere decir que la primera norma lleva a una solución dispersa y por lo

tanto reduciendo el tamaño de los pesos y la segunda evitando que los pesos crezcan demasiado. La inicialización se basó en la densidad de mielina multiplicada por una distribución uniforme aleatoria. El primer PM tiene un enfoque en la precisión y el costo del entrenamiento con un ritmo de aprendizaje bajo y más épocas, sin añadir optimizador, mientras que el segundo se enfoca en cambios estructurales de las matrices de pesos y requiere menos épocas

$$\textit{Inicialización} = \textit{He Uniforme} \cdot d \quad (5.2.4)$$

Donde los pesos aleatorios *He Uniforme* son una matriz de pesos inicializados con una distribución uniforme. Considerando esto, el primer PM tiene un enfoque en la precisión y el costo del entrenamiento con un ritmo de aprendizaje bajo y más épocas, sin añadir optimizador, mientras que el segundo se enfoca en cambios estructurales de las matrices de pesos y requiere menos épocas.

Finalmente, la eficacia del PM se midió por las métricas de costo y pérdida durante el entrenamiento y la precisión del modelo, con especial interés en el impacto de las modificaciones en los valores de inicialización sobre el rendimiento.

Capítulo 6: Resultados

De las ocho áreas usadas para obtener valores de densidades y probar su eficacia, se encontró que solo una, aquella asociada al área 4 de Brodmann causó un error de *overflow* posiblemente por que terminó generando un valor demasiado grande o pequeño. Para el resto de redes neuronales modificadas se obtuvieron los siguientes valores:

Tabla 2.
Resultados de precisión y costo del entrenamiento y las pruebas posteriores

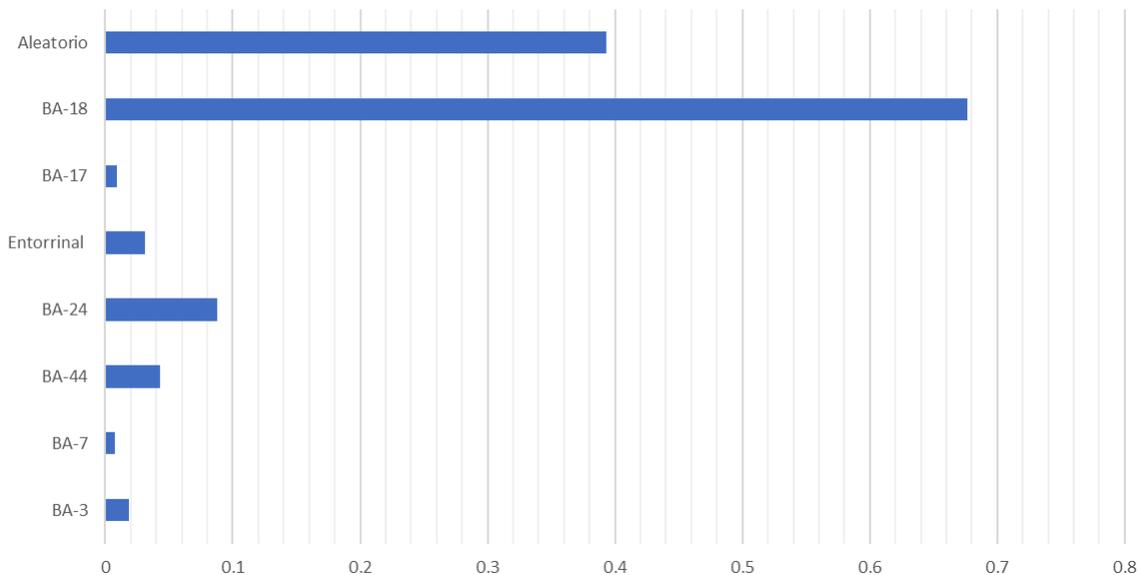
	Precisión	Precisión Entrenamiento	Precisión Prueba	Costo
BA-3	94.74%	99.82%	94.31%	0.018760302
BA-7	95.75%	99.94%	95.39%	0.007331622
BA-44	93.86%	99.07%	93.93%	0.043094539
BA-24	96.64%	97.71%	96.27%	0.087525965
λ 15d Entorrinal	95.59%	99.26%	95.50%	0.030935232
BA-17	95.92%	99.91%	95.56%	0.008944256
BA-18	83.30%	85.08%	81.98%	0.676308609
Aleatorio	91.71%	91.78%	91.11%	0.393382449

Nota. La precisión nos habla de la capacidad del modelo de clasificar las imágenes con el conjunto de datos de validación. (Un conjunto de datos de validación es una muestra de datos retenidos del entrenamiento de su modelo que se utiliza para dar una estimación de la habilidad del modelo mientras se entrena el modelo). La precisión del entrenamiento significa que se utilizan imágenes idénticas tanto para el entrenamiento como para la prueba, mientras que la precisión de la prueba representa que el modelo entrenado identifica imágenes independientes que no se utilizaron en el entrenamiento. El costo es una medida de "lo bien" que lo ha hecho una red neuronal con respecto a su muestra de entrenamiento y la salida esperada.

El área visual secundaria (BA-18) fue la única que obtuvo un resultado subóptimo debajo del 90% de precisión, y aunque al inicio del entrenamiento parecía que podía estabilizarse, no fue el caso. Al contrario, el área visual primaria (BA-17) fue de las mejores en rendimiento, al obtener los mejores resultados en predicción de la imagen que estaba procesando con el menor costo. El área BA-24 que justo se dedica al control motor (es decir, el área motora

cingulada) y la atención obtuvo las mejores predicciones posibles pero con un valor de costo bastante alto. Interesantemente, aquella con valores aleatorizados no obtuvo resultados similares. Sus resultados estuvieron debajo del 3-4% de precisión respecto al resto, y teniendo un costo mucho mayor, exceptuando al área visual secundaria (fig.19).

Figura 19.
Comparativa de costos



Nota. Comparativa de valores de costos después de 400 *epochs*.

Aunque en un principio parecía que el error de *overflow* podría explicarse simplemente por la distribución de los valores, se hizo un pequeño análisis de las características de los valores de densidades estimados de las láminas usados para obtener los siguientes resultados:

Tabla 3.
Análisis de los valores de inicialización

	BA-3	BA-7	BA-44	BA-24	λ_{15d}	BA-4	BA-17	BA-18	Aleatorio
Varianza	0.053	0.029	0.084	0.053	0.012	0.049	0.059	0.068	0.121
Max	0.684	0.613	0.851	0.735	0.423	0.801	0.695	0.758	0.951
Min	0.117	0.113	0.086	0.072	0.085	0.244	0.093	0.122	0.037
Media	0.341	0.296	0.351	0.188	0.265	0.470	0.299	0.374	0.406

Nota. Datos de varianza de los valores de inicialización de las densidades estimadas de mielina.

En primera instancia pareciera no haber algo absolutamente obvio respecto al posible origen del fallo, a excepción posiblemente de los valores de media, ya que justamente los valores aleatorizados, los del área visual secundaria y el área motora primaria tuvieron el peor desempeño teniendo los valores de media mayores.

Tabla 4.
Datos de correlaciones para valores de precisión y valores de medias.

			Precisión	Precisión Prueba	Precisión Entrenamiento	Media
Rho de Spearman	Precisión	Coefficiente de correlación	1.000	0.214	.929**	-0.119
		Sig. (bilateral)		0.610	0.001	0.779
		N	8	8	8	8
	Precisión Prueba	Coefficiente de correlación	0.214	1.000	0.524	-.857**
		Sig. (bilateral)	0.610		0.183	0.007
		N	8	8	8	8
	Precisión Entrenamiento	Coefficiente de correlación	.929**	0.524	1.000	-0.405
		Sig. (bilateral)	0.001	0.183		0.320
		N	8	8	8	8
Media	Coefficiente de correlación	-0.119	-.857**	-0.405	1.000	
	Sig. (bilateral)	0.779	0.007	0.320		
	N	8	8	8	9	

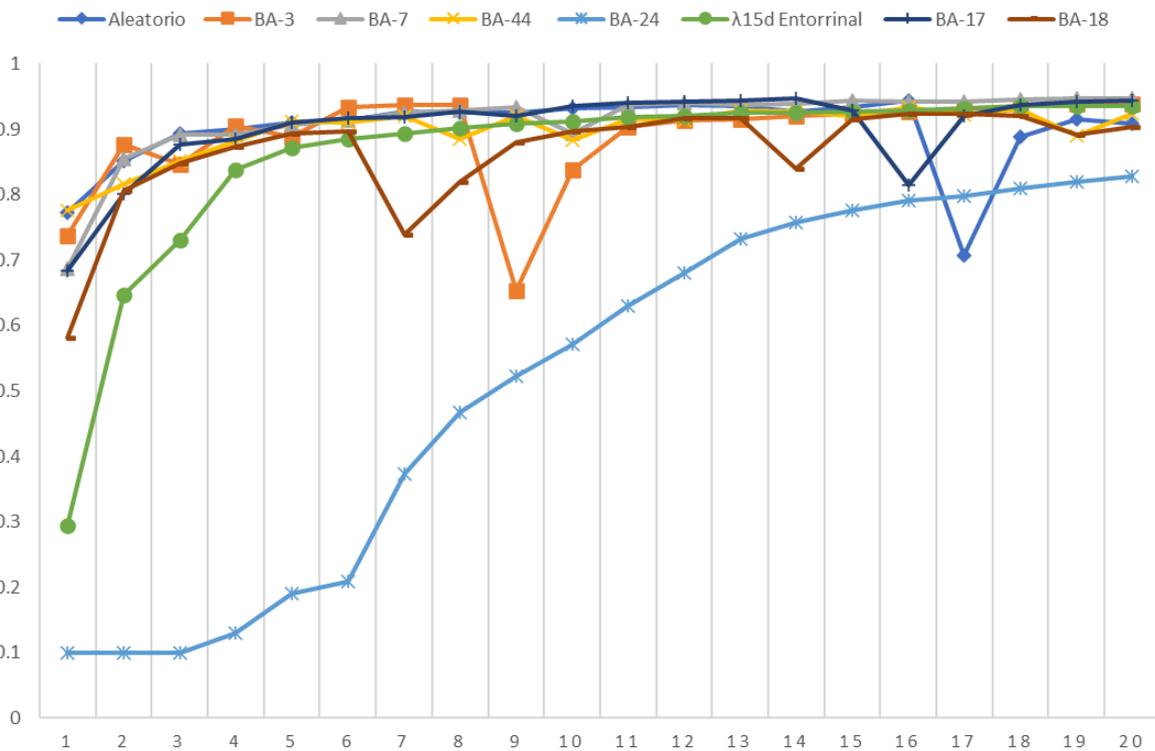
** La correlación es significativa en el nivel 0,01 (bilateral).

Nota. Se puede observar la correlación negativa entre la precisión de prueba y los valores de media, así como correlaciones parciales no significativas del resto de valores de precisión.

Los análisis de correlación revelan una relación inversa entre la media de valores de inicialización y la precisión de prueba, sugiriendo que la media influye considerablemente en alcanzar altas precisiones de entrenamiento. Sin embargo, iniciar la red con valores dentro de un rango estrecho como 0.2 ± 0.1 conduce a un problema de simetría que obstaculiza el aprendizaje, debido a la oscilación de la red. Esta simetría podría estar relacionada con las variaciones observadas en la densidad de mielina entre capas y justificaría el éxito de los valores estimados de mielina como puntos de inicio para las redes neuronales.

El análisis del comportamiento de las redes muestra que en solo 20 epochs, las más eficientes rozan el 95% de precisión en predicciones. Aunque con más epochs sus resultados se estabilizan, la red BA-24, a pesar de predecir mejor imágenes no vistas tras 400 epochs, exhibe una tasa de aprendizaje más lenta y un alto valor de costo, lo que indica que sus valores de inicialización pueden ser efectivos para generalizar, pero no están optimizados para la clasificación de imágenes.

Figura 20.
Gráfica del desarrollo de la precisión de las redes neuronales.

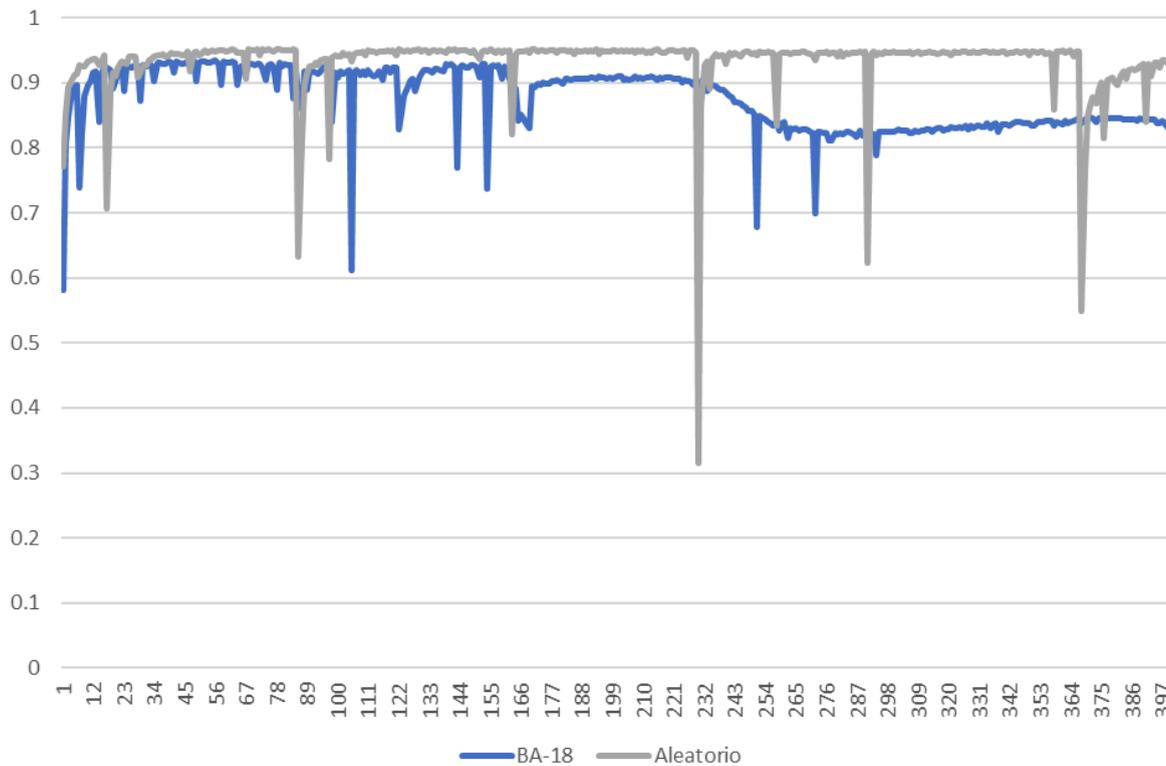


Nota. Gráfica de los resultados de la precisión a lo largo de 20 epochs. Se puede observar como el área BA 24 respecto al resto tiene un aprendizaje más lento.

La naturaleza de la precisión de los valores aleatorizados cambia bastante y termina después de los 300 epochs reduciéndose, y mostrando un comportamiento un poco inestable comparándolo con el resto de los valores producidos por aquellas redes neuronales basadas en las densidades de mielina. Inclusive, cuando se compara con los valores de precisión del área visual secundaria, no se observan cambios tan drásticos, aunque este tenga un valor de media a sus valores de inicialización menor a los valores aleatorizados por lo que podríamos esperar

que tuviera mejores resultados, esto no fue así. Los valores aleatorizados obtuvieron a pesar de ello una mejor respuesta al entrenamiento (fig. 22).

Figura 21.
Gráfica de comparación en la precisión del área visual secundaria y valores aleatorios.



Nota. Gráfica de los resultados de la precisión a lo largo de 400 *epochs*. Se puede observar cómo los cambios de valores más drásticos ocurrieron con los valores aleatorizados.

Posiblemente se hubieran visto mayores diferencias entre ambas redes neuronales si hubieran sido entrenadas por una mayor cantidad de *epochs*. También podemos ver en cierto grado una mayor estabilidad en las redes neuronales con valores estimados de densidad de mielina a largo plazo. Con el segundo grupo de perceptrones multicapa los resultados fueron algo diferentes pero esto ya era esperado debido a la mayor cantidad de cambios y al uso del optimizador Adam. Pero aún así se puede observar como la mayoría de los PM modificados con los valores de mieloarquitectura obtuvieron mejores resultados en en el entrenamiento y en los datos de validación (Tabla 5).

La Tabla 5 muestra que los Perceptrones Multicapa (PM) inicializados con valores de mielina de las áreas de Brodmann 7, 24 y 18 tuvieron una eficiencia comparable a la de aquellos con

valores aleatorios, lo que sugiere una posible resistencia al entrenamiento asociada con estas áreas específicas. Sin embargo, con un número mayor de epochs y un incremento en los nodos por capa, esta resistencia fue superada. Esto indica que los valores de mielina influyen significativamente en el aprendizaje de los modelos. Además, el modelo no modificado presentó un rendimiento inferior, aun cuando se sometió al mismo proceso de regularización y optimización, confirmando el impacto de la mielina en el rendimiento. El empleo de este segundo conjunto de PM tenía el propósito de explorar cómo la manipulación de los valores iniciales de los pesos antes y después del entrenamiento afecta los cambios en la red, proporcionando así una visión clara de la dinámica de los pesos durante el aprendizaje.

Tabla 5.
Resultados de precisión y costo del entrenamiento y las pruebas posteriores

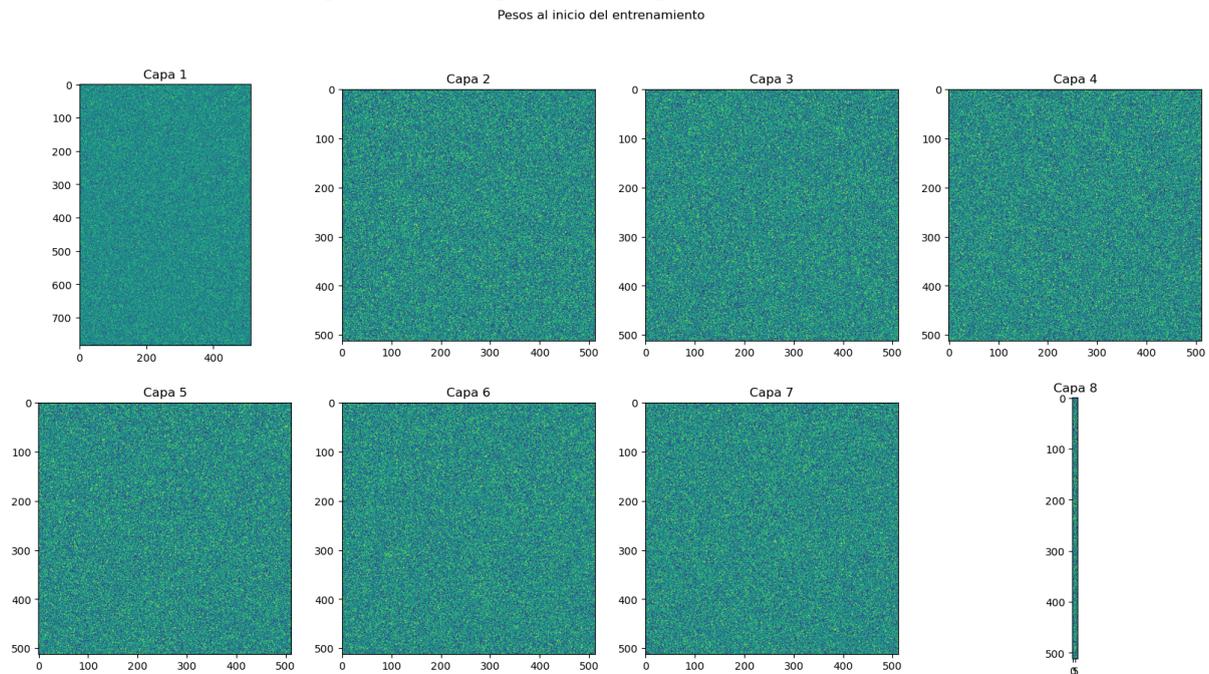
	Costo entrenamiento	Precisión entrenamiento	Costo prueba	Precisión prueba
BA-17	0.0292	99.22%	0.0976	98.19%
BA-4	0.0315	99.19%	0.1005	98.06%
λ 15d Entorrinal	0.031	99.15%	0.1036	98.05%
BA-44	0.0324	99.15%	0.0941	98.05%
BA-3	0.03	99.20%	0.1142	97.80%
BA-7	0.0307	99.21%	0.1192	97.79%
BA-24	0.0366	99.12%	0.1017	97.69%
BA-18	0.0285	99.22%	0.138	97.67%
Aleatorio	0.0269	99.27%	0.1237	97.42%
Sin modificaciones	0.0342	99.15%	0.1154	97.33%

Nota. Se mantuvo el mismo patrón donde el PM que usó los valores del área 17 de brodmann obtuvo los mejores resultados en la precisión de prueba y los mejores valores de precisión con valores bajos de pérdida en ambos casos.

Cómo se logra observar en la figura 22, al inicio no parece tener mayor efecto el multiplicar los valores de inicialización de cada capa de forma independiente por un valor diferente. Pero aunque no es visible al inicio, el proceso de escalar los valores de inicialización de cada capa de forma independiente, cambia el punto de partida del algoritmo de optimización para cada capa. Lo que nos da diferentes rutas de optimización y teóricamente soluciones

potencialmente mejores, ya que cada capa aprende a su propio ritmo y magnitud de manera separada, emulando de esta manera los efectos de la mielinización diferenciada del cerebro humano.

Figura 22.
Visualización de los pesos del Perceptrón Multicapa al inicio del entrenamiento (B-24).



Nota. Representación visual de los valores de los pesos por capas del Perceptrón Multicapa, donde el color amarillo representa un valor más grande y el azul un valor pequeño.

La comparación de los cambios en entropía y negentropía antes y después del entrenamiento ilustra cómo los modelos se transforman y se adecuan a las propiedades mieloarquitectónicas. Un delta de entropía negativo sugiere un incremento en el orden estructural del modelo, mientras que un delta de negentropía positivo indica un aumento en el orden y estructura.

Al evaluar los modelos, se nota que aquel sin modificaciones muestra cambios mínimos en entropía y negentropía, sugiriendo un pequeño cambio estructural en los pesos y una inmunidad a las alteraciones que podrían afectar su estructura y distribución. Por otro lado, la adición de valores aleatorios de mielina no produce efectos tan beneficiosos como el uso de datos reales de un encéfalo, y no conduce a cambios estructurales significativos en los pesos como lo hacen áreas específicas del cerebro. Un cambio sustancial en la entropía podría estar ligado a cómo el modelo se ajusta durante el entrenamiento y, en paralelo, la capacidad de

generalización podría estar conectada a la estructura final de los pesos. Estos cambios estructurales, implicados por los deltas de negentropía y entropía, se pueden visualizar claramente en la figura 23, donde los ajustes en los valores de los pesos son evidentes a simple vista.

Tabla 6.
Datos de entropía y negentropía de los PM.

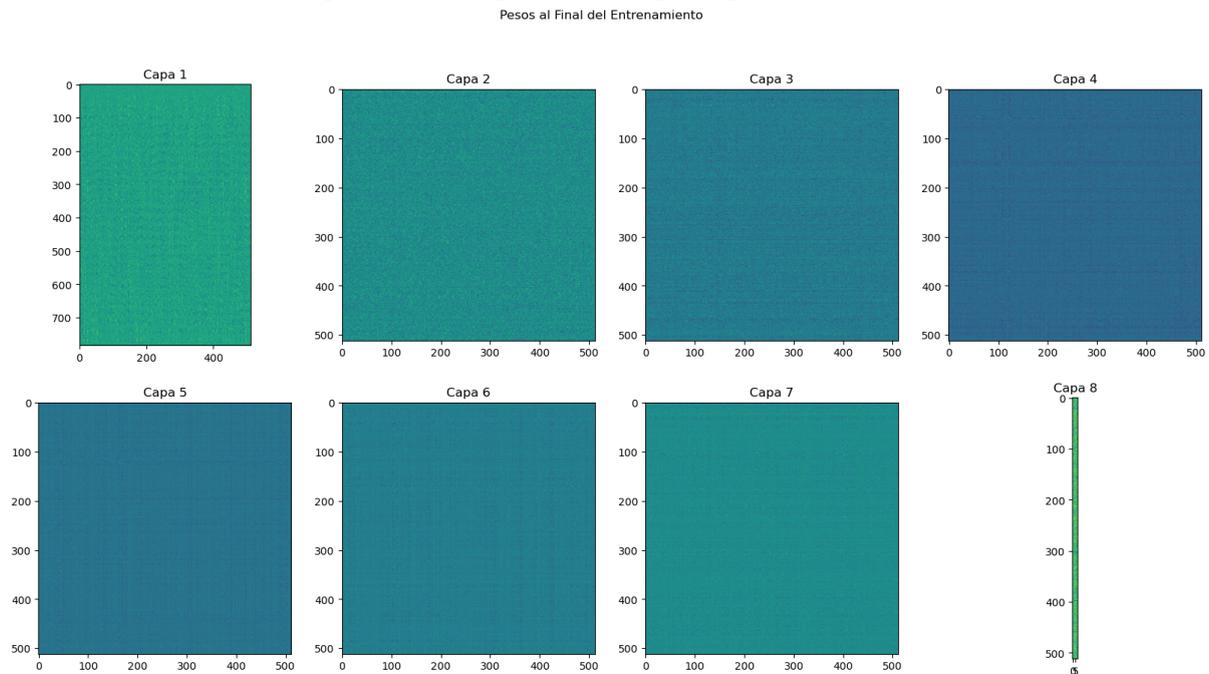
Modelo	Δ Entropía	Δ Negentropía
BA-24	-3.43584199436048	3.43146848678589
λ 15d Entorrinal	-2.27121752686824	2.26847457885742
BA-17	-2.17194366268839	2.16917586326599
BA-7	-2.01734388247155	2.01507949829102
BA-44	-1.86306377872846	1.86076641082764
BA-3	-1.83851248584708	1.8363561630249
BA-18	-1.70907727442701	1.70665717124939
Aleatorio	-1.5447011683138	1.54349637031555
BA-4	-1.18874921463443	1.18751978874207
Sin modificaciones	-0.421484932303493	0.421185970306397

Nota. Deltas de los valores obtenidos de la aproximación al cálculo de la entropía y la neguentropía de los perceptrones multicapa. Se puede observar como el valor de delta es muy pequeño para la entropía y la neguentropía del PM sin modificaciones y el aleatorio, y al mismo tiempo el PM (BA-24) siendo el que tiene el mayor valor de cambio de entropía y negentropía, junto con aquel que uso el área entorrinal y el área 17 de Brodmann.

Es posible que los altos valores de delta de neguentropía y entropía indiquen un significativo cambio estructural en el modelo, reflejando una posible resistencia al cambio. Al analizar los perceptrones multicapa con datos del Área de Brodmann 24, se nota una baja densidad de mielina en las capas II y III, con un impacto considerable en los pesos de las capas 5 y 6. Esto ralentizó el aprendizaje del primer perceptrón y aumentó la resistencia al aprendizaje del segundo. Considerando estos datos, la distribución de pesos en el perceptrón multicapa parece excesivamente densa, potencialmente causando estancamiento durante el entrenamiento. No obstante, los resultados indican que la red fue capaz de aprender y

generalizar nuevos datos eficazmente. Como muestra la figura 24, el modelo puede reconocer y clasificar correctamente datos no vistos previamente, demostrando su capacidad de elección sin ambigüedad.

Figura 23.
Visualización de los pesos del Perceptrón Multicapa después del entrenamiento (B-24).



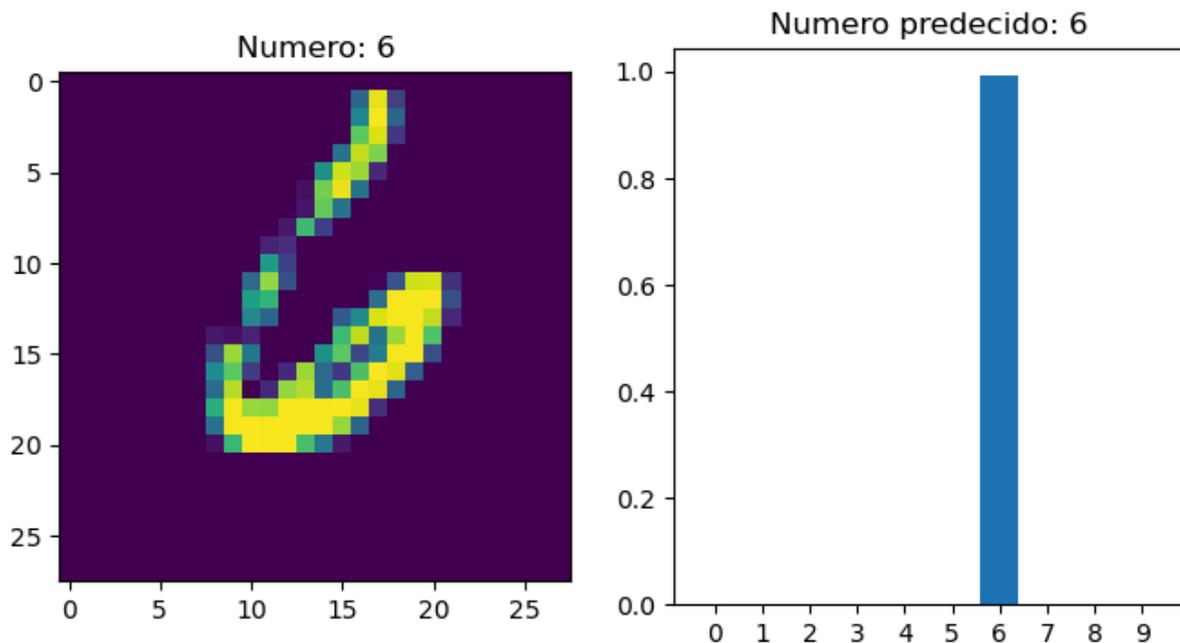
Nota. Representación visual de los valores de los pesos por capas del Perceptrón Multicapa, después del entrenamiento. Se puede observar un cambio claramente visible en las capas 4 y 5, pero estos no ayudaron a mejorar considerablemente las habilidades predictivas del PM (B-24)

Los efectos del entrenamiento en los PM que pudieron de forma más eficiente generalizar los datos en los pesos son más que obvios, alterando la estructura de los pesos y mostrando una distribución más homogénea (e.g. 23). Los valores de los pesos se normalizaron hasta cierto punto en cada capa, algo comparable con las estructuras de mieloarquitectura.

El segundo conjunto de perceptrones multicapa se entrenó usando el optimizador Adam, que integra técnicas de momentum y tasas de aprendizaje adaptativas para mejorar la actualización de los pesos durante el entrenamiento. Esto optimiza la convergencia al superar los problemas de tasas fijas y acelerar ante desafíos como curvas pronunciadas o ruido en los gradientes. Aunque los valores iniciales pueden afectar la rapidez de la convergencia, Adam ajusta los pesos constantemente, incluso cuando cambios en la densidad de mielina pueden alterar estos efectos, los cuales requieren un análisis más profundo. La alta entropía en los

pesos sugiere una distribución uniforme de los valores singulares, mientras que la baja entropía indica una concentración en pocos pesos significativos. La negentropía, por otro lado, refleja que el modelo, a través del entrenamiento, ha organizado la información de entrada en estructuras más complejas, mejorando la representación de la información. Esto se evidencia en los cambios observados en entropía y negentropía, según lo demostrado en la tabla 6.

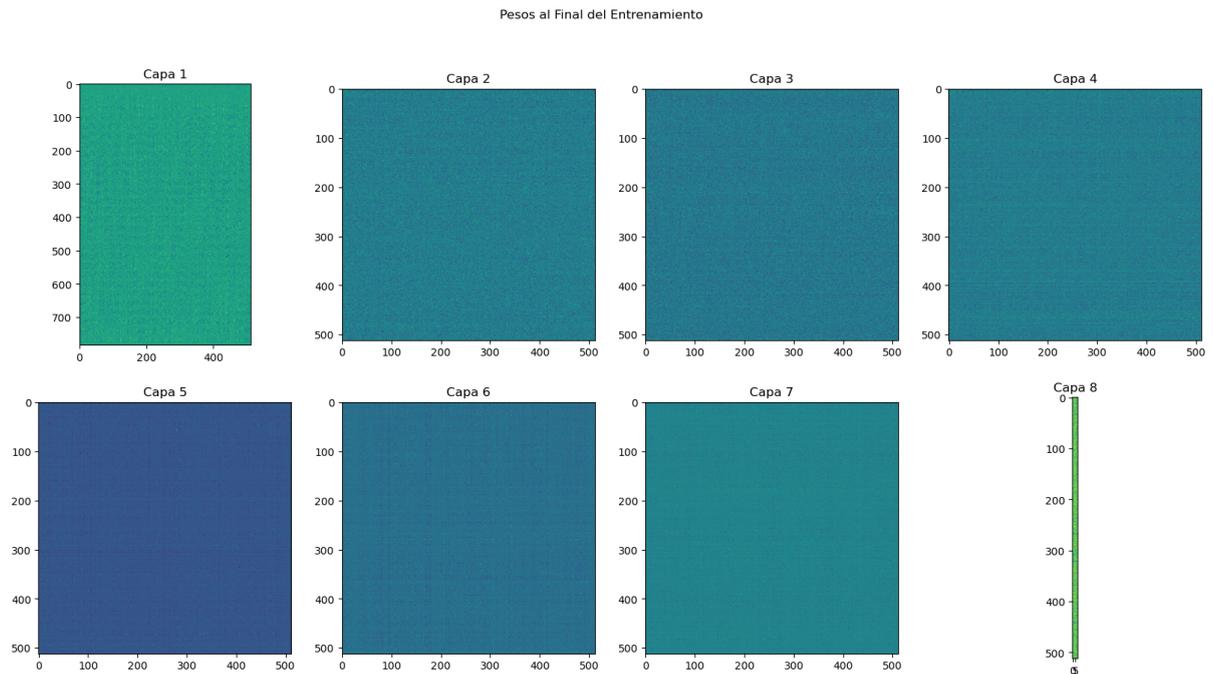
Figura 24.
Prueba del PM (B24) para predecir el número presentado de la base de datos MNIST.



Nota. A la izquierda se puede observar la visualización del número de la base MNIST, y a la derecha la evaluación del PM (B-24) para predecir de qué número se trata. Puede observarse como solo el número 6 tiene un valor muy alto, lo que quiere decir que el PM(B24) logró exitosamente aprender a diferenciar y generalizar adecuadamente la identificación de los números de la base de datos de prueba.

El uso de entropía y negentropía proporciona una comprensión más detallada de los procesos internos del modelo. Se revela que una distribución homogénea de los pesos no necesariamente implica un aprendizaje efectivo, y que la falta de cambios visibles en la distribución de los pesos puede indicar la formación de patrones asociados con el aprendizaje del modelo.

Figura 25.
Visualización de los pesos del Perceptrón Multicapa después del entrenamiento (B-17).



Nota. Representación visual de los valores de los pesos por capas del Perceptrón Multicapa, después del entrenamiento. Aquí se puede observar un cambio bastante grande en las capas 5 y 6, donde el color azul nos dice que tiene valores de pesos muy pequeños.

Capítulo 7: Discusión y Conclusiones

La teoría de la información proporciona una base para comprender las funciones corticales, respaldando el uso de redes neuronales artificiales en la investigación de la conciencia y la memoria. Estos modelos replican la comunicación entre neuronas y cómo la mielinización afecta la señalización. Se ha demostrado que las características de la mieloarquitectura pueden integrarse en redes neuronales artificiales, y se ha propuesto ajustar los nodos según las proporciones de mielina basadas en las áreas de Brodmann para mejorar su eficacia.

A pesar de ser preliminares, los cálculos de entropía y negentropía parecen prometedores para refinar estas redes, aunque son representaciones simplificadas de la complejidad de los perceptrones multicapa. Visualizar los cambios en los pesos de las capas ofrece una comprensión más clara de los procesos subyacentes y del impacto de las condiciones iniciales. Este enfoque, aunque no definitivo, proporciona una visión valiosa de cómo la información se integra en el modelo. Las estructuras de mielina podrían ser claves para mejorar la inicialización de pesos en las redes neuronales artificiales. La formación sinusoidal de las capas de mielina durante la neurogénesis se sugiere como un método de inicialización superior, aunque su aplicación como función de activación ha resultado compleja y con muchas problemáticas en la práctica (Parascandolo et al., 2016).

Es crucial diferenciar entre las redes neuronales artificiales y las biológicas, en particular, en mecanismos como la retropropagación. Aunque hay similitudes en la comunicación entre áreas corticales (Keysers & Perrett, 2004), los procesos biológicos del cerebro difieren de los algoritmos de aprendizaje automático. La variabilidad celular entre las áreas corticales afecta la topología de la red cerebral y puede influir en su comportamiento, considerándose un tipo de hiperparámetro. Finalmente, aunque emular el sistema nervioso central es intrincado, las simulaciones avanzadas podrían proporcionar una comprensión más profunda de las enfermedades mentales, potencialmente guiando hacia nuevos enfoques para su origen y tratamiento.

Referencias

- Amunts, K., & Zilles, K. (2015). Architectonic mapping of the human brain beyond Brodmann. *Neuron*, 88(6), 1086-1107.
- Amunts, K., Schleicher, A., Bürgel, U., Mohlberg, H., Uylings, H. B., & Zilles, K. (1999). Broca's region revisited: cytoarchitecture and intersubject variability. *Journal of Comparative Neurology*, 412(2), 319-341.
- Angevine, J. B., & Sidman, R. L. (1961). Autoradiographic study of cell migration during histogenesis of cerebral cortex in the mouse. *Nature*, 192(4804), 766-768.
- Arbib, M. A. (1975). From automata theory to brain theory. *International Journal of Man-Machine Studies*, 7(3), 279-295.
- Arendt, D., & Nübler-Jung, (1994). Inversion of dorsoventral axis?. *Nature*, 371(6492), 26-26.
- Arendt, D., Musser, J. M., Baker, C. V. H., Bergman, A., Cepko, C., Erwin, D. H., ... & Wagner, G. P. (2016). Evolution of sister cell types by individuation. *Nat Rev Genet*, 17, 744-757.
- Armstrong, R. C., Dorn, H. H., Kufta, C. V., Friedman, E., & Dubois-Dalcq, M. E. (1992). Pre-oligodendrocytes from adult human CNS. *Journal of Neuroscience*, 12(4), 1538-1547.
- Ashby, W. R. (1954). *Design for a Brain*. London: Chapman & Hall.
- Ashby, W. R. (1991). Principles of the self-organizing system. En *Facets of systems science* (pp. 521-536). Springer, Boston, MA.
- Barcia, J. A., Sanz, A., Balugo, P., Alonso-Lera, P., Brin, J. R., Yus, M., ... & Oliviero, A. (2012). High-frequency cortical subdural stimulation enhanced plasticity in surgery of a tumor in Broca's area. *Neuroreport*, 23(5), 304-309.
- Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. *Sensory communication*, 1, 217-234.
- Bateson, W. (1886). Memoirs: the ancestry of the Chordata. *Journal of Cell Science*, 2(104), 535-572.
- Berger, T., & Levy, W. B. (2010). A mathematical theory of energy efficient neural computation and communication. *IEEE Transactions on Information Theory*, 56(2),

852-874.

- Bettle, R., & Rosati, A. G. (2021). The primate origins of human social cognition. *Language Learning and Development*, 17(2), 96-127.
- Bishop, G. H. (1959). The relation between nerve fiber size and sensory modality: phylogenetic implications of the afferent innervation of cortex. *The Journal of Nervous and Mental Disease*, 128(2), 89-114.
- Blasco, O. (1996). Radon-Nikodym versus Fatou. En 29th National Congress of the Mexican Mathematical Society (F. Avila Murillo et al., eds.), San Luis Potosí, México (pp. 71-76).
- Borst, A., & Theunissen, F. E. (1999). Information theory and neural coding. *Nature neuroscience*, 2(11), 947.
- Braitenberg, V. (1962). A note on myeloarchitectonics. *Journal of comparative Neurology*, 118(2), 141-156.
- Brazel, C. Y., Romanko, M. J., Rothstein, R. P., & Levison, S. W. (2003). Roles of the mammalian subventricular zone in brain development. *Progress in neurobiology*, 69(1), 49-69.
- Brillouin, L. (1953). The negentropy principle of information. *Journal of Applied Physics*, 24(9), 1152-1163.
- Brodmann, K. (1906). Beiträge zur histologischen Lokalisation der Grosshirnrinde: Fünfte Mitteilung: Über den allgemeinen Bauplan des Cortex Pallii bei den Mammaliern und zwei homologe Rindenfelder im besonderen. Zugleich ein Beitrag zur Furchenlehre [Contribuciones a la localización histológica de la corteza cerebral: Quinto informe: Sobre el plan de construcción general de la corteza del pallium en los mamíferos y dos campos corticales homólogos en particular. Al mismo tiempo, una contribución a la doctrina de las circunvoluciones]. JA Barth.
- Brodmann, K. (1909). Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues [Doctrina comparativa de la localización de la corteza cerebral en sus principios, representada sobre la base de la estructura celular.]. Barth.
- Brodmann, K. (1914). *Physiologie des gehirns* [La fisiología del cerebro]. Druck der Union deutsche Verlagsgesellschaft.
- Brown, E. N., Frank, L. M., Tang, D., Quirk, M. C., & Wilson, M. A. (1998). A statistical

- paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *Journal of Neuroscience*, 18(18), 7411-7425.
- Brown, G. L., Dale, H. H., & Feldberg, W. (1936). Reactions of the normal mammalian muscle to acetylcholine and to eserine. *The Journal of Physiology*, 87(4), 394.
- Brunel, N., Hakim, V., Isope, P., Nadal, J. P., & Barbour, B. (2004). Optimal information storage and the distribution of synaptic weights: perceptron versus Purkinje cell. *Neuron*, 43(5), 745-757.
- Buehlmann, A., & Deco, G. (2008). The neuronal basis of attention: rate versus synchronization modulation. *Journal of Neuroscience*, 28(30), 7679-7686.
- Bullier, J., Hupé, J. M., James, A. C., & Girard, P. (2001). The role of feedback connections in shaping the responses of visual cortical neurons. *En Progress in brain research* (Vol. 134, pp. 193-204). Elsevier.
- Burkhardt, P. (2015). The origin and evolution of synaptic proteins—choanoflagellates lead the way. *Journal of Experimental Biology*, 218(4), 506-514.
- Burkhardt, P., Grønberg, M., McDonald, K., Sultur, T., Wang, Q., & King, N. (2014). Evolutionary insights into premetazoan functions of the neuronal protein homer. *Molecular biology and evolution*, 31(9), 2342-2355.
- Bystron, I., Blakemore, C., & Rakic, P. (2008). Development of the human cerebral cortex: Boulder Committee revisited. *Nature Reviews Neuroscience*, 9(2), 110-122.
- Cackowski, Z. (1982). Information, Regulation, Negentropy. *En Polish Essays in the Philosophy of the Natural Sciences* (pp. 249-264). Springer, Dordrecht.
- Chien, C. B. (2005). Guidance of Axons and Dendrites. In *Developmental Neurobiology* (pp. 241-267). Springer, Boston, MA.
- Chien, C. B., Rosenthal, D. E., Harris, W. A., & Holt, C. E. (1993). Navigational errors made by growth cones without filopodia in the embryonic *Xenopus* brain. *Neuron*, 11(2), 237-251.
- Chih, B., Engelman, H., & Scheiffele, P. (2005). Control of excitatory and inhibitory synapse formation by neuroligins. *Science*, 307(5713), 1324-1328.
- Chitnis, A. B., & Kuwada, J. Y. (1990). Axonogenesis in the brain of zebrafish embryos. *Journal of Neuroscience*, 10(6), 1892-1905.
- Clausius, R. (1854). Über eine veränderte Form des zweiten Hauptsatzes der mechanischen

- Wärmetheorie. *Annalen der Physik*, 169(12), 481-506.
- Collell, G., & Fauquet, J. (2015). Brain activity and cognition: a connection from thermodynamics and information theory. *Frontiers in psychology*, 6, 818.
- Corning, P. A. (2007). Control information theory: The 'missing link' in the science of cybernetics. *Systems Research and Behavioral Science: The Official Journal of the International Federation for Systems Research*, 24(3), 297-311.
- Corning, P. A., & Kline, S. J. (1998). Thermodynamics, information and life revisited, Part II: 'Thermoeconomics' and 'Control information'. *Systems Research and Behavioral Science: The Official Journal of the International Federation for Systems Research*, 15(6), 453-482.
- Cover, T. M., & Thomas, J. A. (2012). *Elements of information theory*. John Wiley & Sons.
- Cowan, J. D. (1990). Discussion: McCulloch-Pitts and related neural nets from 1943 to 1989. *Bulletin of mathematical biology*, 52(1-2), 73-97.
- Cowan, W. M., Fawcett, J. W., O'Leary, D. D., & Stanfield, B. B. (1984). Regressive events in neurogenesis. *Science*, 225(4668), 1258-1265.
- Crossley, M., Staras, K., & Kemenes, G. (2016). A two-neuron system for adaptive goal-directed decision-making in *Lymnaea*. *Nature communications*, 7, 11793.
- Curto, C. (2017). What can topology tell us about the neural code?. *Bulletin of the American Mathematical Society*, 54(1), 63-78.
- Cuvier, G. (1825). *Recherches sur les ossements fossiles*.
- Dale, H. H., Feldberg, W., & Vogt, M. (1936). Release of acetylcholine at voluntary motor nerve endings. *The Journal of physiology*, 86(4), 353.
- Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience: computational and mathematical modeling of neural systems*. Computational Neuroscience Series.
- Dayan, P., Hinton, G. E., Neal, R. M., & Zemel, R. S. (1995). The helmholtz machine. *Neural computation*, 7(5), 889-904.
- Decharms, R. C., & Merzenich, M. M. (1996). Primary cortical representation of sounds by the coordination of action-potential timing. *Nature*, 381(6583), 610-613.
- Denes, A. S., Jékely, G., Steinmetz, P. R., Raible, F., Snyman, H., Prud'homme, B., ... & Arendt, D. (2007). Molecular architecture of annelid nerve cord supports common origin of nervous system centralization in bilateria. *Cell*, 129(2), 277-288.
- Deoni, S. C., Dean III, D. C., Remer, J., Dirks, H., & O'Muircheartaigh, J. (2015). Cortical

- maturation and myelination in healthy toddlers and young children. *Neuroimage*, 115, 147-161.
- Dohrn, A. (1875). *Der ursprung der wirbelthiere und das princip des functionswechsels: genealogische skizzen* [El origen de los vertebrados y el principio de cambio de función: esbozos genealógicos.]. W. Engelmann.
- Dulesov, A. S., Karandeev, D. J., & Dulesova, N. (2018). Improving the operation quality of technical systems using information theory models.
- Dulesov, A., Karandeev, D., & Krasnova, T. (2017). The evaluation of the correlation between entropy and negentropy in the structure of a technical system. In *MATEC Web of Conferences* (Vol. 129, p. 03005). EDP Sciences.
- Ebbesson, S. O. (1984). Evolution and ontogeny of neural circuits. *Behavioral and Brain Sciences*, 7(3), 321-331.
- Eccles, J. C., & Sherrington, C. S. (1931). Studies on the flexor reflex. VI.—Inhibition. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character*, 109(761), 91-113.
- El Gamal, A., & Kim, Y. H. (2011). *Network information theory*. Cambridge university press.
- Elias, P., Feinstein, A., & Shannon, C. (1956). A note on the maximum flow through a network. *IRE Transactions on Information Theory*, 2(4), 117-119.
- Elliott, G. R., & Leys, S. P. (2010). Evidence for glutamate, GABA and NO in coordinating behaviour in the sponge, *Ephydatia muelleri* (Demospongiae, Spongillidae). *Journal of Experimental Biology*, 213(13), 2310-2321.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870), 429-433.
- Falk, G., & Fatt, P. (1964). Linear electrical properties of striated muscle fibres observed with intracellular electrodes. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 160(978), 69-123.
- Fodor, J. A. (1985). Precis of the modularity of mind. *Behavioral and brain sciences*, 8(1), 1-5.
- Ford, G. W., & O'Connell, R. F. (2006). A quantum violation of the second law?. *Physical review letters*, 96(2), 020402.
- Ford, L. R., & Fulkerson, D. R. (1956). Maximal flow through a network. *Canadian journal of Mathematics*, 8, 399-404.

- Frigg, R., & Werndl, C. (2011). Entropy-a guide for the perplexed.
- Friston, K. (2010). The free-energy principle: a unified brain theory?. *Nature reviews neuroscience*, 11(2), 127-138.
- Friston, K., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. *Journal of Physiology-Paris*, 100(1-3), 70-87.
- Frostick, S. P., Yin, Q., & Kemp, G. J. (1998). Schwann cells, neurotrophic factors, and peripheral nerve regeneration. *Microsurgery: Official Journal of the International Microsurgical Society and the European Federation of Societies for Microsurgery*, 18(7), 397-405.
- Fujita, S. (1960). Mitotic pattern and histogenesis of the central nervous system. *Nature*, 185(4714), 702-703.
- Fukushima, K. (1988). Neocognitron: A hierarchical neural network capable of visual pattern recognition. *Neural networks*, 1(2), 119-130.
- Gabor, D. (1946). Theory of communication. Part 1: The analysis of information. *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering*, 93(26), 429-441.
- Gauck, V., & Jaeger, D. (2003). The contribution of NMDA and AMPA conductances to the control of spiking in neurons of the deep cerebellar nuclei. *Journal of Neuroscience*, 23(22), 8109-8118.
- Genewein, T., Leibfried, F., Grau-Moya, J., & Braun, D. A. (2015). Bounded rationality, abstraction, and hierarchical decision-making: An information-theoretic optimality principle. *Frontiers in Robotics and AI*, 2, 27.
- Glasser, M. F., Goyal, M. S., Preuss, T. M., Raichle, M. E., & Van Essen, D. C. (2014). Trends and properties of human cerebral cortex: correlations with cortical myelin content. *Neuroimage*, 93, 165-175.
- Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 249-256). *JMLR Workshop and Conference Proceedings*.
- Granit, R. (1982). Interactions between Pavlov and Sherrington. *Trends in Neurosciences*, 5, 184-186.
- Gundersen, R. W., & Barrett, J. N. (1980). Characterization of the turning response of dorsal

- root neurites toward nerve growth factor. *The Journal of cell biology*, 87(3), 546-554.
- Halmos, P. R. (2013). *Measure theory* (Vol. 18). Springer.
- Halmos, P. R., & Savage, L. J. (1949). Application of the Radon-Nikodym theorem to the theory of sufficient statistics. *The Annals of Mathematical Statistics*, 20(2), 225-241.
- Hamburger, V., & Levi-Montalcini, R. (1949). Proliferation, differentiation and degeneration in the spinal ganglia of the chick embryo under normal and experimental conditions. *Journal of Experimental Zoology*, 111(3), 457-501.
- Hanslmayr, S., Staudigl, T., & Fellner, M. C. (2012). Oscillatory power decreases and long-term memory: the information via desynchronization hypothesis. *Frontiers in human neuroscience*, 6, 74.
- Hartenstein, V., & Stollewerk, A. (2015). The evolution of early neurogenesis. *Developmental Cell*, 32(4), 390-407.
- Hartley, R. V. (1928). Transmission of information. *Bell System technical journal*, 7(3), 535-563.
- Hatakeyama, J., Bessho, Y., Katoh, K., Ookawara, S., Fujioka, M., Guillemot, F., & Kageyama, R. (2004). Hes genes regulate size, shape and histogenesis of the nervous system by control of the timing of neural stem cell differentiation. *Development*, 131(22), 5539-5550.
- Hawkins, A., & Olszewski, J. (1957). Glia/nerve cell index for cortex of the whale. *Science*, 126(3263), 76-77.
- Henson, R. N., & Gagnepain, P. (2010). Predictive, interactive multiple memory systems. *Hippocampus*, 20(11), 1315-1326.
- Herrick, C. J. (1948). *The brain of the tiger salamander, Ambystoma tigrinum*. University of Chicago Press.
- Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of experimental psychology*, 4(1), 11-26.
- Ho, M. W. (1994). What is (Schrödinger's) negentropy?. *Modern Trends in BioThermoKinetics*, 3(1994), 50-61.
- Hodgkin, A. L. (1947). The membrane resistance of a non-medullated nerve fibre. *The Journal of physiology*, 106(3), 305.
- Hodgkin, A. L., & Huxley, A. F. (1952). The components of membrane conductance in the giant axon of *Loligo*. *The Journal of physiology*, 116(4), 473-496.

- Hodgkin, A. L., & Huxley, A. F. (1952a). A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*, 117(4), 500.
- Hodgkin, A. L., & Huxley, A. F. (1952b). The components of membrane conductance in the giant axon of *Loligo*. *The Journal of physiology*, 116(4), 473.
- Hodgkin, A. L., & Rushton, W. A. H. (1946). The electrical constants of a crustacean nerve fibre. *Proceedings of the Royal Society of London. Series B-Biological Sciences*, 133(873), 444-479.
- Hodgkin, A. L., Huxley, A. F., & Katz, B. (1952). Measurement of current-voltage relations in the membrane of the giant axon of *Loligo*. *The Journal of physiology*, 116(4), 424.
- Hofman, M. A. (1982). Encephalization in mammals in relation to the size of the cerebral cortex. *Brain, Behavior and Evolution*, 20(1-2), 84-96.
- Holland, N. D. (2016). Nervous systems and scenarios for the invertebrate-to-vertebrate transition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1685), 20150047.
- Holloway Jr, R. L. (1966). Cranial Capacity, Neural Reorganization, and Hominid Evolution: A Search for More Suitable Parameters 1. *American anthropologist*, 68(1), 103-121.
- Hopf, A. (1954). Die myeloarchitektonik des isocortex temporalis beim menschen [La mieloarquitectónica del isocórtex temporal en el ser humano].
- Hopf, A. (1968). Photometric studies on the myeloarchitecture of the human temporal lobe. *Journal fur Hirnforschung*, 10(4), 285-297.
- Hopf, A. (1968b). Registration of the myeloarchitecture of the human frontal lobe with an extinction method. *Journal fur Hirnforschung*, 10(3), 259-269.
- Hopf, A. (1969). Photometric studies on the myeloarchitecture of the human parietal lobe. I. Parietal region. *Journal fur Hirnforschung*, 11(4), 253-265.
- Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5), 359-366.
- Howard, J. (1996). The movement of kinesin along microtubules. *Annual review of physiology*, 58(1), 703-729.
- Hu, Y. H., & Hwang, J. N. (Eds.). (2002). *Handbook of neural network signal processing*.
- Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology*, 148(3), 574-591.

- Hunt, B. A., Tewarie, P. K., Mougin, O. E., Geades, N., Jones, D. K., Singh, K. D., ... & Brookes, M. J. (2016). Relationships between cortical myeloarchitecture and electrophysiological networks. *Proceedings of the National Academy of Sciences*, 113(47), 13510-13515.
- Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of experimental psychology*, 45(3), 188.
- Ince, R. A., Senatore, R., Arabzadeh, E., Montani, F., Diamond, M. E., & Panzeri, S. (2010). Information-theoretic methods for studying population codes. *Neural Networks*, 23(6), 713-727.
- Inzlicht, M., Bartholow, B. D., & Hirsh, J. B. (2015). Emotional foundations of cognitive control. *Trends in cognitive sciences*, 19(3), 126-132.
- Inzlicht, M., Shenhav, A., & Olivola, C. Y. (2018). The effort paradox: Effort is both costly and valued. *Trends in cognitive sciences*, 22(4), 337-349.
- Irvine, J., & Harle, D. (2001). *Data communications and networks: an engineering approach*. John Wiley & Sons.
- Jacobs, D. K., Nakanishi, N., Yuan, D., Camara, A., Nichols, S. A., & Hartenstein, V. (2007). Evolution of sensory structures in basal metazoa. *Integrative and Comparative Biology*, 47(5), 712-723.
- Jeffreys, H. (1998). *The theory of probability*. OUP Oxford.
- Jezek, K., Henriksen, E. J., Treves, A., Moser, E. I., & Moser, M. B. (2011). Theta-paced flickering between place-cell maps in the hippocampus. *Nature*, 478(7368), 246-249.
- Keysers, C., & Perrett, D. I. (2004). Demystifying social cognition: a Hebbian perspective. *Trends in cognitive sciences*, 8(11), 501-507.
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Kolmogorov, A. N. (1963). On tables of random numbers. *Sankhyā: The Indian Journal of Statistics, Series A*, 369-376.
- Krause, W. (1860). *Die terminalen Körperchen der einfach sensiblen Nerven*. Hahn.
- Kriegeskorte, N., & Douglas, P. K. (2018). Cognitive computational neuroscience. *Nature neuroscience*, 21(9), 1148-1160.
- Kristan Jr, W. B. (2016). Early evolution of neurons. *Current Biology*, 26(20), R949-R954.
- Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The annals of*

- mathematical statistics, 22(1), 79-86.
- La Cerra, P. (2003). The first law of psychology is the second law of thermodynamics: the energetic evolutionary model of the mind and the generation of human psychological phenomena. *Human Nature Review*, 3, 440-447.
- Lachapelle, F., Duhamel-Clerin, E., Gansmüller, A., Evercooren, A. B. V., Villarroya, H., & Gumpel, M. (1994). Transplanted transgenically marked oligodendrocytes survive, migrate and myelinate in the normal mouse brain as they do in the shiverer mouse brain. *European Journal of Neuroscience*, 6(5), 814-824.
- Landauer, R. (1961). Irreversibility and heat generation in the computing process. *IBM journal of research and development*, 5(3), 183-191.
- Lebowitz, J. L. (1993). Boltzmann's entropy and time's arrow. *Physics today*, 46, 32-32.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436-444.
- Lein, P., Johnson, M., Guo, X., Rueger, D., & Higgins, D. (1995). Osteogenic protein-1 induces dendritic growth in rat sympathetic neurons. *Neuron*, 15(3), 597-605.
- Lennie, P. (2003). The cost of cortical computation. *Current biology*, 13(6), 493-497.
- Letourneau, P. C. (1975). Cell-to-substratum adhesion and guidance of axonal elongation. *Developmental biology*, 44(1), 92-101.
- Levi-Montalcini, R., & Levi, G. (1944). Correlazioni nello sviluppo tra varie parti del sistema nervoso. *Acta Pontificia Academia Scientiarum*, 7, 527-575.
- Liddell, E. G. T., & Sherrington, C. S. (1925). Recruitment and some other features of reflex inhibition. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character*, 97(686), 488-518.
- Liebeskind, B. J., Hillis, D. M., Zakon, H. H., & Hofmann, H. A. (2016). Complex homology and the evolution of nervous systems. *Trends in ecology & evolution*, 31(2), 127-135.
- Loewi, O. (1921). Über humorale übertragbarkeit der Herznervenwirkung. *Pflüger's Archiv für die gesamte Physiologie des Menschen und der Tiere*, 189(1), 239-242.
- Luce, R. D. (2012). *Individual choice behavior: A theoretical analysis*. Courier Corporation.
- Luria, A. R. (2011). *Las funciones corticales superiores del hombre*. Fontamara.
- Luskin, M. B., & Shatz, C. J. (1985). Studies of the earliest generated cells of the cat's visual cortex: cogeneration of subplate and marginal zones. *Journal of Neuroscience*, 5(4), 1062-1075.
- Marcer, P., & Schempp, W. (1998). The brain as a conscious system. *International Journal of*

- general system, 27(1-3), 131-248
- Martin, R. D. (1981). Relative brain size and basal metabolic rate in terrestrial vertebrates. *Nature*, 293(5827), 57-60.
- Martin, R. D. (1983). Human brain evolution in an ecological context (James Arthur lecture on the evolution of the human brain, no. 52, 1982). New York, NY: American Museum of Natural History.
- Maynard, R. L., & Downes, N. (2019). *Anatomy and Histology of the Laboratory rat in Toxicology and Biomedical Research*. Academic Press.
- McCarthy, J. (1979). *Ascribing Mental Qualities to Machines* (No. STAN-CS-79-725). Stanford University, California department of computer science.
- McClare, C. W. F. "Chemical machines, Maxwell's demon and living organisms." *Journal of theoretical biology* 30.1 (1971): 1-34.
- McCulloch, W., & Pitts, W. (1943). A Logical Calculus of the Ideas Imminent in Nervous Activity. *Bulletin of Mathematical Biophysics*, 5, 115-33.
- McGee, A. W., & Strittmatter, S. M. (2003). The Nogo-66 receptor: focusing myelin inhibition of axon regeneration. *Trends in neurosciences*, 26(4), 193-198.
- Menn, B., Garcia-Verdugo, J. M., Yaschine, C., Gonzalez-Perez, O., Rowitch, D., & Alvarez-Buylla, A. (2006). Origin of oligodendrocytes in the subventricular zone of the adult brain. *Journal of Neuroscience*, 26(30), 7907-7918.
- Mitchison, T., & Kirschner, M. (1988). Cytoskeletal dynamics and nerve growth. *Neuron*, 1(9), 761-772.
- Monderer, R. S., Harrison, D. M., & Haut, S. R. (2002). Neurofeedback and epilepsy. *Epilepsy & Behavior*, 3(3), 214-218.
- Moroz, L. L. (2015). Convergent evolution of neural systems in ctenophores. *Journal of Experimental Biology*, 218(4), 598-611.
- Moroz, L. L., & Kohn, A. B. (2016). Independent origins of neurons and synapses: insights from ctenophores. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1685), 20150041.
- Moroz, L. L., Kocot, K. M., Citarella, M. R., Dosung, S., Norekian, T. P., Povolotskaya, I. S., ... & Ptitsyn, A. (2014). The ctenophore genome and the evolutionary origins of neural systems. *Nature*, 510(7503), 109-114.
- Nicholls, J. G., Martin, A. R., Wallace, B. G., & Fuchs, P. A. (2001). *From neuron to brain*

- (Vol. 271). Sunderland, MA: Sinauer Associates.
- Nieuwenhuys, R. (2013). The myeloarchitectonic studies on the human cerebral cortex of the Vogt–Vogt school, and their significance for the interpretation of functional neuroimaging data. *Brain Structure and Function*, 218(2), 303-352.
- Nikodym, O. (1930). Sur une généralisation des intégrales de MJ Radon [Sobre una generalización de las integrales de MJ Radon]. *Fundamenta Mathematicae*, 15(1), 131-179.
- Northcutt, R. G. (1984). Evolution of the vertebrate central nervous system: patterns and processes. *American Zoologist*, 24(3), 701-716.
- Nowakowski, R. S., & Hayes, N. L. (2005). Cell proliferation in the developing mammalian brain. In *Developmental Neurobiology* (pp. 21-39). Springer, Boston, MA.
- Nübler-Jung, K., & Arendt, D. (1996). Enteropneusts and chordate evolution. *Current Biology*, 6(4), 352-353.
- O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map: preliminary evidence from unit activity in the freely-moving rat. *Brain research*.
- Ohlenforst, B., Zekveld, A. A., Lunner, T., Wendt, D., Naylor, G., Wang, Y., ... & Kramer, S. E. (2017). Impact of stimulus-related factors and hearing impairment on listening effort as indicated by pupil dilation. *Hearing Research*, 351, 68-79.
- Palm, G., & Aertsen, A. (Eds.). (2012). *Brain Theory: Proceedings of the First Trieste Meeting on Brain Theory, October 1–4, 1984*. Springer Science & Business Media.
- Palomero-Gallagher, N., & Zilles, K. (2019). Cortical layers: Cyto-, myelo-, receptor-and synaptic architecture in human cortical areas. *Neuroimage*, 197, 716-741.
- Parascandolo, G., Huttunen, H., & Virtanen, T. (2016). Taming the waves: sine as activation function in deep neural networks.
- Park, I. M., & Pillow, J. W. (2017). Bayesian efficient coding. *bioRxiv*, 178418.
- Parrish, A. E., Brosnan, S. F., & Beran, M. J. (2015). Do you see what I see? A comparative investigation of the Delboeuf illusion in humans (*Homo sapiens*), rhesus monkeys (*Macaca mulatta*), and capuchin monkeys (*Cebus apella*). *Journal of Experimental Psychology: Animal Learning and Cognition*, 41(4), 395.
- Patterson, P. H. (1988). On the importance of being inhibited, or saying no to growth cones. *Neuron*, 1(4), 263-267.
- Patton, B., & Burgess, R. W. (2005). Synaptogenesis. In *Developmental Neurobiology* (pp.

- 269-316). Boston, MA: Springer US.
- Pavão, R., Saviotto, J. P., Sato, J. R., Xavier, G. F., & Helene, A. F. (2016). On sequence learning models: Open-loop control not strictly guided by Hick's law. *Scientific reports*, 6, 23018.
- Pearce, J. M. S. (2004). Sir Charles Scott Sherrington (1857–1952) and the synapse. *Journal of Neurology, Neurosurgery & Psychiatry*, 75(4), 544-544.
- Pfrieger, F. W. (2009). Roles of glial cells in synapse development. *Cellular and molecular life sciences*, 66(13), 2037-2047.
- Pfrieger, F. W., & Barres, B. A. (1997). Synaptic efficacy enhanced by glial cells in vitro. *Science*, 277(5332), 1684-1687.
- Pias, C. (2016). *Cybernetics-the Macy Conferences 1946-1953: the complete transactions*. In Macy Conferences. Chicago University Press.
- Pierce, J. R. (1980). *An introduction to information theory: Symbols, Signals and Noise E*, 2.
- Pilbeam, D., & Gould, S. J. (1974). Size and scaling in human evolution. *Science*, 186(4167), 892-901.
- Pribram, K. H., & Carlton, E. H. (1986). Holonomic brain theory in imaging and object perception. *Acta Psychologica*, 63(2), 175-210.
- Proakis, J. G., Salehi, M., Zhou, N., & Li, X. (1994). *Communication systems engineering* (Vol. 2). New Jersey: Prentice Hall.
- Rakic, P., & Sidman, R. L. (1968). Subcommissural organ and adjacent ependyma: autoradiographic study of their origin in the mouse brain. *American Journal of Anatomy*, 122(2), 317-335.
- Ramón y Cajal, S. (1890). A quelle époque apparaissent les expansions des cellules nerveuses de la moelle épinière du poulet [¿En qué época aparecen las expansiones de las células nerviosas de la médula espinal del pollo?]. *Anatomischer Anzeiger*, 5, 609-613.
- Ramon y Cajal, S. (1911). *Histologie du système nerveux de l'homme et des vertèbres* [Histología del sistema nervioso del hombre y de los vertebrados] (Vol. 2). Paris: A. Maloine.
- Rényi, A. (1960). Some Fundamental Questions of Information Theory. *MTA III Oszt. Közl.*, 251 (1960); On measures of information and entropy. En *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability* (p. 547).
- Rényi, A. (1961). On measures of entropy and information. En *Proceedings of the Fourth*

- Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics (Vol. 4, pp. 547-562). University of California Press.
- Richards, G. S., Simionato, E., Perron, M., Adamska, M., Vervoort, M., & Degnan, B. M. (2008). Sponge genes provide new insight into the evolutionary origin of the neurogenic circuit. *Current Biology*, 18(15), 1156-1161.
- Rinzel, J. (1990). Discussion: Electrical excitability of cells, theory and experiment: Review of the Hodgkin-Huxley foundation and an update. *Bulletin of Mathematical Biology*, 52(1-2), 3-23.
- Robert, C. P., Chopin, N., & Rousseau, J. (2009). Harold Jeffreys's theory of probability revisited. *Statistical Science*, 24(2), 141-172.
- Robertson, R. G., Rolls, E. T., Georges-François, P., & Panzeri, S. (1999). Head direction cells in the primate pre-subiculum. *Hippocampus*, 9(3), 206-219.
- Rolls, E. T. (2008). Memory, attention, and decision-making. Chapter 2. OUP.
- Rolls, E. T., & Treves, A. (1990). The relative advantages of sparse versus distributed encoding for associative neuronal networks in the brain. *Network: computation in neural systems*, 1(4), 407-421.
- Rolls, E. T., & Treves, A. (1998). *Neural networks and brain function* (Vol. 572). Oxford: Oxford university press.
- Rolls, E. T., & Treves, A. (2011). The neuronal encoding of information in the brain. *Progress in neurobiology*, 95(3), 448-490.
- Ronan, M., & Northcutt, R. G. (1990). Projections ascending from the spinal cord to the brain in petromyzontid and myxinoid agnathans. *Journal of Comparative Neurology*, 291(4), 491-508.
- Rosenblatt, F. (1957). *The perceptron, a perceiving and recognizing automaton* (Project Para). Cornell Aeronautical Laboratory.
- Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6), 386.
- Rosenblatt, F. (1961). *Principles of neurodynamics. perceptrons and the theory of brain mechanisms*. Cornell Aeronautical Lab Inc Buffalo NY.
- Rosso, O. A. (2007). Entropy changes in brain function. *International Journal of Psychophysiology*, 64(1), 75-80.

- Rubin, O., & Meiran, N. (2005). On the origins of the task mixing cost in the cuing task-switching paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(6), 1477.
- Ruffini, A. (1906). Le espansioni nervose periferiche alla luce dell'analisi moderna. *Monit. zool. ital.*, 17, 68-87.
- Sakarya, O., Armstrong, K. A., Adamska, M., Adamski, M., Wang, I. F., Tidor, B., ... & Kosik, K. S. (2007). A post-synaptic scaffold at the origin of the animal kingdom. *PloS one*, 2(6), e506.
- Sakmann, B., Noma, A., & Trautwein, W. (1983). Acetylcholine activation of single muscarinic K⁺ channels in isolated pacemaker cells of the mammalian heart. *Nature*, 303(5914), 250-253.
- Sarampalis, A., Kalluri, S., Edwards, B., & Hafter, E. (2009). Objective measures of listening effort: Effects of background noise and noise reduction. *Journal of Speech, Language, and Hearing Research*.
- Sarnat, H., & Netsky, M. (1981). Telencephalon. *Evolution of the nervous system*, 321-378.
- Sawant, H. K., & Jalali, Z. (2010). Detection and classification of EEG waves. *Oriental Journal of Computer Science and Technology*, 3(1), 207-213.
- Sayood, K. (2018). Information Theory and Cognition: A Review. *Entropy*, 20(9), 706.
- Schaper, A. & Cohen, C., (1905). Beiträge Zur analyse des tierischen Wachstums. *Archiv für Entwicklungsmechanik der Organismen*, 19(3), 348-445.
- Schonewille, M., Khosrovani, S., Winkelman, B. H., Hoebeek, F. E., De Jeu, M. T., Larsen, I. M., ... & De Zeeuw, C. I. (2006). Purkinje cells in awake behaving animals operate at the upstate membrane potential. *Nature neuroscience*, 9(4), 459.
- Schrödinger, E. (1944). *What is life? The physical aspect of the living cell*. Cambridge University Press
- Schyns, P. G., Thut, G., & Gross, J. (2011). Cracking the code of oscillatory activity. *PLoS biology*, 9(5).
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and brain sciences*, 3(3), 417-424.
- Sgonina, K. (1938). Zur vergleichenden anatomie der entorhinal-und präsubikularregion [Para la anatomía comparada de la región entorrinal y presubicular]. *J. Psychol. Neurol*, 48, 56-163.

- Shams, Z., Norris, D. G., & Marques, J. P. (2019). A comparison of in vivo MRI based cortical myelin mapping using T1w/T2w and R1 mapping at 3T. *PloS one*, 14(7).
- Shannon, C. (1953). The lattice theory of information. *Transactions of the IRE professional Group on Information Theory*, 1(1), 105-107.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell system technical journal*, 27(3), 379-423.
- Shannon, C. E. (1949). Communication in the presence of noise. *Proceedings of the IRE*, 37(1), 10-21.
- Sherrington, C. (1952). *The integrative action of the nervous system*. CUP Archive.
- Skaggs, W. E., McNaughton, B. L., & Gothard, K. M. (1993). An information-theoretic approach to deciphering the hippocampal code. In *Advances in neural information processing systems* (pp. 1030-1037).
- Smith, L. N. (2018). A disciplined approach to neural network hyper-parameters: Part 1 - learning rate, batch size, momentum, and weight decay. *arXiv preprint arXiv:1803.09820*.
- Srivastava, M., Begovic, E., Chapman, J., Putnam, N. H., Hellsten, U., Kawashima, T., ... & Signorovitch, A. Y. (2008). The *Trichoplax* genome and the nature of placozoans. *Nature*, 454(7207), 955-960.
- Srivastava, M., Simakov, O., Chapman, J., Fahey, B., Gauthier, M. E., Mitros, T., ... & Larroux, C. (2010). The *Amphimedon queenslandica* genome and the evolution of animal complexity. *Nature*, 466(7307), 720-726.
- Staley, K. (2015). Molecular mechanisms of epilepsy. *Nature neuroscience*, 18(3), 367-372.
- Stone, J. V. (2015). *Information theory: a tutorial introduction*. Sebtel Press.
- Stuart, D. G. (2005). Integration of posture and movement: contributions of Sherrington, Hess, and Bernstein. *Human Movement Science*, 24(5-6), 621-643.
- Switkes, E., Bradley, A., & De Valois, K. K. (1988). Contrast dependence and mechanisms of masking interactions among chromatic and luminance gratings. *JOSA A*, 5(7), 1149-1162.
- Szilard, L. (1929). Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen [Sobre la reducción de la entropía en un sistema termodinámico por la intervención de seres inteligentes]. *Zeitschrift für Physik*, 53(11), 840-856.

- Takahashi, T., Nowakowski, R. S., & Caviness, V. S. (1995). The cell cycle of the pseudostratified ventricular epithelium of the embryonic murine cerebral wall. *Journal of Neuroscience*, 15(9), 6046-6057.
- Tanksley, D., & Wunsch II, D. C. (2019). Reproducibility via Crowdsourced Reverse Engineering: A Neural Network Case Study With DeepMind's Alpha Zero. arXiv preprint arXiv:1909.03032.
- Tishby, N., Pereira, F. C., & Bialek, W. (2000). The information bottleneck method. arXiv preprint physics/0004057.
- Tobias, P. V. (1971). The brain in hominid evolution (James Arthur lecture on the evolution of the human brain, no. 38, 1969).
- Vallée, R. (1990). Plato's cave revisited. *Kybernetes*, 19(3), 37-42.
- Vallstedt, A. (2004). The role of Nkx proteins neuronal and glial specification. Institutionen för cell-och molekylärbiologi (CMB)/Department of Cell and Molecular Biology.
- Vetter, M. L., & Dorsky, R. I. (2005). Neurogenesis. In *Developmental Neurobiology* (pp. 129-150). Springer, Boston, MA.
- Vigotski, L. S. (2001). *Psicología pedagógica; Un curso breve*. Aique.
- Vogt, C., & Vogt, O. (1919). *Allgemeine ergebnisse unserer hirnforschung [Resultados generales de nuestra investigación cerebral]* (Vol. 21). JA Barth.
- Vogt, O. (1911). Die myeloarchitektonik des isocortex parietalis [La mieloarquitectónica del isocórtex parietal]. *J Psychol Neurol*, 18, 379-390.
- Vogt, O., & Vogt, C. (1903). Zur anatomischen Gliederung des Cortex cerebri [Sobre la estructuración anatómica de la corteza cerebral]. *J Psychol Neurol*, 2, 160-180.
- Von Economo, C. F., & Koskinas, G. N. (1925). *Die cytoarchitektonik der hirnrinde des erwachsenen menschen [La citoarquitectura de la corteza cerebral del ser humano adulto]*. J. Springer.
- Von Foerster, H. (1968). Idea, image and memory. En *The future of the brain sciences* (pp. 89). Springer, Boston, MA.
- Von Foerster, H. (2002). *Understanding understanding: Essays on cybernetics and cognition*. Springer Science & Business Media.
- Von Neumann, J. (1993). First Draft of a Report on the EDVAC. *IEEE Annals of the History of Computing*, 15(4), 27-75.
- Von Neumann, J. (2017). The general and logical theory of automata. En *Systems Research*

- for Behavioral Sciencesystems Research (pp. 97-107). Routledge.
- Vos, B. P., Maex, R., Volny-Luraghi, A., & De Schutter, E. (1999). Parallel fibers synchronize spontaneous activity in cerebellar Golgi cells. *The Journal of neuroscience*, 19(11), RC6.
- Walter, J. T., & Khodakhah, K. (2006). The linear computational algorithm of cerebellar Purkinje cells. *Journal of Neuroscience*, 26(50), 12861-12872.
- Wang, X., & Poor, H. V. (2004). *Wireless communication systems: Advanced techniques for signal reception*. Prentice Hall Professional.
- Warf, B. C., Fok-Seang, J., & Miller, R. H. (1991). Evidence for the ventral origin of oligodendrocyte precursors in the rat spinal cord. *Journal of Neuroscience*, 11(8), 2477-2488.
- Wiener N. (1948). *Cybernetics: or Control and Communication in the Animal and the Machine*. M.I.T. Press.
- Wiener, N. (1961). *Cybernetics: or control and communication in the animal and the machine*. M.I.T. Press.
- Zenon, A., Solopchuk, O., & Pezzulo, G. (2018). An information-theoretic perspective on the costs of cognition. *Neuropsychologia*, 123, 5-18.
- Zilles, K., & Amunts, K. (2010). Centenary of Brodmann's map—conception and fate. *Nature Reviews Neuroscience*, 11(2), 139-145.
- Zucker, R. S., Kullmann, D. M., & Kaeser, P. S. (2014). Release of neurotransmitters. In *From Molecules to Networks* (pp. 443-488). Academic Press.