

UNIVERSIDAD DE MÁLAGA

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA DE
TELECOMUNICACIÓN

PROGRAMA DE DOCTORADO EN INGENIERÍA DE
TELECOMUNICACIÓN

Integrative Signal Processing and Explainable Artificial Intelligence for Functional Connectivity Modeling in Language Disorders

Tesis Doctoral

Marco A. Formoso Trigo

Supervisor: Andrés Ortiz García



December 2024

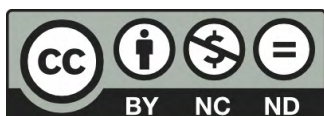


UNIVERSIDAD
DE MÁLAGA

AUTOR: Marco Antonio Formoso Trigo

 <http://orcid.org/0000-0001-8969-1941>

EDITA: Publicaciones y Divulgación Científica. Universidad de Málaga



Esta obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional:

Cualquier parte de esta obra se puede reproducir sin autorización
pero con el reconocimiento y atribución de los autores.

No se puede hacer uso comercial de la obra y no se puede alterar, transformar o hacer
obras derivadas.

<http://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>

Esta Tesis Doctoral está depositada en el Repositorio Institucional de la Universidad de
Málaga (RIUMA): riuma.uma.es

Declaration of Original and Sole Authorship

I, Marco A. Formoso Trigo, a student in the Telecommunication Engineering doctoral program at the University of Málaga, author of the thesis submitted for the degree of Doctor at the University of Málaga, titled:

Integrative Signal Processing and Explainable Artificial Intelligence for Functional Connectivity Modeling in Language Disorders

Conducted under the supervision of Andrés Ortiz García,

DECLARE THAT:

The presented thesis is an original work that does not infringe upon intellectual property rights, industrial property rights, or any other rights, in accordance with the current legal framework (Royal Legislative Decree 1/1996, of April 12, approving the revised text of the Intellectual Property Law, regularizing, clarifying, and harmonizing the existing legal provisions on the matter), modified by Law 2/2019, of March 1. I also assume, before the University of Málaga and any other authority, the responsibility that may arise in the event of plagiarism in the submitted thesis, in accordance with the current legal framework.

December 2024

Authorization by the supervisor for the thesis defense

Dr. Andrés Ortiz García, professor at the University of Málaga in the Department of Communication Engineering, of the Higher Technical School of Telecommunication Engineering:

Hereby authorize the doctoral candidate, Marco A. Formoso Trigo, to proceed with the presentation and defense of his Doctoral Thesis titled:

Integrative Signal Processing and Explainable Artificial Intelligence for Functional Connectivity Modeling in Language Disorders,

of which he is the supervisor. Furthermore, they confirm that the articles used to support this work have not been utilized in any other thesis.

To attest to this, they issue and sign this report in:

December 2024

Resumen

La investigación que presenta esta tesis doctoral se centra en un tema de gran relevancia dentro de la neurociencia y de la aplicación de la inteligencia artificial al ámbito de la salud, especialmente para abordar el diagnóstico y la comprensión de los mecanismos cerebrales implicados en la dislexia. Esta condición neurológica, que repercute en una proporción significativa de la población infantil, se caracteriza por dificultades persistentes en el aprendizaje de la lectura pese a contar con una inteligencia normal, una instrucción adecuada y un entorno socioeducativo favorable. El impacto de la dislexia no se limita al desempeño académico de los niños afectados, sino que también incide en su desarrollo personal, emocional y social. Por ello, el establecimiento de herramientas diagnósticas más tempranas y precisas resulta fundamental para el abordaje integral de este trastorno y la implementación de intervenciones adecuadas que mitiguen sus efectos en el largo plazo. Con esta perspectiva, la tesis integra técnicas avanzadas de procesamiento de señales cerebrales con metodologías de inteligencia artificial, en particular con un énfasis en la inteligencia artificial explicable, con el fin de lograr tanto la identificación de patrones de conectividad funcional cerebral asociados a la dislexia como la confección de modelos computacionales interpretables que puedan aportar un sustento objetivo al diagnóstico clínico.

La problemática principal a la que se responde nace de las limitaciones de los métodos diagnósticos tradicionales, los cuales dependen en gran medida de la observación de la conducta lectora en un momento determinado, de la evaluación de pruebas estandarizadas y de la interpretación subjetiva del profesional. Estas pruebas, además de requerir que el sujeto presente ya determinadas competencias lectoras, pueden verse influidas por factores emocionales, motivacionales o contextuales, lo cual puede generar incertidumbre sobre la precisión en la identificación de niños con dislexia. El planteamiento de la tesis es que, si se pudieran determinar ciertos biomarcadores objetivos a partir de la actividad cerebral, no solo se mejoraría la exactitud diagnóstica, sino que también se podría llevar a cabo el proceso de detección en edades más tempranas, facilitando así la instauración de intervenciones preventivas más eficaces. En tal sentido, se parte de la hipótesis de que las alteraciones en la conectividad funcional y, específicamente, en la sincronización neuronal en bandas de frecuencia bajas como delta y theta, pueden constituir marcadores significativos de la dislexia. Estas oscilaciones neuronales de baja frecuencia intervienen en procesos de segmentación del habla y parecen estar estrechamente vinculadas con la fluidez y la precisión lectora.

El interés por estudiar la conectividad funcional se enmarca en la creciente evidencia que muestra que el cerebro trabaja como un sistema coordinado de múltiples regiones que se comunican de manera dinámica para ejecutar tareas cognitivas. En el caso del

lenguaje, se han identificado redes cerebrales muy distribuidas en ambos hemisferios, con regiones específicas que participan en la decodificación fonológica, la comprensión semántica, la articulación del habla y el control atencional. Cuando se habla de conectividad funcional, no se hace referencia exclusivamente a la anatomía de las regiones cerebrales, sino, sobre todo, a cómo dichas regiones se sincronizan temporalmente para intercambiar información de forma eficaz. Si esta coordinación está alterada en determinadas bandas de frecuencia o en redes concretas, es muy posible que se presenten problemas en la ejecución de tareas relacionadas con el lenguaje, como la lectura y la escritura.

Para estudiar esta conectividad funcional, esta tesis se basa en el uso de señales de electroencefalografía (EEG), una técnica que ofrece una alta resolución temporal y que permite capturar, casi en tiempo real, la actividad eléctrica del cerebro a través de múltiples electrodos colocados en el cuero cabelludo. A pesar de las ventajas de la EEG, como su carácter no invasivo y su capacidad para registrar oscilaciones cerebrales en distintas bandas de frecuencia, el análisis de este tipo de señales presenta grandes desafíos. Son señales de muy baja relación señal-ruido (SNR) y sujetos a múltiples artefactos (oculares, musculares, interferencias eléctricas) que pueden sesgar el estudio si no se realiza un preprocesamiento cuidadoso. Dentro de esta tesis, se desarrollan metodologías avanzadas para filtrar y limpiar estas señales, combinando técnicas clásicas de procesamiento de la señal e inteligencia artificial. Además, se utilizan técnicas estadísticas, como por ejemplo, análisis de componentes independientes para identificar y eliminar artefactos, o utilizando filtrados de banda bien definidos que preserven las frecuencias de interés.

Con el objetivo de detectar patrones neuronales de la dislexia, se exploran diferentes métricas de conectividad funcional. Además de la coherencia espectral y del acoplamiento fase-fase, se pone un énfasis especial en el acoplamiento fase-amplitud (PAC). El PAC permite determinar en qué medida la fase de una oscilación de baja frecuencia modula la amplitud de una oscilación de alta frecuencia en regiones concretas del cerebro. Este tipo de acoplamiento es relevante para la integración de información a múltiples escalas temporales, ya que las oscilaciones lentas, como las bandas delta y theta, suelen actuar como marcos de referencia que organizan la actividad de oscilaciones más rápidas en bandas alfa, beta o gamma. La hipótesis de trabajo es que, en la dislexia, ocurriría un desajuste en esta modulación, lo cual afectaría la capacidad de codificar y procesar de manera fluida los rasgos fonológicos del lenguaje, originando o reforzando las dificultades lectoras.

Esta perspectiva de estudiar el procesamiento del lenguaje desde la sincronía neuronal está sostenida teóricamente por marcos como el llamado Temporal Sampling Framework, el cual enfatiza la importancia de las oscilaciones de baja frecuencia en la segmentación del discurso. Según este marco, el cerebro utiliza dichas oscilaciones para adaptar su actividad al ritmo del habla. Cuando la señal de habla presenta picos de energía que coinciden con sílabas o con pausas naturales en el discurso, las oscilaciones neuronales se ajustan a esos tiempos y facilitan así la extracción de unidades fonológicas. En niños con dislexia, se ha observado que esta sincronización se ve interrumpida o es menos estable, lo cual supone una dificultad adicional para identificar las fronteras silábicas o para asociar las letras con sus correspondientes sonidos.

La tesis también introduce la importancia de la lateralización cerebral. Desde la

tradición clásica, se ha sostenido que el lenguaje se encuentra predominantemente en el hemisferio izquierdo, con áreas como las de Broca y Wernicke relacionadas con la producción y la comprensión lingüística, respectivamente. Sin embargo, estudios recientes con técnicas de neuroimagen han revelado que la actividad lingüística comprende redes muy distribuidas y que, además, factores como la dislexia pueden correlacionarse con patrones atípicos de lateralización. Algunos individuos con dislexia podrían mostrar un menor predominio del hemisferio izquierdo o, incluso, una participación excesiva del hemisferio derecho, lo cual se reflejaría en un equilibrio diferente de la actividad neuronal durante tareas de lectura o de procesamiento fonológico. Para capturar este fenómeno, dentro del trabajo se plantean modelos computacionales que comparan la actividad de ambos hemisferios y que buscan encontrar diferencias estadísticas en la forma en que la fase y la amplitud de las oscilaciones se acoplan dentro de cada hemisferio o entre ambos.

Con la finalidad de traducir estos hallazgos en herramientas diagnósticas prácticas, la tesis se adentra en el uso de algoritmos de aprendizaje automático y, de manera más sofisticada, en el empleo de redes neuronales profundas (Deep Learning). A través de estos modelos, se busca la clasificación automática de sujetos entre aquellos que presentan dislexia y aquellos sin dificultades lectoras, basándose en los patrones de conectividad funcional extraídos de las señales EEG. Se exploran tanto redes convolucionales, que son eficientes para reconocer patrones espacio-temporales y que pueden utilizarse para datos EEG multicanal, como redes recurrentes, que resultan adecuadas para el tratamiento de secuencias de datos y, por ende, para la captura de dependencia temporal en la actividad cerebral. De igual modo, se diseñan arquitecturas híbridas que incluyen capas de atención, las cuales permiten asignar ponderaciones distintas a canales o segmentos concretos de la señal según su relevancia para la clasificación. Dichas capas de atención, además, facilitan la interpretabilidad del modelo, ya que dejan entrever cuáles son los rasgos de la señal que tienen mayor impacto en la decisión final de diagnóstico.

El aspecto de la interpretabilidad resulta primordial en aplicaciones clínicas, pues, si bien los modelos de redes neuronales profundas pueden ofrecer altos porcentajes de acierto, la práctica médica requiere justificaciones claras de cada diagnóstico. Los profesionales de la salud necesitan comprender el proceso que ha seguido un modelo para realizar una predicción (dislexia o no), y qué variables o bandas de frecuencia están contribuyendo a dicha predicción. Para ello, en esta tesis se utilizan metodologías de inteligencia artificial explicable (XAI), entre las que destaca el uso de métodos como SHAP (SHapley Additive exPlanations). Estas técnicas descomponen la salida de la red en contribuciones individuales de cada característica, de tal manera que se puede trazar un mapa que muestre la importancia de cada canal de EEG o de cada métrica de conectividad en la predicción. Con ello, se provee no solo un diagnóstico objetivo, sino también un conjunto de indicadores que orientan al clínico sobre cuáles regiones cerebrales o cuáles patrones rítmicos están desempeñando un papel central en las dificultades lectoras.

La tesis, además, evalúa la posibilidad de que esta clase de biomarcadores y modelos explicables ayuden a la detección temprana de la dislexia. Puesto que las evaluaciones conductuales convencionales requieren que los sujetos (en el caso de esta tesis, además, son niños) hayan alcanzado un determinado nivel de competencia lectora para ser

evaluados, se considera que, si se detectan anomalías en la sincronización neuronal o alteraciones de la conectividad funcional desde tempranas etapas (incluso antes de que la lectura esté plenamente desarrollada), se podrían diseñar estrategias de intervención y rehabilitación más eficaces y personalizadas. Esto cobra particular relevancia al tener en cuenta que la dislexia, al no ser diagnosticada, puede derivar en problemas de autoestima, desmotivación escolar y fracaso educativo.

En los aspectos metodológicos, la tesis describe detalladamente el proceso de adquisición de la señal EEG, recalcando la necesidad de seguir protocolos estrictos de colocación de electrodos, control de las condiciones de grabación y reducción de ruido ambiental. Tras la recolección de los datos, se aplican los pasos de preprocesamiento, incluyendo filtros pasa-banda para conservar el rango de frecuencias de interés, algoritmos de análisis de componentes independientes para identificar y eliminar componentes asociados a parpadeos, movimientos oculares o tensiones musculares, y segmentaciones de la señal para aislar los periodos de actividad cortical relacionados específicamente con la tarea lingüística o con la estimulación auditiva aplicada. En algunos de los protocolos experimentales, se presentan estímulos auditivos rítmicos modulados en amplitud a frecuencias propias del habla, con el fin de examinar si los sujetos con dislexia pueden sincronizar sus oscilaciones neuronales a dichos ritmos. La evidencia empírica sugiere que, en el caso de los niños con dislexia, la sincronización delta y theta con el estímulo se ve mermada, lo que podría explicar sus dificultades para identificar límites silábicos y para realizar una correcta correspondencia entre grafemas y fonemas.

La tesis también se adentra en un estudio pormenorizado de la fase, la amplitud y la coherencia entre distintas bandas de frecuencia del EEG. De especial interés son las correlaciones fase-fase (que miden en qué medida dos señales de regiones distintas comparten la misma fase en cada momento) y las correlaciones fase-amplitud (que determinan cómo la amplitud de las oscilaciones rápidas se ve modulada por la fase de las oscilaciones lentas). Para procesar estos datos, se requiere no solo de algoritmos robustos de análisis de series temporales, sino también de métodos estadísticos que ayuden a validar la relevancia de los hallazgos. Se utilizan, por ejemplo, pruebas de permutación para corroborar que los patrones de conectividad o de sincronía difieren de manera significativa entre los grupos con dislexia y los grupos de control. Igualmente, se aplican modelos probabilísticos, como el modelado bayesiano para la identificación de anomalías, que posibilitan cuantificar de forma más flexible las diferencias en los patrones de conectividad.

El estudio se complementa con el análisis de cómo estas anomalías se traducen en la lateralización atípica. Algunos modelos de redes profundas descritos en la tesis están diseñados específicamente para comparar la actividad de ambos hemisferios, estimar índices de lateralización y examinar cómo varían en función de la tarea o del tipo de estímulo lingüístico. Se utiliza, por ejemplo, una arquitectura de red neuronal siamés que recibe en paralelo la señal EEG del hemisferio izquierdo y del derecho, con el fin de aprender una representación conjunta que permita clasificar la dislexia tomando como insumo el grado de desbalance entre hemisferios. Los resultados señalan que, en efecto, los sujetos con dislexia muestran menor predominio del hemisferio izquierdo y exhiben patrones de conectividad más simétricos o atípicos, lo cual concuerda con diversas teorías que habían postulado una lateralización menos marcada en este trastorno.

Los resultados obtenidos con los métodos desarrollados en esta tesis reportan,

de manera consistente, que la combinación de métricas de conectividad (coherencia, acoplamiento fase-fase y fase-amplitud) con redes neuronales profundas diseñadas para reconocer patrones espacio-temporales incrementa notablemente la precisión diagnóstica en comparación con enfoques más tradicionales de aprendizaje automático (como SVM o árboles de decisión). Al implementar capas de atención, se mejora la interpretabilidad de los modelos, pues se puede visualizar qué canales o qué ventanas temporales dentro de la señal EEG resultan más relevantes para distinguir a los sujetos con dislexia de los sujetos de control. Esto tiene implicaciones importantes en entornos clínicos, ya que los profesionales podrían focalizarse, por ejemplo, en la actividad de regiones temporoparietales o frontales, o bien en intervalos de tiempo específicos vinculados a la percepción de sílabas o al procesamiento fonológico. Esta claridad explicativa no solo potencia la aceptación de las herramientas de inteligencia artificial en contextos médicos, sino que también estimula la colaboración entre investigadores, clínicos y desarrolladores de tecnología para la creación de protocolos de intervención y rehabilitación neuronal basados en retroalimentación acerca de la actividad cerebral.

Las aportaciones centrales de la tesis consisten, por tanto, en desarrollar metodologías avanzadas para preprocesar y analizar señales EEG centradas en la detección de biomarcadores de dislexia, aplicar y optimizar modelos de aprendizaje profundo para la clasificación y la estimación de la conectividad funcional en bandas relevantes al lenguaje, y proponer métodos de inteligencia artificial explicable que garanticen la transparencia y la interpretabilidad necesarias en el ámbito clínico. Todo ello concluye con el planteamiento de que la integración de la neurociencia cognitiva, el procesamiento de señales y la inteligencia artificial es un camino prometedor para transformar la manera en que se diagnostican y tratan los trastornos del lenguaje. De esta convergencia surge la posibilidad de identificar marcadores cerebrales tempranos de dislexia y, al mismo tiempo, ofrecer a los especialistas herramientas objetivas, fiables y explicables que sirvan de apoyo a la práctica clínica.

Además de los hallazgos con aplicación directa para la dislexia, el conocimiento generado en esta tesis proporciona métodos que podrían extrapolarse a otros trastornos del desarrollo, como el trastorno por déficit de atención e hiperactividad (TDAH) o los trastornos del espectro autista, que también implican alteraciones en la conectividad cerebral y en los ritmos de procesamiento cognitivo, o incluso a enfermedades neurológicas como la enfermedad de Alzheimer, la enfermedad de Parkinson o la epilepsia. La flexibilidad de los algoritmos de aprendizaje profundo y la potencia de las métricas de sincronización neuronal pueden permitir, si se diseña adecuadamente la experimentación, la elaboración de biomarcadores diferenciados y altamente específicos para cada trastorno. Así, se contribuye no solo al entendimiento y tratamiento de la dislexia, sino al refinamiento general de las estrategias de diagnóstico de diversa índole, mejorando la capacidad diagnóstica de diferentes trastornos y enfermedades neurológicas, y por tanto, mejorando la vida la calidad de vida personas y sus familias.

El conjunto de resultados que describe la tesis enfatiza la evidencia de que los sujetos con dislexia presentan deficiencias significativas en la sincronía de fase, en el acoplamiento fase-amplitud y en la capacidad de sus redes neuronales para sintonizarse con los ritmos del habla. Tales déficits se localizan, principalmente, en bandas de baja frecuencia (delta y theta), las cuales desempeñan un rol crítico en la segmentación fonológica. También se observan alteraciones en la lateralización cerebral, en la que la

actividad en el hemisferio izquierdo no resalta con la misma fuerza que en los lectores típicos. El análisis de estos datos se realiza con métodos robustos de aprendizaje profundo, aprovechando redes neuronales convolucionales y recurrentes, así como mecanismos de atención que mejoran la interpretación clínica de los resultados. La investigadora o el investigador responsable del trabajo subraya la relevancia de validar estos modelos con muestras más extensas y de diferentes edades, así como de profundizar en las aproximaciones multimodales que integren registros de EEG con otras técnicas de neuroimagen, como la resonancia magnética funcional o la espectroscopía de infrarrojo cercano, para obtener una panorámica más completa de la dinámica cerebral en la dislexia.

Por otro lado, los métodos, desarrollos, procedimientos y herramientas propuestos permitirían un cambio de paradigma en entornos clínicos, superando el enfoque exclusivamente conductual de las pruebas tradicionales. Una de las fortalezas radica en la posibilidad de estimar la conectividad en distintos niveles de frecuencia y de trazar, gracias a los algoritmos de IA explicable, un mapa de los elementos concretos que determinan la presencia de un patrón disléxico. Con ello se aporta una gran ventaja en la práctica médica, ya que el especialista no únicamente recibiría un veredicto binario de “dislexia o no dislexia”, sino que obtendría además un perfil detallado de las regiones, bandas y tiempos de procesamiento en los que se observa una desviación respecto a la norma. Dicho perfil podría orientar intervenciones específicas, como entrenamientos auditivos rítmicos, ejercicios de conciencia fonológica, estimulación de la atención a nivel neurológico o un acompañamiento pedagógico más focalizado. En este sentido, el trabajo abre vías de investigación futuras que van más allá del diagnóstico, aventurándose en el terreno de la rehabilitación y la neurofeedback: dado que se pueden detectar y cuantificar métricas de sincronía neuronal alteradas, también cabría plantear protocolos que refuercen o reeduchen esa sincronía, midiendo de manera continua la actividad EEG y realimentando al sujeto para que logre una mejor coordinación rítmica con los patrones propios del lenguaje.

Otra innovación que surge de esta tesis es el método de modelado de la lateralización cerebral mediante el diseño de arquitecturas neuronales específicas, como redes siamesas o mecanismos atencionales bilaterales, que comparan directamente la actividad de los hemisferios. Este enfoque constituye un aporte relevante al debate científico sobre la base neurobiológica de la dislexia y cómo los desequilibrios de la lateralización pueden interactuar con el desarrollo cognitivo del lenguaje. Al ofrecer evidencia cuantitativa y algoritmos diseñados para evaluar automáticamente dicha lateralización, la tesis contribuye a un conocimiento más sólido acerca de la complejidad del procesamiento lingüístico y las posibles vías de intervención cuando la lateralización típica se ve comprometida.

En resumen, el conjunto de aportaciones reunidas en este trabajo posiciona la convergencia entre neurociencia y aprendizaje profundo como un pilar fundamental para la innovación en el campo de los trastornos del lenguaje. No se limita a caracterizar la dislexia desde la perspectiva de la conectividad funcional, sino que también se consolida la idea de que la dislexia conlleva déficits en la sincronización neuronal a nivel cortical, que pueden ser objetivamente medidos y explotados por técnicas de inteligencia artificial para proveer un diagnóstico temprano, fiable y explicable. Este panorama es esperanzador, pues fomenta la posibilidad de diseñar intervenciones basadas en

biomarcadores fiables, lo cual podría suponer un paso decisivo para que la comunidad educativa y clínica cuente con herramientas más objetivas que las actuales. De igual forma, se realza el valor de la investigación básica y el desarrollo tecnológico aplicado, al poner de manifiesto que, a través de la integración de disciplinas, se alcanzan resultados que tendrían un impacto directo en la calidad de vida de las personas con dislexia. El enfoque multidisciplinar que combina la adquisición y el análisis detallado de señales EEG, la estimación de la conectividad funcional cerebral y la aplicación de técnicas estasíticas, de inteligencia artificial y especialmente, modelos de aprendizaje profundo con la meta de crear herramientas diagnósticas innovadoras y explicables. Se demuestra que las dificultades lectoras asociadas a la dislexia podrían rastrearse en alteraciones específicas de la sincronización neuronal, en particular en las bandas de frecuencia delta y theta, y en un patrón de lateralización cerebral menos marcado de lo habitual. El empleo de redes neuronales y técnicas de inteligencia artificial explicable potencia la precisión diagnóstica y facilita la comprensión de los mecanismos subyacentes, brindando a los profesionales sanitarios y a los educadores un soporte objetivo que complementa la valoración clínica tradicional. De este modo, se sientan las bases para una transformación progresiva de la forma en que se identifica y aborda la dislexia, abriendo también el camino para futuros estudios que apliquen estos enfoques en otros trastornos del lenguaje o del neurodesarrollo. Esta integración de neurociencia, procesamiento avanzado de señales y aprendizaje automático contribuye a la evolución del diagnóstico asistido por computadora, marcando un hito significativo en la investigación de la dislexia y ofreciendo nuevas perspectivas para mejorar la detección temprana, el tratamiento personalizado y el entendimiento más profundo de los procesos cognitivos que sustentan la lectura y el lenguaje.

Abstract

Integrative Signal Processing and Explainable Artificial Intelligence for Functional Connectivity Modeling in Language Disorders

This thesis explores the integration of advanced signal processing techniques and [Explainable Artificial Intelligence \(XAI\)](#) methodologies to enhance the understanding and modeling of functional connectivity of EEG signals in the context of language disorders, particularly [Developmental Dyslexia \(DD\)](#). Developmental dyslexia is a neurobiological condition affecting reading skills linked to neural synchronization deficits, making it a prime candidate for innovative diagnostic approaches.

The approach uses non-invasive brain recording techniques, mainly [Electroencephalogram \(EEG\)](#), to assess connectivity through [Cross Frequency Coupling \(CFC\)](#) metrics. A methodological framework is established, combining classical signal processing with the design of [Deep Learning \(DL\)](#) architectures, including [Convolutional Neural Network \(CNN\)](#) and [Recurrent Neural Network \(RNN\)](#). These models are designed to capture complex patterns in EEG data while maintaining interpretability through XAI techniques like [SHapley Additive exPlanations \(SHAP\)](#) or attention layers in the network architecture. These methods are also used to evaluate the lateralization effect in language processing.

Key contributions include the development of novel deep learning architectures tailored for [Phase-To-Amplitude Coupling \(PAC\)](#) estimation, classification methods for differential diagnosis, and explainable tools to aid clinicians in understanding EEG-based connectivity metrics. The proposed models demonstrate improved accuracy in identifying dyslexic neural patterns, highlighting the role of neural synchronization deficits in language processing anomalies. Moreover, different entrainment profiles to auditory stimuli were found when studying local and long-range neural coupling by means of CFC.

The findings advance the field of neurodiagnostics by offering robust, interpretable AI-driven methods to study functional connectivity. This thesis sets the stage for early diagnosis and individualized interventions for language disorders, with implications that extend to other neurological conditions.

Acknowledgment

En primer lugar, me gustaría expresar mi más profunda gratitud a mis padres, Marco y Ana, por su apoyo y amor inquebrantables. Su confianza en mí ha sido una fuente constante de fuerza y motivación.

Estoy profundamente agradecido a mi pareja, Rocío, por su paciencia, ánimo, comprensión y por darme en breve el mejor regalo que un hombre puede desear. Su apoyo ha sido inestimable y no podría haber alcanzado este hito sin ella a mi lado.

También estoy infinitamente agradecido a mi supervisor, Andrés, por su experta orientación, sus comentarios constructivos, su infinita paciencia y su compromiso con mi crecimiento académico y profesional. Sin él esto no hubiera sido posible a ningún nivel.

Agradecer a Pakito el que me reclutara y por enseñarme los primeros pasos en el mundo académico. Los comienzos en algo nuevo siempre son difíciles, pero con él al lado también son divertidos.

A mis amigos, en especial a Victor $\times 2$ y Leyva, por preocuparse durante estos años, estar ahí en los momentos de bajón y tener siempre lista la maquinaria de las risas.

Y por último, pero no por ello menos importante, a todos los miembros del grupo Biosip, especialmente a: Nico, Juan Eloy, Diego e Ignacio $\times 2$. Sin ellos este camino hubiera sido demasiado aburrido y mucho más complicado.

Asimismo, agradecimientos al grupo Leeduca de la Universidad de Málaga por la adquisición de los datos y el soporte prestado para la realización de este trabajo.

Funding

This thesis has been supported by Marco A. Formoso PRE2019-087350 grant, funded by MICIU/AEI/10.13039/501100011033 «ESF Investing in your future», in the framework of the research projects PGC2018-098813-B-C32 and PID2022-137461NB-C32 funded by MICIU/AEI/10.13039/501100011033 and FEDER/UE (Ministerio de Ciencia, Innovación y Universidades y fondos FEDER), as well as UMA20-FEDERJA-086 project (Proyectos de I+D+i en el marco del Programa Operativo FEDER. Junta de Andalucía) and TIC251-G-FEDER (BioSiP) project, funded by ERDF/EU.

Acronyms

AAC Amplitude-To-Amplitude Coupling. 16

AI Artificial Intelligence. 1, 5, 6, 20, 21, 45, 118

AUC Area Under the Receiver Operating Characteristic Curve. 67, 70

BCI Brain Computer Interface. 14, 24

BMU Best Matching Unit. 93–96

CFC Cross Frequency Coupling. xv, 11, 16, 35, 90

CNN Convolutional Neural Network. xv, 60, 117, 118, 125

DD Developmental Dyslexia. xv, 1, 2, 6, 16, 81, 87, 139

DL Deep Learning. xv, 6, 20, 117, 118

EEG Electroencephalography. 23

EEG Electroencephalogram. xv, 1, 6, 14, 15, 21, 23, 45, 106, 118

FIR Finite Impulse Response. 121

fNIRS Near-Infrared Spectroscopy. 1, 21

GPU Graphical Processing Unit. 1, 60

HT Hilbert Transform. 35

ICA Independent Component Analysis. 30

LSTM Long Short-Term Memory. 62, 117, 118

MEG Magnetoencephalography. 1

MI Modulation Index. 37

MLP Multi Layer Perceptron. 128

- MRI** Magnetic resonance imaging. [118](#)
- MVL** Mean Vector Length. [16](#), [37](#)
- PAC** Phase-To-Amplitude Coupling. [xv](#), [6](#), [7](#), [16](#), [36](#), [106](#)
- PLV** Phase-Locking Value. [16](#), [38](#), [39](#)
- PPC** Phase-To-Phase Coupling. [16](#), [38](#)
- Qe** Quantization Error. [93](#), [95](#), [97](#)
- Qe** Quantization Error. [94](#)
- RNN** Recurrent Neural Network. [xv](#), [60](#), [117](#), [118](#)
- SC** Spectral Coherence. [90](#)
- SHAP** SHapley Additive exPlanations. [xv](#), [76](#)
- SLI** Speech and Language Impairment. [81](#)
- SNR** Signal to noise ratio. [29](#)
- SOM** Self-organizing map. [7](#), [89](#), [92](#), [93](#), [95](#), [139](#)
- TSF** Temporal Sampling Framework. [4](#), [11](#)
- XAI** Explainable Artificial Intelligence. [xv](#), [6](#), [76](#)

El maná no cae del cielo.

Contents

Declaration of Original and Sole Authorship	iii
Authorization by the supervisor for the thesis defense	v
Resumen	vii
Abstract	xv
Acknowledgment	xvii
Acronyms	xix
Contents	xxiii
List of Figures	xxvii
List of Tables	xxxix
1 Introduction	1
1.1 Aims and Objectives	6
1.2 Organization of this Thesis	7
1.3 Contributions	7
1.3.1 Articles	7
1.3.2 Conferences	8
2 Foundations	11
2.1 Neural synchronization in sensory perception	11
2.2 Language processing models	13
2.3 Electroencephalography	14
2.3.1 Neural Synchronization	15
2.3.2 Temporal Sampling Framework: synchronization in language processing	16
2.4 Artificial Intelligence in healthcare	20

3	Electroencephalography and functional connectivity	23
3.1	The electrical brain	24
3.2	EEG acquisition	25
3.2.1	Artifacts in EEG signals	29
3.3	EEG Preprocessing	30
3.4	Connectivity from EEG signals	32
3.4.1	Connectivity Estimation and Modeling	33
3.4.2	Surrogate Analysis	40
4	Artificial Intelligence	45
4.1	Machine Learning	46
4.1.1	Training Process	47
4.2	Overview of Machine Learning Algorithms	49
4.2.1	Supervised Learning	49
4.2.2	Unsupervised Learning	57
4.3	Deep Learning	58
4.3.1	Convolutional Neural Networks	60
4.3.2	Recurrent Neural Networks	60
4.3.3	Activation functions	63
4.3.4	Training	64
4.4	Model Evaluation	67
4.4.1	Confusion Matrix	67
4.4.2	Area Under ROC Curve	69
4.5	Model Validation	70
4.5.1	Cross Validation	73
4.5.2	Permutation Test	73
4.5.3	Permutation Test vs Cross Validation	75
4.6	Explainability	76
5	Database	81
5.1	Study participants	81
5.2	EEG Recording Protocol	83
5.3	EEG Preprocessing	84
5.4	Stimuli and Language Processing	85
5.4.1	Relevant EEG Bands in language disorders	85
6	Phase-synchrony Connectivity Anomalies in EEG signals	89
6.1	Functional Connectivity from EEG signals	90
6.2	Pearson's Circular Correlation	91
6.3	Outlier Detection	92
6.3.1	Self-Organizing Maps	92
6.3.2	Band relevance using quantization error distribution	93
6.3.3	Bayesian Anomaly Detection for SOM	96
6.4	Results	97
6.5	Discussion	100

7	EEG Entrainment to Speech Rhythms	105
7.1	EEG Entrainment and Auditory Steady-State Response	105
7.2	CFC-based synchronization between EEG and auditory stimuli	106
7.3	Significance Testing via Permutation	107
7.4	Feature Aggregation and Classification	108
7.5	Evaluation and Interpretability	108
7.6	Results	112
7.7	Discussion	114
8	Cerebral Lateralization Assessment using Deep Learning	117
8.1	Brain Lateralization	118
8.2	Siamese Networks	119
8.3	Synthetic EEG signal generation	120
8.4	Siamese Architecture for Lateralization Analysis	124
8.4.1	PAC estimation with Deep Learning	125
8.4.2	Channel Attention Layer	130
8.4.3	Siamese network	132
8.5	Results	133
8.6	Discussion	134
	Conclusion	139
	References	143

List of Figures

1.1	Spoken language segmentation for extracting different phonemic units. Adaptation from [144]	4
2.1	Sensorial Processing. During brain development, neurons in the cerebral cortex cluster into regions specialized for sensory processing. Brain networks generate abstract economic representations (prototypes). Figure by P.J. López and A. Ortiz	12
2.2	Wernike-Geschwind functional model. Areas involved in language production and comprehension [179]	13
2.3	EEG Example	15
2.4	PAC Example	16
2.5	Brainwaves	19
3.1	When dealing with EEG data, we should consider volume conduction. In volume conduction, one event can be captured by multiples electrodes at the same time, resulting in spurious correlations.	25
3.2	International 10-20 system.	26
3.3	EEG Montage	27
3.4	Common EEG setups. (a) Bipolar (a) and (b) Unipolar.	28
3.5	Examples of EEG artifacts [146]	29
3.6	EEG Preprocessing Pipeline	31
3.9	Phase-Amplitude Coupling Examples	40
3.7	Hilbert Transform	43
3.8	PLV	44
3.10	Histograms showing the null distribution of a connectivity metric with the observed metric value (red line) and corresponding z-scores. (a) represents a low z-score ($z = 0.3$), indicating no significant deviation from the null distribution, while (b) shows a higher z-score ($z = 2.3$), suggesting statistical significance. The z-scores are displayed rotated and positioned along the red lines for clarity.	44
4.1	AI Field	45
4.2	Models explainability. Opening the black box.	46
4.3	Supervised Learning	47
4.4	Supervised Learning	48
4.5	Illustration of a decision tree. Each decision split the data in several branches depending on a feature value [151]	49

4.6	Illustration of a boosting ensemble. Each new classifier tries to correct the errors in the previous one [70]	50
4.7	KNN classification task. Knowing the K nearest neighbors points, the new point will be classified as the same class as the majority of those K nearest neighbors.	53
4.8	Hyperplane (red) solution for a SVM. The margins are maximized between the plan and the closest datapoints. The datapoints bolded are the support vectors [118]	55
4.9	In the kernel trick, the input data is transformed into a higher dimensional space where a hyperplane can separate the data. Different types of kernels result in different types of spaces [214]	55
4.10	In one class SVM the origin is taken as a negative class sample. Anything that falls outside the boundary of decision of the positive class is classified as an outlier (ξ_i, ξ_j) [113]	58
4.11	Perceptron, the basic element of the neural network. The inputs are multiplied by the weights (the learnable parameters) and summed together. Then, the previous result is passed through an activation function, giving the final result.	59
4.12	The multilayer perceptron is an evolution of a perceptron capable to deal with non-linearities. There is no inherent limit to how many layers or neurons per layer can be included in the network structure.	60
4.13	(a) A typical CNN network. In each layer, the network learns to extract a hierarchical feature of the training images. (b) The convolution operation used in CNNs. The filter parameters are what the network try to learn from the data [76]	61
4.14	The recurrent network is a network block in which the input vector x is feed iteratively to the architecture. A hidden state h is maintained in every step, working as <i>memory</i> . [105]	62
4.15	The LSTM block architecture showing input, forget, and output gates, along with the cell state [132]	63
4.16	Three of the most used activation functions: a) Sigmoid, b) Tanh, c) ReLu. These functions introduce the non-linearity part that allows the networks to learn complex patterns.	64
4.17	Learning rate examples and its outcomes [111].	66
4.18	Example of a confusion matrix for a binary example	68
4.19	(a) AUC example. (b) Distribution of TN, FP, FN, TP based on the chosen threshold. Every point in the AUC represents the True Positive Rate against the False Positive Rate for a certain threshold.	69
4.20	Fitting Validation	71
4.21	Bias/Variance tradeoff. When bias is high and variance is low, we are probably underfitting the data, whereas the variance is high but bias is low, we are probably overfitting the data. The goal is to find a model that sits between these two extremes, achieving a balance where the bias and variance are minimized enough to ensure the model generalizes well to new, unseen data, resulting in optimal performance. Source: [141]	72

4.22	Illustration of k-fold cross-validation with $k=5$. The data is split into 5 folds, and the model is trained and validated 5 times, each time using a different fold as the validation set. Source: [180]	74
4.23	Permutation Test Distribution. The histogram represents the distribution of performance metrics (AUC) obtained from models trained on permuted data. The vertical red line indicates the observed performance from the model trained on the original data. A p-value is calculated as the proportion of permuted performance metrics equal to or exceeding the observed performance.	76
4.24	Example of SHAP summary plot.	78
5.1	Screenshot of the Leeduca platform login page. https://leeduca.es/	82
5.2	Laboratory view. actiCAP headset and NIRS acquisition systems	83
5.3	White noise modulated at different frequencies as used in the experiment	84
6.1	SOM showing different classes. Each input data instance is represented by the model vector of its corresponding BMU [210].	94
6.2	Quantization error distribution for Controls and Dyslexia subjects. The (a) 4.8 Hz and (b) 16 Hz stimulus.	95
6.3	SOM map activation for each data sample, according to the corresponding mean quantization error. Example for EEG Delta band using 4.8 Hz stimulus. (a) Control (b) Dyslexia.	96
6.4	Statistically significant connections for meaningful bands. The axes represent the channels names according to the extended 10–20 EEG montage. Only connections with FDR-corrected $p < 0.01$ are shown. (a) 4.8Hz and (b)16 Hz stimulus.	98
6.5	Classification results for a) 4.8 Hz and b) 16 Hz.	99
6.6	ROC Curves corresponding to the classification results for (a) 4.8 Hz and (b) 16 Hz stimulus.	99
6.7	Permutation test results to assess the statistical significance of the results. Blue distribution corresponds to the null distribution and red line corresponds to the accuracy obtained with true labels. (a) 4.8 Hz stimulus and (b) 16 Hz stimulus.	100
7.1	PAC aggregation to compute the final score for a subject. For each segment and every band a ZPAC score is computed and then averaged across all segments.	109
7.2	10-20 placement of the EEG electrodes in the Lambert azimuthal representation. Below: brief explanation of the azimuthal projection of the region areas.	111
7.3	Comparison of performance measures, a) balanced accuracy and b) sensitivity, in experiments with different hyperparameters (C , penalty) for each stimulus and PAC estimation method.	112
7.4	Parallel plot tracing the performance of the different combinations of stimuli, PAC measures, and SVC parameters: penalty and C .	113

7.5	Classification weights of the linear SVC trained to discriminate between DLX and CTL subjects using the HR entrainment between the stimuli (16 and 40 Hz) and different bands.	113
7.6	SHAP values for the DLX class using Gamma-average HR PAC for the 40 Hz stimulus. Color indicates the actual PAC value.	114
8.1	A Siamese network is actually the same model in both situations. But the input is considered in pairs to apply some special loss function between the output vectors, commonly called embeddings.	119
8.2	EEG Power Law	121
8.3	EEG Colored Noise	121
8.4	Synthetic generation	123
8.5	Siamese Neural Network. This network is able to process the two brain hemispheres at the same time while selecting the more important channels.	124
8.6	Neural Network for PAC calculation. The parameters of this network are frozen once trained, and only the layers until the embeddings are used in the siamese network (red part).	125
8.7	In 1D Convolution a segment of the signal is multiplied by a kernel resulting in a single value. This value contains information about the whole segment while reducing the signal length.	126
8.8	Depiction of a working LSTM layer. Each blue rectangle is an LSTM unit. We can see where each output and input are connecting inside the network.	127
8.9	Distribution of the strength value α of the synthetic signals vs the output of the network. The network is able to correctly stimate the level of synchrony of the signals. This value is highly correlated with PAC metrics like MVL, PLV. The higher the strength, the higher these metrics.	129
8.10	The channel attention layer facilitates channel selection by processing input EEG channels through multiple Conv1D and Average Pooling layers, ending in a Softmax layer. This final layer represents each channel's significance, enabling the consolidation of all channels into a single, information-rich channel by multiplying each by its importance and summing the results.	130
8.11	Best permutation results a)4.8 Hz, b)16 Hz and c) 40 Hz) stimuli. Null distributions obtained over 5000 permutations are shown in blue. The red line represents the AUC obtained when using the actual labels.	135
8.12	Topographical maps showing the significant differences ($p < 0.05$) between controls and dyslexics in channel importance computed by the attention layer for the a) 4.8 Hz, b) 16 Hz and c) 40 Hz stimuli used in this work.	136

List of Tables

2.1	Summary of results from different studies on dyslexia and brainwave synchronization.	17
5.1	EEG Preprocessing Pipeline	85
6.1	Summary of the best classification results obtained for the different EEG bands. Bands providing the best accuracy values are highlighted in bold.	97
6.2	Comparison of the proposed method with state-of-the-art methods for ADHD classification. The * denotes that the value was not reported in the original paper. The best results are highlighted in bold.	102
7.1	Comparison of the performance of this and other state-of-the-art methodologies for the differential diagnosis of DLX.	115
8.1	Results obtained for different input stimuli and brainwaves. The * denotes no permutation test were done for that band combination.	133
8.2	Comparison among works implementing different techniques for dyslexia classification. *Data not reported by the source.	134

Chapter 1

Introduction

The integration of computer science and medicine has transformed healthcare, most notably treatment plans and diagnosis. New diagnostic techniques and machines have been developed at the same time as computing power has increased. Think for example of [Electroencephalogram \(EEG\)](#), [Near-Infrared Spectroscopy \(fNIRS\)](#) or, [Magne-toencephalography \(MEG\)](#) and the amount of computer power and storage capacity required to acquire them. Extracting insights from the data also demands a substantial amount of computational resources; before acquiring any information, it is necessary to detect and eliminate signal artifacts, prepare the signal and compute pertinent statistical data or measures. Additionally, a thorough expert evaluation plays a role in assisting with the diagnosis, being susceptible to inaccuracies and frequently consumes a significant amount of time. Currently, being able to process all this data and construct some statistical measure was all there was with the computing power of the moment. It has taken another breakthrough before computers can also be used to accurately assist the experts.

In recent years, the world has witnessed the rise of [Artificial Intelligence \(AI\)](#). Although most algorithms have been with us since the last century, it is only now, with the exponential development of computing power, dedicated hardware such as [Graphical Processing Unit \(GPU\)](#), and massive data processing, that we have been able to use these algorithms to their full potential. [AI](#) has transformed the field of medical diagnostics, enhancing the capabilities of traditional methods and developing new ways to perform it. The evolution of artificial intelligence is intricately linked to the growth of computational power, which serves as the backbone for driving [AI](#) innovations across various domains. The synergy between human expertise, robust computing hardware, extensive datasets, and sophisticated algorithms underscores the transformative potential of [AI](#) technologies. As computational power continues to advance, the capabilities of [AI](#) systems are poised to expand, revolutionizing industries and enhancing decision-making processes across diverse sectors.

The irruption of [AI](#) in healthcare is undeniable and a great advantage for both professionals and patients. As this field grows, new opportunities arise to better understand disorders or diseases that were previously very difficult to analyze or even absolute mysteries. The case of [Developmental Dyslexia \(DD\)](#) is one of the disorders that benefit from the advantages of [AI](#), since it is capable of capturing the subtleties that are generated in the brains of people suffering from this disorder.

DD is a specific difficulty in the acquisition of reading skills, of neurobiological and genetic origin and not related to the intellectual quotient, mental age, or inadequate schooling. Its prevalence is estimated between 5% and 14% of the population [161], depending on the reference used. It has serious effects on the self-esteem of affected children. Early diagnosis and prognosis to initiate an appropriate, early, and individualized intervention are decisive for the personal and intellectual development of these children. Dyslexia prevalence is estimated to be the cause of more than 40% of school failure [140]. Therefore, it is a problem with a significant socioeconomic impact.

Currently, the causes and neurobiological processes of DD are unknown and are usually diagnosed through specifically designed tests to measure different behavioral variables involved in the reading process. Examples of these variables are reading efficiency or the ability to split words into their constituent syllables. These tests, performed by specialists, require considerable time to be thoroughly analyzed and to issue an accurate diagnosis, as well as to determine the most appropriate intervention. On the other hand, the diagnosis is usually based on cutoff points calculated, in most cases, on a population not sufficiently extensive. Additionally, test results depend on the motivation and mood of the subject when performing the reference tasks. Consequently, classical diagnostic methods require considerable time for manual analysis and are prone to error, often leading to children with specific learning difficulties not being correctly diagnosed or treated, affecting their cognitive and emotional development. Furthermore, most reference criteria are designed for readers, which limits the minimum age for early diagnosis. For this reason, research aimed at obtaining results that allow for early diagnosis and individualized intervention would have a great impact.

Dyslexia is often associated with impaired phonological processing, which involves recognizing, identifying, and manipulating phonological structures within words [82, 208]. The causes of developmental dyslexia are multifaceted and may involve deficits in auditory and speech perception processes. Some studies point to genetic factors [99][170]. Moreover, neuroanatomical precursors identified in the brain from pre-reading stages through age 11 have been linked to dyslexia [33]. Also, it has been suggested that reduced and suboptimal reading experience may lead to deficits associated with dyslexia [32].

The above has led to significant research activity in the search for objective markers that improve diagnostic accuracy and, eventually, reveal unknown aspects of DE related to its neural basis. Furthermore, research into the biological causes of dyslexia can provide valuable information for a better understanding of how controls and dyslexics process auditory information and, thus, language, with special application in designing individualized intervention tasks. However, studying language processing in the brain to find markers that highlight differences between controls and subjects with a disorder (such as dyslexia) is a complex task involving several disciplines. On the one hand, it is necessary to have theoretical or semi-empirical models of language processing to guide the search for markers. This knowledge is essential to determine the biomedical signals to acquire, the specific acquisition technique to use, and to establish the type of stimulus from which to study its response. On the other hand, it is necessary to develop specific signal processing techniques capable of capturing differential information between control subjects and dyslexics. Finally, marker selection, pattern recognition, and, in general, methods that allow acquired biomedical information to be transformed

into knowledge rely on interpretable artificial intelligence techniques that explain the models.

For studying brain activity related to cognitive processes, non-invasive techniques for recording local cerebral metabolic activity are often used. These techniques can be classified into two groups:

- Techniques providing temporal resolution (1-3 ms resolution), EEG and MEG. EEG records the electrical activity generated as a consequence of postsynaptic potentials of neurons. The ion flow between the neuron and the extracellular space activated by a particular neurotransmitter produces a potential difference (dipoles) that generates an electric field around the neuronal body. Specifically, EEG records the signal due to radially oriented dipoles. The electric field generated by a single neuron is too weak to be measured by a contact electrode placed on the skull. However, it is possible to measure the field generated by thousands or millions of neurons synchronously activated, normally considering 5 frequency bands: Delta (0.5-4 Hz), Theta (4-8 Hz), Alpha (8-12 Hz), Beta (12-20 Hz), and Gamma (>20Hz). The activity recorded by MEG is basically the same as with EEG, but MEG relies on recording magnetic activity generated by intracellular currents (excitatory and inhibitory postsynaptic potentials, tangential orientation dipoles) and provides superior spatial resolution to EEG. In both cases, cortical activity is recorded, providing temporal resolution on the order of ms.
- Techniques providing spatial resolution. Among the non-invasive techniques most used to study brain functional activity is functional magnetic resonance imaging (fMRI), based on detecting local variations in blood flow. Thus, fMRI allows obtaining an image that reflects local oxygen levels. This technique enables monitoring activity in subcortical regions. The fMRI functional image provides spatial resolution between 1 to 3 mm. More recently, NIRS is based on analyzing the reflectance spectrum of NIR radiation (between 750nm and 2600nm wavelength). Thus, it is possible to measure changes in oxygenated and deoxygenated hemoglobin levels using the Beer-Lambert law. Using this technique to study functional activity has led it to be called fNIRS (functional NIRS). Although spatial resolution is inferior to fMRI, the acquisition cost is much lower, making it, in some cases, a compromise solution between fMRI and EEG.

The analysis of the auditory signal of spoken language reveals that the different units that compose it are produced regularly over time, following various rhythms corresponding to the temporal characteristics of each phonological unit [71]. Thus, while phonemes are produced at a frequency between 20 and 40 Hz, syllables and words are produced at a frequency between 4 and 7 Hz. The slower rhythms (0.5-2 Hz) correspond to prosody or intonation. The exact production frequency of these units depends on each language, hence the establishment of the previous intervals. Speech comprehension requires correctly segmenting the auditory signal, i.e., correctly identifying the boundaries between different phonological units. This segmentation process is carried out in the brain by synchronizing different brain rhythms with the characteristic frequency of production of different phonological units. Oscillations in the Delta band (0.5 – 2 Hz) allow extracting information about prosody, oscillations in the Theta band (4 – 7 Hz)

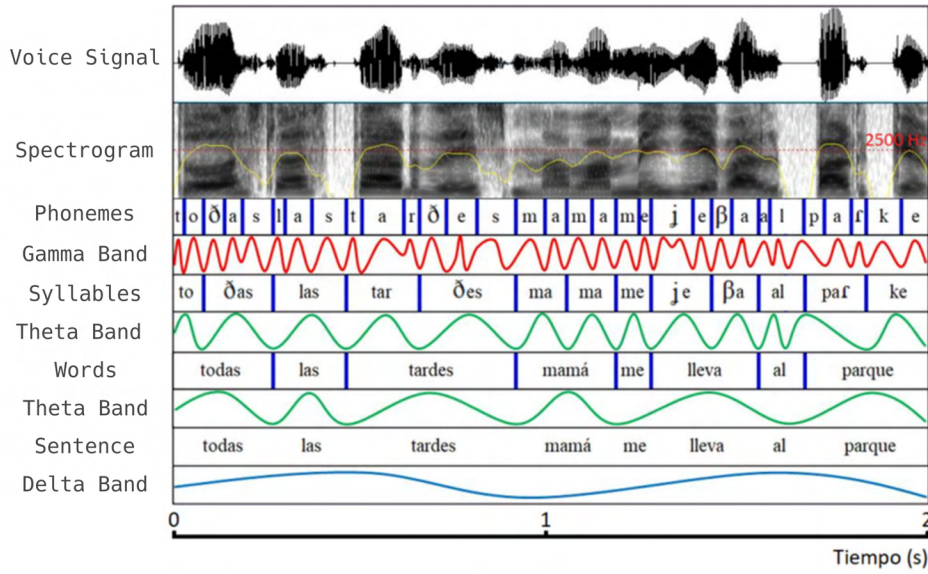


Figure 1.1: Spoken language segmentation for extracting different phonemic units. Adaptation from [144]

allow extracting words and syllables, and oscillations in the Gamma band (25 – 40 Hz) allow extracting phonemes [71, 74, 90, 163]. Furthermore, the auditory regions of the left and right hemispheres play different roles in speech segmentation. While the right hemisphere is responsible for segmenting slow units via Delta and Theta rhythms, the left hemisphere handles fast unit segmentation (phonemes) through Gamma rhythms [163]. This synchronization process occurs between groups of neurons and is known as neuronal synchronization or neural entrainment. Fig. 1.1 shows an example of spoken language segmentation for extracting different phonemic units. The previously described segmentation process follows a parallel and distributed processing model, allowing such segmentation processes to be stable and robust in the presence of noise.

The previous synchronization process responds to the theoretical framework developed in [78] and called the **Temporal Sampling Framework (TSF)**, which postulates that the lower frequency bands synchronize with the lower frequency phonemic units, while the higher frequency bands serve to sample segments where phonemes are located. Thus, neuronal activity in the Theta band is phase-locked to the frequency of syllabic units. On the other hand, Gamma band activity is phase-locked to phonemic sounds. The higher-order phonological processing that requires coordinated participation from different brain areas would require a coupling process known as neuronal entrainment. Studies by Poeppel [131] and Simon [52] using EEG and MEG signal recordings show that the brain response synchronizes with the auditory stimulus envelope during spoken language comprehension. Specifically, in [52], this synchronization process is shown in the Delta and Theta bands to capture the syllabic rhythm (necessary to decompose words into their constituent syllables). This suggests the existence of groups of neurons responsible for this task. However, the experiments conducted by Poeppel and Simon do not show whether this synchronization phenomenon also occurs to extract structures at other higher levels of abstraction (words, phrases). In a subsequent experiment, they used sequences of words with syllables of the same duration (250ms) in two languages.

From these words, pseudophrases are generated at two levels (2 Hz and 1 Hz). The analysis of the MEG signals captured during the cited stimulus application allowed observing brain activity synchronized at 4 Hz (syllabic rhythm) in all cases. However, activity at 2 Hz and 1 Hz was only observed when the stimulus corresponded to the subject's native language. Two important conclusions can be drawn from this study: 1) the brain can generate abstract representations from unknown combinations of words (pseudophrases) and 2) only the first level (phonological) is language-independent, so this representation capacity lies in specific connections between groups of neurons that map the language's semantic rules (generated through a learning process). The model proposed in [92] suggests that speech processing involves a segmentation process and a prediction process. According to this model, the segmentation process operates directly on auditory stimuli without prior knowledge of the temporal properties of the stimulus [102]. Thus, the coupling between low-frequency bands (Delta - Theta) and the Gamma band allows the hierarchical coordination necessary for encoding/decoding phonemes into syllables. The prediction process that enables continuous speech recognition is generated by an internal model that integrates sensory and motor information and depends on contextual cues to anticipate the content and temporal structure of the input. This prediction process seems to be driven by sets of neurons oscillating in the Beta band. Therefore, an atypical neuronal entrainment at slow frequencies should produce cascading effects in faster bands, affecting, for example, the correct sampling of phonemes.

The aforementioned studies have provided considerable knowledge about the brain processes responsible for language segmentation. However, it has only been in recent years that theoretical and empirical relationships between neurodevelopmental disorders, such as developmental dyslexia, and deficits in the synchronization between brain rhythms in certain brain areas and the production frequency of different phonological units have been established [78]. Although studies using functional brain activity acquisition techniques (mainly EEG and MEG) have shown evidence of asynchrony deficits in specific bands (especially in the Delta and Theta bands) in dyslexic subjects, a complete and coherent explanation identifying the effects of such desynchronization on language processing networks is still lacking, as most studies focus on identifying localized asynchrony deficits.

The convergence of classical signal processing techniques and artificial intelligence represents a transformative paradigm in biomedical signal analysis, with particular significance in neurosciences. Traditional methodologies like Fourier transforms, wavelet decomposition, and advanced filtering algorithms provide critical preprocessing mechanisms that normalize and extract fundamental features from intricate neural signals. Machine learning models, particularly deep neural networks and convolutional architectures, can then strategically leverage these preprocessed signals to develop sophisticated pattern recognition and classification frameworks. This interdisciplinary approach enables more robust noise reduction, precise artifact elimination, and multi-dimensional feature extraction from complex EEG datasets. By bridging mathematical signal processing principles with computational intelligence, researchers can achieve unprecedented insights into neurological dynamics, enhancing diagnostic capabilities for conditions ranging from epilepsy and brain-computer interfaces to neurodegenerative disease monitoring. The synergy between signal processing and AI ultimately enables

more nuanced, accurate, and comprehensive neurological interpretation.

1.1 Aims and Objectives

The aim of this thesis is the development of computer aid techniques and tools as well that help in the context of language disorders, and, particularly, in [DD](#). Also, the same techniques can be applied in other areas where the inclusion of [AI](#), [EEG](#) and connectivity is needed:

- Development of new signal processing and [AI](#) methods for estimating functional connectivity by means of synchronization among brain areas using [EEG](#) signals. In the quest for a method to quantify the connectivity, we differentiate between local connectivity and long-range connectivity. The first provides information about the underlying neural (biological) mechanisms that locally process sensorial information. The second shows the interaction between different groups of neurons to integrate the result of local processing.
- Exploration of Connectivity estimation methods based on cross-frequency coupling, and specifically [Phase-To-Amplitude Coupling \(PAC\)](#), which has been proposed the most plausible mechanism for the dynamic coordination of brain activity over multiple spatial scales.
- Development of classification methods with enough accuracy for differential diagnosis, avoiding the problem of the inherent variability of the experimental group, the limited sample size and the data dimensionality. This includes the use of semi-supervised methods such as anomaly detection techniques.
- Development of interpretable and visual-meaningful Artificial intelligence techniques providing exploratory information useful for differential diagnosis of reading difficulties. Explainability is crucial in biomedical signal processing, especially for complex, multi-channel data like EEG. Analytical methods should not only predict but also explain relevant features and their interactions. Enhancing the transparency of machine learning models is vital for understanding brain processes. Visualization and interpretable tools will be developed to illustrate channel interactions.
- Development of new [DL](#) architectures that could capture complex patterns from EEG data without the use of predefined statistical features. The development of deep learning architectures to process EEG signals for inter-channel PAC computation will involve the use of several known neural network layers like CNN, LSTM and the development of new blocks. Additionally, the developed model will be [Explainable Artificial Intelligence \(XAI\)](#), providing information about the model operation with clinical implications.

1.2 Organization of this Thesis

This thesis is organized mainly in three parts along six main chapters. The first part corresponds to chapter 1, where we introduce the motivations and primary aims of the study, exploring the neural basis of DD and providing foundational insights into EEG and its relevance in brain connectivity analysis.

The second part corresponds to methodology used and include chapters 3, 4 and 5. Chapter 3 reviews EEG principles, preprocessing steps, and connectivity metrics essential for understanding brain connectivity patterns associated with DD. In chapter 4 we delve into AI methodologies applicable to EEG analysis, particularly focusing on supervised and unsupervised machine learning, along with deep learning architectures like CNNs and LSTMs. Chapter 5 further details the experimental setup, participant selection, EEG recording protocols, and preprocessing techniques, setting a solid framework for the methodology applied in this study.

The last part correspond to how this methodology is applied to obtain different results. Chapter 8 is dedicated to deep learning approaches for cerebral lateralization analysis, introducing the Siamese neural network architecture and PAC estimation techniques, which help identify patterns specific to dyslexia. This chapter contains one of the main contributions of this Thesis, providing a neural approach for PAC computation and a full architecture for quantifying the lateralization by means of EEG signals. Moreover, we contribute a channel selection layer inspired in the attention principle that considerably reduces the computational burden and provides explainability. Chapter 6 presents phase-synchrony connectivity analysis to detect anomalies in EEG signals associated with DD, utilizing advanced methods such as Self-organizing map (SOM) to extract differential phase-amplitude patterns and Bayesian anomaly detection for accurate classification.

Finally, in the concluding section, we provide a comprehensive discussion of the findings, summarizing the study's contributions to dyslexia diagnostics and proposing future directions for refining connectivity-based EEG analysis through AI advancements.

1.3 Contributions

This section showcases a series of research efforts focused on advancements in the analysis of EEG signals, especially related to dyslexia diagnosis. These contributions, divided into articles and conferences, highlight innovative methodologies for identifying neural connectivity anomalies, leveraging complex network models, deep learning, and explainable AI approaches to enhance the understanding of brain signal processing.

1.3.1 Articles

1. Formoso, M.A., Ortiz, A., Martinez-Murcia, F.J., Gallego, N., Luque, J.L.: "Detecting Phase-synchrony Connectivity Anomalies in EEG signals. Application to Dyslexia Diagnosis". *Sensors*, 21(7061), pp.1-20

2. Gallego-Molina, N.J., Ortiz, A., Martinez Murcia, F.J., Formoso, A., Giménez, A. "Complex network modelling of EEG band coupling in dyslexia: an exploratory analysis of auditory processing and diagnosis". Knowledge-Based Systems. 2022, 108098. pp.1-26.
3. Diego Aquino-Brítez, Andrés Ortiz, Julio Ortega, Javier León, Marco Formoso, John Q. Gan, Juan J. Escobar.:” Optimization of Deep Architectures for EEG Signal Classification: An AutoML Approach Using Evolutionary Algorithms”. Sensors, 21(6), pp. 1-21, 2021
4. Górriz, J.M., Álvarez-Illán, I., Álvarez-Marquina, A., Arco, J.E., Atzmueller, M., Ballarini, F., Barakova, E., Bologna, G., Bonomini, P., Castellanos-Dominguez, G. and Castillo-Barnes, D., 2023. Computational approaches to explainable artificial intelligence: advances in theory, applications and trends. Information Fusion, 100, p.101945.
5. Rodríguez-Rodríguez, I., Ortiz, A., Gallego-Molina, N. J., Formoso, M. A., Woo, W. L. (2023). EEG Interchannel Causality to Identify Source/Sink Phase Connectivity Patterns in Developmental Dyslexia. International Journal of Neural Systems, 33(04), 2350020.
6. Diego Castillo-Barnes, Nicolás J. Gallego-Molina, Marco A. Formoso, Andrés Ortiz, Patricia Figueiredo, Juan L. Luque.:”Probabilistic and explainable modeling of Phase-Phase Cross-Frequency Coupling patterns in EEG. Application to dyslexia diagnosis”. Biocybernetics and Biomedical Engineering. 2024
7. Formoso, M.A., Arco, J.E., Ortiz, A. Q. Gan, J., Rodriguez-Rodriguez, I.: "Cerebral lateralization assessment: an explainable deep learning approach with channel attention mechanism". Under review in IEEE Journal of Biomedical and Health Informatics, 2024.

1.3.2 Conferences

1. Ortiz, A., Arco, J.E., Formoso, M.A., Gallego-Molina, N.J. 1,3, Rodriguez-Rodriguez, I., Martinez-Murcia, F.J., Gorriz, J. M., Ramirez, J.: "Towards mixed mode biomarkers: combining structural and functional information by deep learning". 9th International Work-Conference on the Interplay Between Natural and Artificial Computation. IWINAC 2022
2. Formoso, M.A., Ortiz, A., Martinez-Murcia, F.J., Aquino-Britez, D., Escobar, Juan J., Luque, J.L.: "Temporal phase synchrony disruption in Dyslexia: anomaly patterns in auditory processing". 9th International Work-Conference on the Interplay Between Natural and Artificial Computation. IWINAC 2022
3. Rodriguez-Rodriguez, I., Ortiz, A., Formoso, M.A., Gallego-Molina, N.J., Luque, J.L.: "Inter-channel Granger causality for estimating EEG phase connectivity patterns in dyslexia". 9th International Work-Conference on the Interplay Between Natural and Artificial Computation. IWINAC 2022

4. Gallego-Molina, N.J., Ortiz, A., Martinez-Murcia, F.J., Rodriguez-Rodriguez, I.: "Unraveling dyslexia-related connectivity patterns in EEG signals by Holo-Hilbert spectral analysis". 9th International Work-Conference on the Interplay Between Natural and Artificial Computation. IWINAC 2022
5. Formoso, A. Ortiz, A. Martinez-Murcia, F.J., Gallego, N., Luque, J.L., Escobar, J.J., Montoya, Francisco G.: "Modelling Brain Connectivity Networks by Graph Embedding for Dyslexia Diagnosis". International Conference on Bioengineering, Biomedical Signal and Image Processing (BIOMESIP). 2021.
6. Gallego, N., Formoso, M., Ortiz, A., Martinez-Murcia, F.J., Luque, J.: "Temporal EigenPAC for dyslexia diagnosis". International Work-Conference on Artificial Neural Networks, IWANN 2021
7. Ortiz, A., Martinez-Murcia, F.J., Formoso, M.A., Luque, J., Sanchez, A.: "Dyslexia detection from EEG signals using SSA component correlation and Convolutional Neural Networks". The 15th International Conference on Hybrid Artificial Intelligence Systems (HAIS). 2020
8. Martínez-Murcia, F.J., Ortiz, A., Formoso, M.A., Lopez-Zamora, M., Luque, J., Gimenez, A.: "A Neural Approach to Ordinal Regression for the Preventive Assessment of Developmental Dyslexia". The 15th International Conference on Hybrid Artificial Intelligence Systems (HAIS). 2020
9. Formoso, M. A., Gallego-Molina, N. J., Ortiz, A., Rodriguez-Rodriguez, I., Gimenez, A. (2024, May). Explainable Exploration of the Interplay Between HRV Features and EEG Local Connectivity Patterns in Dyslexia. In International Work-Conference on the Interplay Between Natural and Artificial Computation (pp. 45-54). Cham: Springer Nature Switzerland.
10. Gallego-Molina, N., Martinez-Murcia, F. J., Formoso, M. A., Castillo-Barnes, D., Ortiz, A., Ramirez, J., Luque, J. L. (2024, May). A Survey on EEG Phase Amplitude Coupling to Speech Rhythm for the Prediction of Dyslexia. In International Work-Conference on the Interplay Between Natural and Artificial Computation (pp. 161-170). Cham: Springer Nature Switzerland.
11. Gallego-Molina, N. J., Ortiz, A., Formoso, M. A., Martinez-Murcia, F. J., Woo, W. L. (2024, May). Enhancing Neuronal Coupling Estimation by NIRS/EEG Integration. In International Work-Conference on the Interplay Between Natural and Artificial Computation (pp. 24-33). Cham: Springer Nature Switzerland.
12. Rodriguez-Rodriguez, I., Ortiz, A., Formoso, M. A., Gallego-Molina, N. J., Luque, J. L. (2024, May). Causal Mechanisms of Dyslexia via Connectogram Modeling of Phase Synchrony. In International Work-Conference on the Interplay Between Natural and Artificial Computation (pp. 34-44). Cham: Springer Nature Switzerland.

Chapter 2

Foundations

Language processing relies heavily on sensory processing, which is governed by the synchronization between brain areas that handle information at different levels of abstraction. Thus, the study of the synchronization mechanisms during cognitive processes result essential to understand alterations in the brain networks. This chapter explores the foundations of neural synchronization in sensory perception and its implications for language processing. It challenges the traditional view of continuous perception by proposing the [Temporal Sampling Framework \(TSF\)](#), which suggests that perception operates through rhythmic pulses in the brain, depending on the phonological unit to be processed. Language processing is examined in light of brain models, transitioning from traditional localization theories to a network-based perspective enabled by connectivity models. Additionally, EEG is a very important and not invasive acquisition method for studying neural synchronization. Specifically, [Cross Frequency Coupling \(CFC\)](#) constitutes a method to understand brain function across multiple frequency bands, indicating its relevance for exploring neural processes and brain disorders. The TSF implications for language, particularly in conditions like dyslexia, are evaluated, suggesting that alterations in neural rhythms could provide insight into cognitive impairments.

2.1 Neural synchronization in sensory perception

Sensory perception represents a complex neurobiological process that in recent decades has witnessed radical changes in its conceptualization. The TSF emerges from the established paradigm as a novel model challenging the widespread conception of perception as continuous, instead exposing a more complex system of spatial 'pulses' that operate across neural systems [201].

The fundamental premise of TSF is that sensory information in neural systems is represented in distinct rhythmic time slots rather than in a continuous flow. VanRullen and Koch (2003) carried out pioneering work claiming that perceptual awareness occurs in specific neuronal rhythms, suggesting perception happens through oscillating neurons [202]. This paradigm shift fundamentally alters the understanding of sensory information processing. Moreover, neurophysiological studies have identified complex neural synchronization structures enabling temporal sampling. Thus, different senses

exhibit distinct sampling characteristics reflecting their computational requirements. Maass et al. (2002) observed that the visual cortex samples environmental information at 10-100 Hz, facilitating rapid motion and scene orientation [134]. The auditory system demonstrates even greater temporal precision, with sampling rates of 200-500 Hz likely optimized for sound localization and complex speech comprehension [134]. Such sampling rate variations are adaptively utilized by neural systems for computational efficiency. This way, neural synchrony plays a crucial role in information processing and transmission. In the same line, Fries (2005) argues that precise neural assembly activation timing dynamically relays information [63]. Synchronized neural populations facilitate: 1) selective information transmission, 2) integration of temporally disparate brain region information, and 3) internal sensory world modeling. Other works such as [169], explains temporal sampling mechanisms, demonstrating how unexpected stimuli trigger intense neural responses due to predictive neural models. This predictive coding elucidates how neural systems encode complex sensory signals, anticipate environmental changes, and allocate resources efficiently. Additionally, Senkowski et al. (2008) explored cross-modal sensory integration, revealing how sensory systems synchronize sampling rates to create unified perceptual experiences [188]. This synchronization enables multisensory perception, complex cognitive processing, and adaptive environmental interaction. Variations in temporal sampling mechanisms may serve as potential biomarkers for neurodevelopmental disorders.

Advanced acquisition techniques, including high-resolution EEG and magnetoencephalography (MEG), have emerged providing unprecedented insights into neural oscillatory patterns and synchronization mechanisms [181]. Signals acquired using these methods have been used in this Thesis to estimate connectivity through neural synchronization.

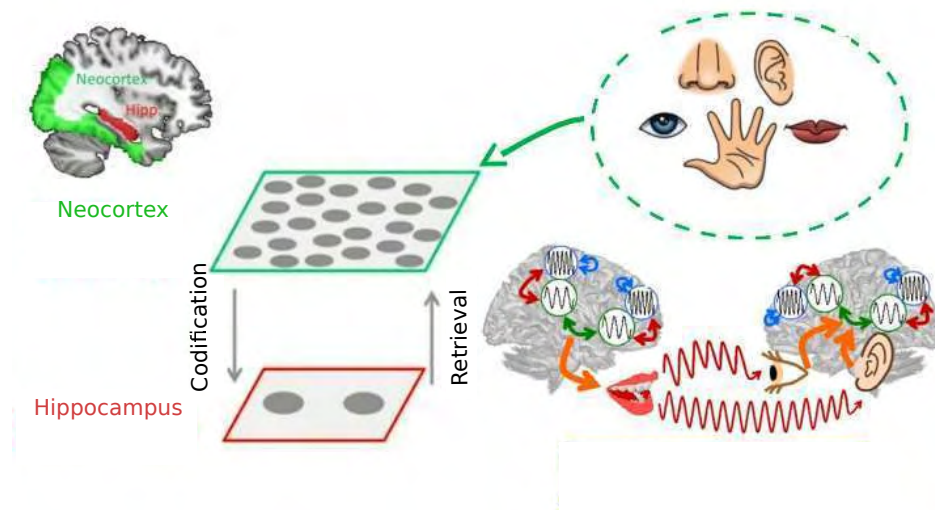


Figure 2.1: Sensorial Processing. During brain development, neurons in the cerebral cortex cluster into regions specialized for sensory processing. Brain networks generate abstract economic representations (prototypes). Figure by P.J. López and A. Ortiz

2.2 Language processing models

Language represents a communication ability innate to humans, with brain mechanisms for linguistic development initiating from birth. Studying these mechanisms is crucial for understanding brain processes involved in language production, comprehension, and associated disorders like aphasia, as well as reading and writing challenges. Although researchers have developed various anatomical and computational models describing brain areas involved in language processing, understanding these models' scope and validity is essential for designing stimuli that can differentiate responses between control subjects and individuals with language disorders.

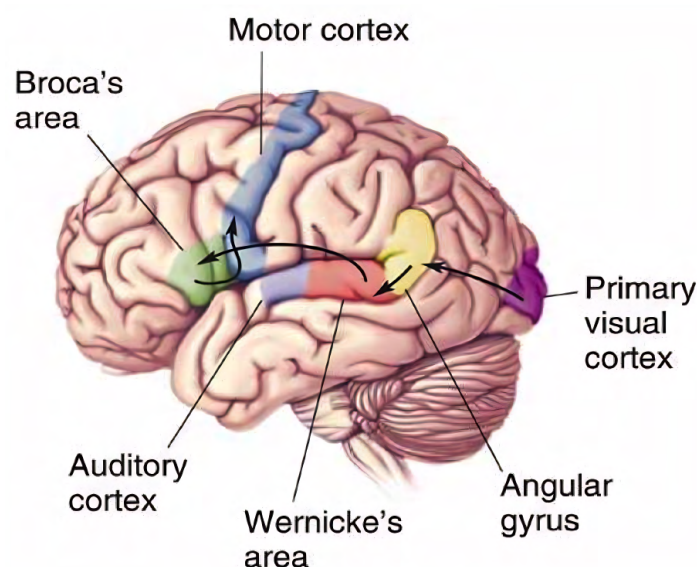


Figure 2.2: Wernicke-Geschwind functional model. Areas involved in language production and comprehension [179]

Early studies on language processing assumed a localisationist model of the brain, arising from postmortem investigations of subjects with language production or comprehension problems. Brain areas were identified and associated with different stages of language processing. These models analyzed the relationship between four brain regions: Broca's area (responsible for language production), Wernicke's area (responsible for language comprehension), the motor area, and the primary auditory cortex. Different types of aphasia correlate with lesions in specific regions (such as Broca's aphasia, characterized by difficulty producing words fluently, or Wernicke's aphasia, associated with the inability to produce intelligible spoken language). The Wernicke-Geschwind model, illustrated in Fig. 2.2, posits an interaction between Wernicke's and Broca's areas via the arcuate fasciculus. According to this model, word meanings stored in Wernicke's area are transmitted to Broca's area through this neural pathway. Broca's area then generates commands to the primary motor cortex for producing the sounds of spoken language. Thus, in spoken language comprehension, the primary auditory cortex first processes speech by breaking it down into identifiable units. These units are then processed in Wernicke's area, where word comprehension occurs and abstract word representations are generated. Reading follows a similar process, originating

in the primary visual cortex (V1). The angular gyrus transforms visual information from V1 into an auditory signal sent to Wernicke's area, after which language processing proceeds analogously to spoken language. As a result, while the model assumes distributed language processing and localization of specific functions, it introduces a crucial concept of inter-area connectivity as essential for language processing. Despite being a primitive localisationist model, it laid groundwork for understanding neural communication in language.

Current functional imaging techniques have enabled real-time observation of brain regions activated during language processing, rendering the Wernicke-Geschwind model obsolete. More sophisticated models now demonstrate that numerous brain regions activate synchronously, forming a network and definitively moving beyond the localisationist approach. Importantly, the exact location and size of Broca and Wernicke areas vary between individuals [101]. Advanced imaging techniques like diffusion tensor tractography (DTI) [190] combined with fMRI have facilitated the construction of activation maps detailing word generation levels and meaning associations (meaning - word form - sound - concept). These studies have also revealed the involvement of subcortical regions, particularly the basal ganglia, in spoken language processing [27, 101], a significant departure from the traditional view that language was exclusively a cortical function.

2.3 Electroencephalography

As explained, the usual way to study neural synchronization consists on the analysis of brain functionality through EEG signals. EEG is a non-invasive way to obtain information about brain activity with high temporal resolution. The subject wears a cap with several electrodes that are as close as possible to the scalp and capture small electric currents. These currents are made by the activity of the neurons. When a group of neurons is activated, they produce a small electrical signal that can be detected by the electrodes placed on the scalp and recording.

These raw signals can be decomposed in five different bands called brainwaves, each with its own frequency range. From low to high we have Delta (1-4Hz), Theta (5-8Hz), Alpha (9-12Hz), Beta (13Hz-30Hz) and Gamma (>30Hz) band. Delta waves, are prominent during deep sleep and play a role in restorative processes [41]. Theta waves, are linked to relaxation, creativity, and meditation [86]. Alpha waves, are present during wakeful relaxation and are associated with calmness and improved cognitive performance [30][142]. Beta waves, are prevalent during active thinking, problem-solving, and concentration [104]. Finally, Gamma waves are involved in higher mental activities, memory formation, and information processing [19].

EEG is a widely used method to measure the electrical activity of the brain. As neurons fire, they generate a small current that is captured by the electrodes placed on the scalp. EEG has contributed hugely to the advance of Brain Computer Interface (BCI), enhancing the quality of life through various applications like controlling a wheelchair or a wearable [5, 89]. EEG has also been used to decode motor functions and associated brain activities, as in [158]. Recent research has shown the potential EEG has studying low-frequency cortical entrainment to speech, reflecting phoneme-

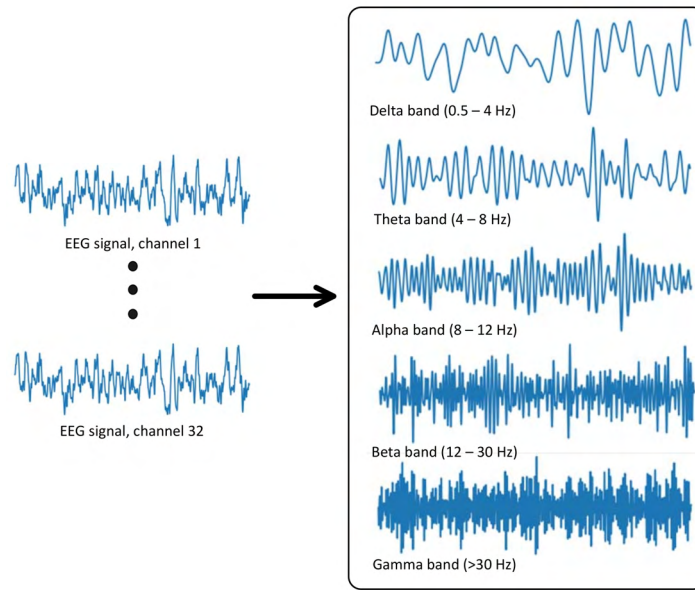


Figure 2.3: Multichannel EEG signal example and frequency bands

level processing [126], demonstrating its importance in investigating neural responses to complex stimuli like speech, which has been key for the development of this thesis. However, working with EEG signals (especially with surface electrodes) is challenging due to its noisy nature, low amplitude, and very easily disturbed by artifacts. For that reason, the preprocessing stage is critical as it may come out quite hard or even challenging to remove noise from these signals with all of the information preserved. Such artifacts can be described as eye movement, muscle movement and quite electrical activity of outside devices that can be present in EEG signals causing incorrect conclusions if not controlled, and have to be removed using filtering and signal separation techniques such as ICA that may be time-consuming.

In the context machine learning, EEG signals are complex, non-periodic and non stationary. The electrical activity in areas of the brain that is picked up by each electrode in the array is only part of the picture of electrical activity in that area of the brain and there is much diversity in such an activity picture depending on age, sex, or brain activity. Such variability makes it harder to apply machine learning models consistently for various subjects. In addition, EEG data is usually quite non dense and non-structured, and has high temporal variations, and so effective modeling is difficult. In summary, the combination of noise, high dimensionality, and subject-specific variability demands robust signal processing and machine learning algorithms, such as deep learning methods, that can handle large amounts of data and identify subtle patterns. These challenges highlight the need for continuous development in signal processing techniques and machine learning methods.

2.3.1 Neural Synchronization

While EEG offers high temporal resolution, the complexity of EEG signals requires advanced signal processing and feature extraction methods for its interpretation. The

study of the interaction between frequency bands and synchronization among populations of neurons can provide functional information regarding the areas involved in the development of cognitive tasks. One of the most used methods to assess local and inter-regional synchronization is CFC. It quantifies how one signal influences or modulates another one. There are several variations of this technique that have been studied: [Phase-To-Phase Coupling \(PPC\)](#), [Phase-To-Amplitude Coupling \(PAC\)](#) and [Amplitude-To-Amplitude Coupling \(AAC\)](#). Canolty et al. [21] reviewed evidence of CFC and the role it plays in brain functioning. CFC is present in multiple brain processes, such as regulating inter-area communication at different spatio-temporal scales, with a strong correlation between CFC strength and learning performance. This suggests that phase–amplitude CFC provides a mechanism for regulating synaptic strength, which is essential for cognitive processes such as attention, learning, and memory. Moreover, studies have shown that CFC is a powerful tool to gain insight into brain behavior in a wide range of brain disorders such as epilepsy [168], schizophrenia [93], obsessive-compulsive disorder [8], Alzheimer’s disease [83] or Parkinson’s disease [43].

This thesis primarily focuses on PAC, in which the phase of one signal influences the amplitude of another. Fig. 2.4 illustrates this phenomenon, showing that the amplitude of the high-frequency signal peaks whenever the phase of the other signal is at the same point. In a real EEG, this behavior is more subtle, and techniques to enhance, validate or remove correlations are needed.

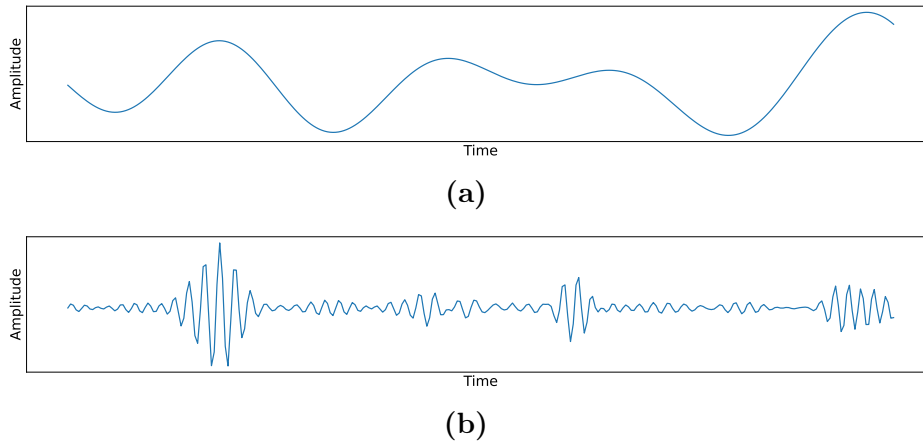


Figure 2.4: Example of coupled signals. The phase of the low frequency (a) signal modulates the amplitude of the high frequency signal (b).

Along the years, several ways of computing the PAC score over two signals have been developed. Some of them are reviewed in [100], among them, the ones used in this thesis are [Phase-Locking Value \(PLV\)](#) and [Mean Vector Length \(MVL\)](#).

2.3.2 Temporal Sampling Framework: synchronization in language processing

Over the past 10 years, an explanation of [Developmental Dyslexia \(DD\)](#) has been developed based on the study of neuronal synchrony with the rhythmic properties of

Table 2.1: Summary of results from different studies on dyslexia and brainwave synchronization.

Study	Age	Participants	Paradigm	Technique	Results
Lehongre et al., 2011 [122]	24	44 (23 with reading difficulties)	ASSRs, White Noise	MEG	Synchrony in <i>planum temporale</i> in gamma band (25–35 Hz) in controls but not in dyslexics. Dyslexics show no asymmetry in drag at 25–35 Hz.
Hamalainen et al., 2012 [68]	27	21 (11 dyslexic)	ASSRs at 2, 4, 10, and 20 Hz	MEG	Reduced drag at 2 Hz in dyslexics and weaker response in Delta. Stronger drag at 10 Hz.
Molinaro et al., 2016 [144]	8–14 and adults	40 (20 dyslexic). 10 adults and 10 children (M=11, 8–14), per group	40 sentences with meaning [Gamma (25–35 Hz), Theta (4–7 Hz), Delta (1–3 Hz)]	MEG/MRI	Adults and children with dyslexia showed Delta band asynchrony. Reduced Delta synchronization in RH. Deterioration of HD-LH coupling.
De Vos et al., 2016 [46]	14	53 (32 dyslexic)	ASSRs at 4, 10, 20, and 40 Hz	EEG	Theta asynchrony, reduced Alpha sync (10 Hz); Beta activity sync (20 Hz) in dyslexics. Correlation between Beta sync and phonological abilities.
Power et al., 2016 [166]	46: 12 dyslexics, CE and YI	14–15	100 sentences. 5 bands (0.4–2, 2–4, 4–6, 6–8, 8–12 Hz)	EEG	Reduced synchronization in dyslexics in CE and YI at 0–2 Hz. No differences at higher frequencies between dyslexics and CE-YI controls.
De Vos et al., 2017 [45]	5,7 and 9	68 (14 dyslexic, 36 CE, 18 CL)	ASSRs at 4 and 20 Hz	EEG	Correlations between Delta/Theta bands and phonological abilities. Major synchrony at 20 Hz at 5 years old, better reading at 9 years old.

language, inspired by the work of Poeppel [74]. Groups of neurons in the auditory cortex have been identified that oscillate synchronously at different frequency bands ranging from 1 Hz to 40 Hz. Lower frequencies have been associated with intonation processing and syllabic segmentation, whereas groups of neurons synchronized at higher frequencies have been related to phoneme segmentation [108]. Alterations in this described synchrony could produce abnormal syllabic and phonemic representations, constituting an interesting potential endophenotype of language disorders and, in particular, dyslexia, that is, a bridge between the expression of genetic anomalies and cognitive alterations in the phonological domain [78].

Speech encoding, which is related to prosody and sensorimotor synchronization problems, can be revealed by searching for patterns in EEG channels across different sub-bands, as it provides sufficient temporal resolution. Although recent, the study of EEG synchronization with speech signals across different bands has already given rise to significant debates and hypotheses. Essentially, synchronization deficits have been found in lower frequency bands associated with prosody and syllabic segmentation or, alternatively, in higher frequency bands associated with intrasyllabic and phonemic segmentation. Table 2.1 shows results obtained with different samples, paradigms, and brain activity recording techniques.

Language is a complex signal that operates across multiple time scales, from rapid phonemes (individual sounds) to slower components like syllables, words, and sentences. The TSF suggests that the brain uses rhythmic neural oscillations synchronized with the different temporal aspects of speech to parse and interpret language effectively. These neural oscillations are reflected in distinct EEG bands, each associated with different linguistic functions, as shown in Fig. 2.5.

The core idea of TSF is that the brain segments and integrates speech using neu-

ral oscillations tuned to different frequency bands. Populations of neurons oscillating at different frequency bands act like *temporal filters* allowing the brain to selectively sample the auditory signal at different rates, depending on the phonological unit to be analyzed. Delta band (1-4 Hz), which is a low-frequency EEG band, plays a crucial role in processing the slower units of language, such as prosody (the rhythm, stress, and intonation patterns of speech). Prosody helps to determine sentence structure, emotional tone, and overall meaning. Delta oscillations align with the rhythm of phrases or sentences, allowing the brain to track broader units of meaning and structure over longer periods, contributing to coherent language comprehension. Theta rhythms (4-8 Hz) are closely aligned with the natural rate of syllables in spoken language, which often falls between 4 and 8 per second. This band enables the brain to chunk speech into tractable units, facilitating syllable segmentation and the recognition of individual words. Theta oscillations are believed to support lexical access (retrieving word meanings from memory) and play a crucial role in speech perception, especially when parsing continuous speech into distinct words. This frequency band is particularly active during tasks involving verbal working memory and phonological processing, both of which are fundamental for language comprehension and production. Beta band (12-20 Hz) is involved in the processing of faster elements of speech, such as the internal segmentation of the syllable and fine-grained phonemic distinctions. Phonemes, the smallest units of sound in language, are produced with a high frequency and beta/gamma oscillations are thought to help track these fast transitions. This band may also support more fine-tuned aspects of language processing, such as grammatical construction and the articulation of precise phonetic details, which are essential for clear speech production. Gamma band (>30 Hz) is linked to higher-order cognitive functions, such as semantic processing and the integration of information from multiple sources. In the context of language, gamma activity is often associated to phoneme segmentation and the binding of syntactic and semantic information.

Thus, the TSF emphasizes the hierarchical nature of language processing in the brain, with different EEG bands synchronizing to different linguistic units. Each frequency band provides a temporal window through which the brain can analyze speech signals, from slow prosodic cues captured by delta oscillations to rapid phonemic details tracked by beta and gamma bands. This multi-scale analysis is crucial because speech contains overlapping information across various timescales.

Goswami [78] proposed, under the TSF theory, the hypothesis that dyslexic children exhibit desynchronization in the Delta and Theta bands, affecting prosody and word segmentation. A subsequent EEG study [166] aimed to test the theory formulated in [78], using complex auditory stimuli (phrases), analyzing EEG and voice signal spectrograms to correlate the EEG envelope in specific bands with the envelope of the voice signal in the same bands. This study demonstrated Delta band desynchronization during speech processing in the right hemisphere's auditory regions. However, no experimental evidence of desynchronization between the Theta band and syllabic rhythms in speech was found. Other studies have used amplitude-modulated white noise at different frequencies in experiments measuring auditory evoked potentials. For instance, [85] demonstrated that dyslexic readers exhibit a weaker response to a 2 Hz stimulus in the auditory regions of the right hemisphere. The signal processing methodology used in this study is similar to that in [166], although white noise-modulated stimuli

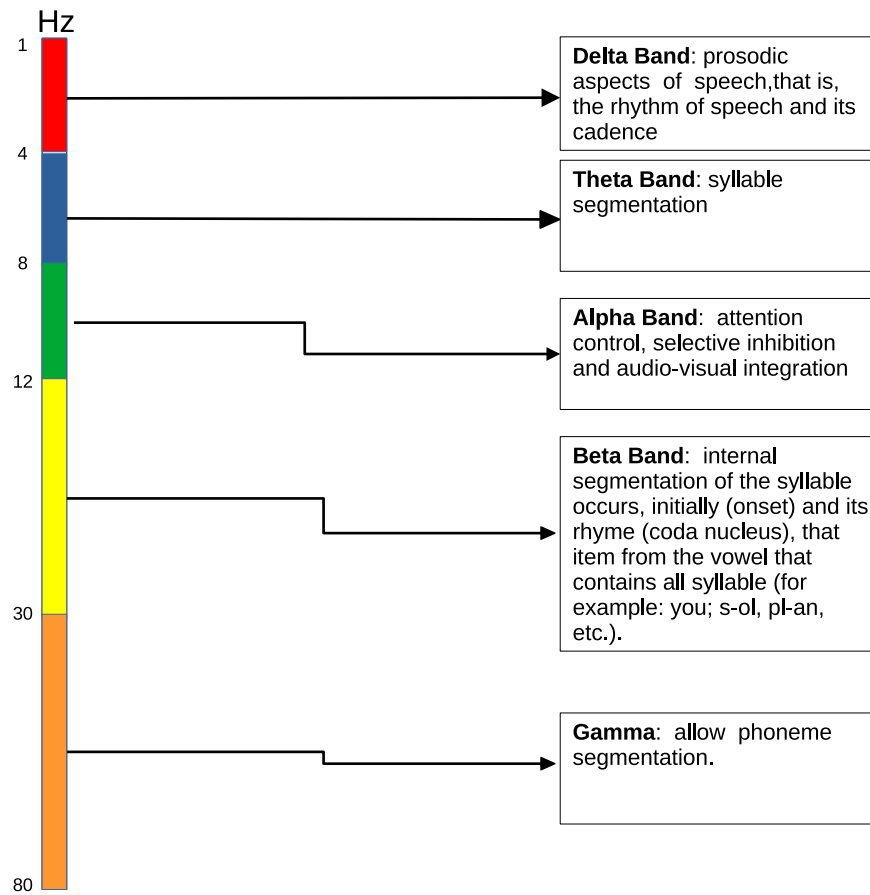


Figure 2.5: EEG Brainwaves and their role in language processing [109]

and MEG acquisition techniques were used instead of EEG.

Based on the results obtained in [165], [35] used amplitude-modulated pure tone stimuli (at 2.4 Hz, as in [165]). This study aimed to observe the significance of phase synchronization in the auditory cortex using statistical techniques (ANOVA). This experiment demonstrated a difference in the mean phase between control and dyslexic groups, which has been linked to cognitive impairment in the phonological representation of word forms. In a more recent experiment [46] conducted with dyslexic and control adults, using the ASSR (Auditory Steady State Response) methodology, positive correlations were found between Alpha band synchronization and phonological skills in normal readers, which were absent in dyslexic readers. However, dyslexic readers exhibited positive correlations between Beta band synchronization and phonological skills.

In more recent works [81, 135], Goswami et al. conducted experiments to validate the theoretical TSF framework [78, 79, 81], focusing on the phonological deficit as the main cause of dyslexia. In this case, complex auditory stimuli (specifically, listening to a story) were again used, involving children aged 7 to 8 years. Specifically, the stimulus application consisted of 5 minutes of gazing into space and 5 minutes of listening to a story. Subsequently, the story's envelope was characterized in different

bands, termed stress, syllable, sub-beat, and phoneme, and the spectral coherence between the MEG bands and the different bands of the speech signal was calculated, searching for differential patterns between controls and dyslexics. In this work, the authors introduced the concept of connectivity (not previously used by them to explain Control/DD differences), but locally and focused on the Delta and Theta bands. The preliminary results obtained in our experiments with ASSR stimuli (simple, without speech stimuli) are consistent with the results presented in this article. However, in our case, we used EEG with non-interactive stimuli and a much simpler experimental methodology.

The use of advanced and specific techniques for signal processing, feature extraction, classification, or modeling has not been the norm among researchers in this field. However, in recent years, the number of research studies incorporating machine learning techniques has grown, especially in neuroimaging studies [159], while EEG and fNIRS studies generally continue to use classical methods. Combining statistical signal processing techniques with machine learning for biomarker selection and automatic diagnosis have been clearly demonstrated in the literature and in the prior work of the investigator in this project. As shown earlier, these techniques have been successfully used for feature extraction, selection, and classification, as well as for building models of various neurodegenerative diseases (such as Alzheimer's and Parkinson's diseases).

As evidenced in most of the previously cited works, which are currently considered reference studies, speech signal stimuli are used, attempting to associate synchronization deficits in specific brain areas guided by theoretical or semi-empirical models. However, the problem has not, as a general rule, been addressed from an integral perspective, considering both local and inter-regional interactions to build computational models of the language processing network and explore differential patterns between control and dyslexic subjects.

2.4 Artificial Intelligence in healthcare

The field of AI is considered part of the computer science field. In one form or another, it has been around since the beginning of computer science, ever since Alan Turing thought about whether machines could someday reason and learn like humans. The day seems to be approaching at great speed since Deep Learning (DL) has taken control of the field.

We could divide the AI field in several subsets like robotics, genetics algorithm and, of course, machine learning. Machine learning can be divided again in various types of algorithms: supervised learning algorithms, unsupervised learning and reinforcement learning. In this thesis, the focus is on the first two types of algorithms, since reinforcement learning is not suitable for the type of problem we try to resolve.

In supervised learning, algorithms are trained using labeled data. This means that each training example includes the correct output. The main objective is for the algorithm to map the relationship between inputs and outputs so it can make predictions for unseen data. Common tasks in learning involve classification, which predicts discrete labels and regression which predicts continuous values. Known algorithms in this category include linear regression, support vector machines and neural networks.

In unsupervised learning, the focus is on data without labels. The aim is to uncover the structure within a dataset. A typical application of learning is clustering, where objects are grouped based on similarities within the same cluster rather than with objects in other clusters. Dimensionality reduction is another task, in this area that aims to simplify data visualization by reducing the number of variables considered. Examples of learning methods include K-Means clustering and principal component analysis (PCA).

In recent years, DL, a branch of machine learning, has made significant impact. Models like convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have excelled in areas such as image recognition, natural language processing and gaming. These models stand out for their utilization of layers of nonlinear processing units (neurons) for extracting and transforming features.

Despite the success of both types of learning, there are still challenges. Supervised learning often demands a volume of labeled data, which can be costly and time consuming to acquire. On the side unsupervised learning algorithms require careful selection and tuning to capture the underlying data structure, which may not always be straightforward.

AI has had a direct impact on the medical field. In the field of medical imaging, it was by improving diagnostic accuracy, efficiency and patient outcomes. It is widely used across almost all imaging modalities such as MRI, PET, SPECT and ultrasound to automate tasks and enhance diagnostic precision [49]. For example, in the realm of MRI, AI plays a role in diagnosing conditions like prostate cancer, breast cancer, cardiac diseases, brain tumors and musculoskeletal disorders [29, 107, 123, 196]. The integration of AI into MRI processes has resulted in accurate disease diagnosis, leading to improved patient care and treatment outcomes. AI is crucial for analyzing MRI data to detect and classify breast lesions accurately [55]. Moreover, it aids in predicting treatment responses assessing risks and enhancing image quality. Additionally, AI is crucial for tumor grading and classification efforts – particularly when distinguishing brain gliomas using various MRI sequences [197]. It has also been utilized to predict the depth of invasion in gynecological conditions [66].

Moving to fNIRS, it is now seen as a tool for brain imaging, allowing researchers to monitor brain activity in a non-invasive way. AI plays a role in analyzing fNIRS data, providing valuable insights into brain function and cognitive processes [133]. Another technique used to analyze brain related data that take advantage of AI is EEG. Research has looked into using machine learning methods to categorize brain disorders, such as bipolar depressive disorders, showing how AI can assist in medical decision-making processes [54]. Furthermore, AI has been utilized for predicting outcomes after arrest based on EEG data, demonstrating its ability to improve neurological clinical practices[215]. The integration of AI with EEG signals has enabled interpretation of electrophysiological information, highlighting the progress made in this field[42].

The future of AI in the medical field holds immense promise, with ongoing research focusing on leveraging AI for better medical diagnosis and treatment. Studies have highlighted the potential of AI-driven diagnostic tools to assist clinicians in interpreting complex biomedical data. However, AI-systems seen as a black box does not provide essential information about the patterns used to produce a specific outcome. This is a space where accurate black-box algorithms that cannot justify their

outputs are at best inadequate and at worst dangerous. To add transparency, explainable methods offer research scientists and clinicians to interpret AI systems for specific conclusions in disease diagnosis, drug discovery or treatment recommendations. Such transparency is crucial for trust between the healthcare professionals and patients, to meet regulatory requirements, as well as to be able to identify any bias or errors in the biological explanation of AI. Furthermore, the XAI can provide insights into previously unknown biological relationships or mechanisms, thus enabling discovery rather than just predictions. As an example, when an AI system recognizes patterns in medical imaging suggestive of a certain disease, unpacking the features that it finds relevant can potentially provide insights on how a particular abnormality develops and/or progresses. This also aids the validation process where domain experts can assess whether the reasoning of AI corresponds to standard medical knowledge thereby meeting the high safety and ethical standards needed for healthcare applications.

Chapter 3

Electroencephalography and functional connectivity

Electroencephalography (EEG) is a non-invasive method for capturing and interpreting the brain electrical activity. EEG has transformed our comprehension of how the brain functions, empowering healthcare professionals and scientists to identify neurological conditions, monitor brain activity and explore the intricacies of human thinking processes.

The roots of EEG can be traced back to the 20th century through the pioneering efforts of German psychiatrist Hans Berger. In 1924 Berger conducted groundbreaking experiments by capturing brain waves using an electroencephalogram device he coined as "brain waves." His discoveries showcased that it was possible to measure and study the brain's rhythms, leading to significant progress, in both clinical practices and scientific investigations. Since then, the applications of EEG to the medical field has been increase exponentially. This technology born from the curiosity of physiologists and psychiatrists eager to unravel the mysteries of brain electrical activity, has evolved into an indispensable tool in neuroscience.

By the mid-20th century, EEG had firmly established itself as a valuable asset in identifying and diagnosing various neurological conditions. Simultaneously, researchers began exploring its potential in experimental psychology, uncovering correlations between specific brain rhythms and mental states. Furthermore, the development of high-density recording techniques and standardized electrode placement systems, such as the International 10-20 System, has greatly facilitated study comparisons and result replication. These innovations have significantly enhanced the reliability and reproducibility of EEG research.

Nowadays, EEG stands as a testament to the power of scientific innovation and curiosity in advancing our understanding of the human brain. It has become an essential technique for studying and diagnosing numerous neurological conditions and investigating brain activity and cognition. The non-invasive nature of EEG, its high temporal resolution, and its relatively low cost make it a valuable tool in various areas of neuroscience and medicine [150]. As a matter of fact, EEG is widely used in clinical neurology for diagnosing and monitoring diseases such as epilepsy, providing a detailed view of brain electrical activity during seizures and facilitating the identification of epileptogenic foci [9]. Furthermore, EEG is useful in detecting sleep disorders and di-

agnosing alterations in consciousness levels and certain neurological pathologies such as encephalopathy, encephalitis, and neurodegenerative diseases [15].

In the research domain, EEG has been employed to explore the neural correlates of numerous cognitive processes, including attention, memory, learning, and decision-making [112]. The patterns of brain electrical activity recorded through EEG can be associated with different mental and emotional states, offering real-time insights into brain activity dynamics. Moreover, EEG has been utilized in studies on consciousness and sleep, enabling the investigation of various sleep stages and changes in brain activity related to the loss and recovery of consciousness [18].

Recently, EEG has emerged as a promising technique in the field of BCI, allowing direct communication between the brain and an external device [209]. These applications can be particularly relevant for individuals with severe motor disabilities, providing a new means of communication and control over assistive devices. The versatility and scope of EEG are reflected in its diverse applications and its importance in both clinical practice and research. EEG has proven to be a powerful and flexible technique, capable of shedding light on the processes underlying brain activity and cognition while supporting clinical decision-making in various neurological contexts.

3.1 The electrical brain

The electrical activity of the brain, which is essential for the function of the central nervous system, arises from the complex interaction of ionic processes in neuronal membranes and synapses. Understanding these dynamics and their manifestations in the EEG is crucial for deciphering the mechanisms behind cognitive phenomena and diagnosing neurological disorders [150]. The source of brain electrical activity lies in the electrochemical properties of neurons. Neuronal activity involves the generation and propagation of action potentials, which are depolarization and repolarization events that enable information transmission along neurons. Each action potential involves a flow of ions across neuronal membranes, creating a small local electric field. EEG record this electrical activity.

One of the primary challenges in EEG is localizing the sources of these dipoles, that is, determining precisely which brain regions the recorded potentials originate from. This is because EEG records electrical activity on the scalp surface, which can be affected by various factors, such as the placement of electrodes, the conductivity of different head structures, and the orientation and activity of the underlying neurons [18]. This problem is known as the volume conduction problem, see Fig. 3.1. This could lead to spurious correlations between several electrodes.

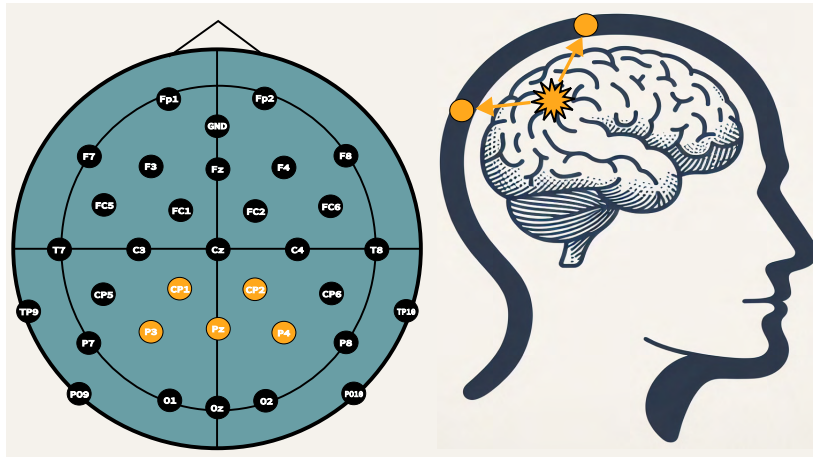


Figure 3.1: When dealing with EEG data, we should consider volume conduction. In volume conduction, one event can be captured by multiples electrodes at the same time, resulting in spurious correlations.

Oscillations of brain electrical activity, known as brain rhythms, are characterized by their frequency, amplitude, and topographic distribution, and are linked to different cognitive and behavioral states. For instance, the Alpha rhythm, which typically dominates during relaxation with eyes closed, is associated with cortical inhibition and decreases during mental tasks. In contrast, the Beta rhythm may be related to mental concentration, while Theta and Delta oscillations are observed during sleep and deep relaxation, respectively [112]. A depiction of these rhythms can be found in Section 2.3

Brain electrical activity can also be altered in various pathological conditions. In epilepsy, for example, high-amplitude paroxysmal discharges can be observed. In neurodegenerative diseases, such as Alzheimer’s disease, a slowing of brain activity is observed, with an increase in slow oscillations (Delta and Theta) and a reduction in fast oscillations (Alpha and Beta) [15]. The interaction of cortical and subcortical systems is another crucial aspect reflected in the EEG. For example, thalamo-cortical activity is essential for the generation of specific brain rhythms, such as Alpha oscillations and sleep spindle and slow wave oscillations [189]. Similarly, the interaction between the prefrontal cortex and other cortical and subcortical regions is reflected in EEG coherence, a measure of synchrony between different brain regions that has been associated with cognitive processes such as attention and memory [178].

Finally, advances in understanding electrophysiology are contributing to the improvement and innovation of EEG techniques. For example, source localization techniques allow inferring the cerebral location of the electrical activity recorded in the EEG, based on electrophysiological and anatomical models of the brain. Similarly, EEG-based connectivity analyses are based on electrophysiological principles to estimate the synchrony and interaction between different brain regions [178].

3.2 EEG acquisition

EEG signals are acquired through voltage detectors called electrodes, which sense the weak electrical signals produced by the simultaneous firing neurons in the brain. These

signals are then amplified and recorded, producing a visual representation of the brain's activity over time.

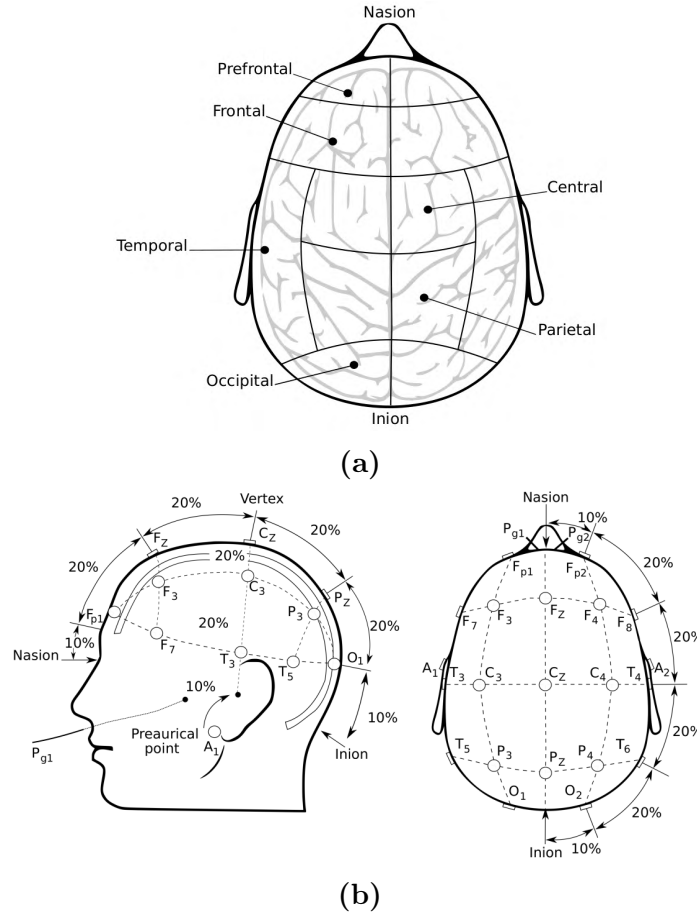


Figure 3.2: (a) Cerebral lobes. Axial view of the skull showing the areas corresponding to the lobes. Image adapted from [143]. (b) EEG the electrodes are placed at 10%,20%, 20%, 20% and 10% from inion to nasion along length. Image adapted from [6].

In order to capture the signal, the electrodes are placed in the scalp in multiple positions according to a specific montage. The choice of EEG montage to capture signals varies depending on the specific application and desired outcomes. In the experiments relative to this thesis, the chosen montage was the international, standardized 10-20 system with 32 channels at specific positions that allows monitoring the auditory cortex. It is named like that because it splits the skull into increments of 10% or 20% to place the electrodes, and every electrode is identified with a letter and a number. Moreover, electrodes at the left hemisphere are designed with odd numbers, while electrodes at the right hemisphere are designed with even numbers (Fig. 3.2b, along with a letter indicating the lobe, as shown in Figure 3.2a. In Fig. 3.3 there is a depiction of this montage and the final result when using it.

The placement of electrodes is crucial for accurately collecting data in EEG. This setup refers to how electrodes are connected to the EEG recording equipment and may follow a monopolar (unipolar) or bipolar method (Fig. 3.4, each with its distinct

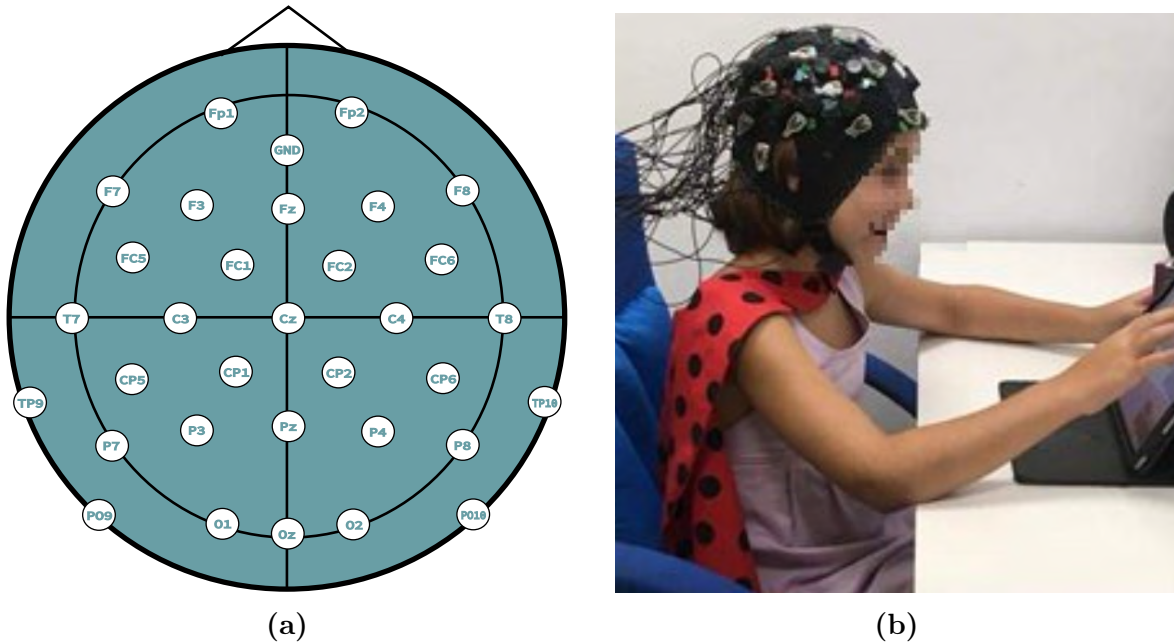


Figure 3.3: (a) Electrode montage in the extended 10–20 system used in the experiments. All 32 channels plus GND. Cz is used as reference. (b) Child with EEG equipment.

characteristics and applications. The bipolar setup uses pairs of electrodes to measure the difference in electrical potential between two points on the scalp. This method is particularly useful for identifying focal electrical activity, that is, activity originating from a localized region of the brain. By employing this method, researchers can identify specific electrical patterns associated with various neurological disorders, such as focal epilepsies [124]. However, the primary limitation of the bipolar setup is that the recorded signal heavily depends on the spatial proximity of the brain areas underlying the two electrodes. Signals generated in areas distant from the two electrodes may not be adequately recorded.

On the other hand, the monopolar setup, also known as a referential montage, uses a single active electrode on the scalp and a reference electrode, which can be placed in a neutral location such as the earlobe or as an average of all other electrodes. This allows measurement of the potential difference between the active electrode and the reference, providing a global view of brain electrical activity. The monopolar montage is particularly useful for identifying generalized electrical activity, that is, activity originating in or propagating across large regions of the brain [150]. This is the setup using in this work.

The electrodes can also be classified into two main categories: passive and active. Passive electrodes are the most common and have been used in EEG for the longest time. They are called "passive" because they simply capture and transmit the signal without doing anything to enhance or modify it. Passive electrodes can be made of various materials, such as silver, gold, or even saline gel or water electrodes. However, signals captured by passive electrodes can be affected by various factors, such as the impedance between the electrode and the skin, environmental noise, or electromagnetic

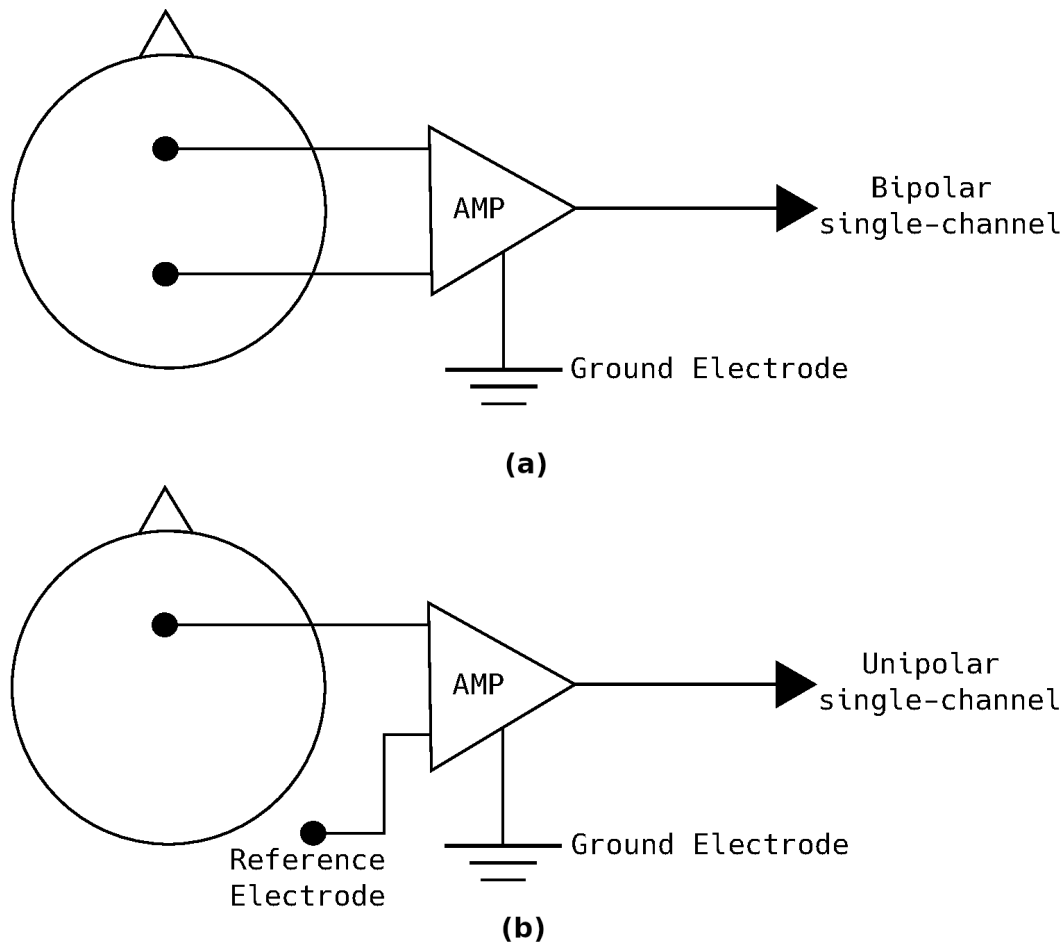


Figure 3.4: Common EEG setups. (a) Bipolar (a) and (b) Unipolar.

interference. On the other hand, active electrodes are a more recent innovation in EEG. Unlike passive electrodes, active electrodes include integrated electronic components, such as amplifiers, that process the signal close to the source. These electrodes amplify the EEG signal immediately after capture, which can help reduce noise and interference and improve signal quality. Active electrodes can be particularly useful in noisy environments or when high spatial resolution is needed. By amplifying the signal directly on the scalp, these electrodes can reduce the influence of noise that may be introduced along the wiring. They can also have a higher input impedance, making them less sensitive to variations in skin and electrode impedance. Moreover, due to their increased sensitivity and noise reduction, active electrodes may be preferable for certain types of research where weaker or more subtle signals need to be captured. However, they are generally more expensive than passive electrodes and may require more maintenance and care, as they contain delicate electronic components. Each electrode in the cap is connected to an amplifier and a digital recording device via a cable. The amplifiers magnify the electrical signal captured by the electrodes, which is then digitized and stored for further analysis.

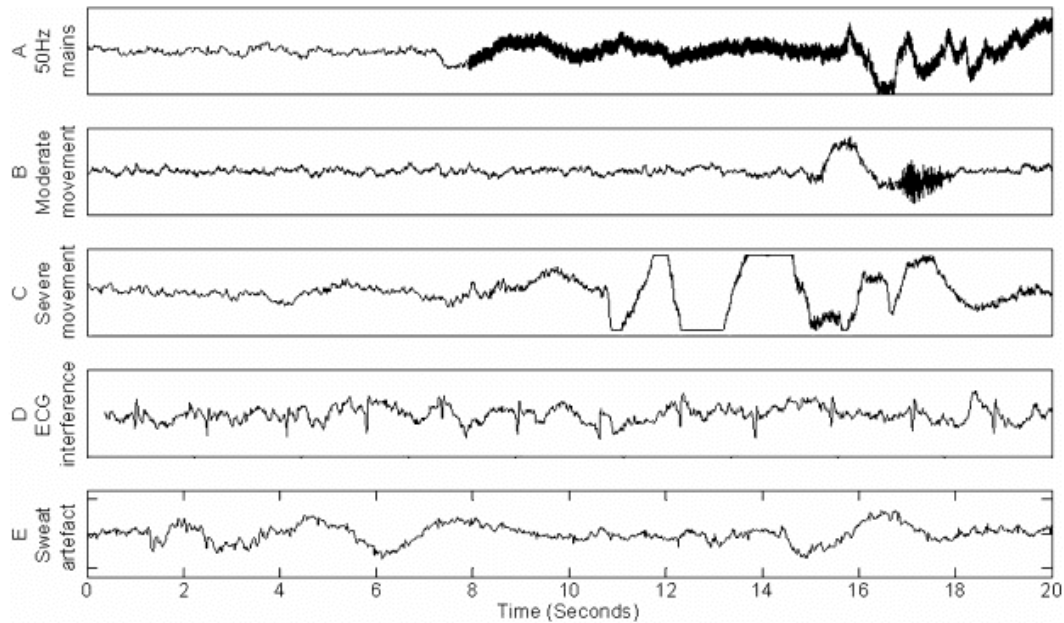


Figure 3.5: Examples of EEG artifacts [146]

3.2.1 Artifacts in EEG signals

Artifacts in EEG are considered one of the main challenges in data collection and analysis in clinical practice and research. Artifacts are distortions or signals that do not reflect genuine brain activity and can be generated by various sources. Their elimination is a crucial aspect of EEG signal acquisition due to its importance in increasing the [Signal to noise ratio \(SNR\)](#). They can mix with true brain signals and introduce noise, obscuring signals of interest and decreasing measurement quality. According to their origin, they are classified into external and internal artifacts, each with particular characteristics and challenges [39, 56]

The influence of artifacts on EEG recordings is a fundamental and complex issue in neurophysiology and cognitive neuroscience, as they can lead to misinterpretation of data, generating confusion in the diagnosis and analysis of brain function. Artifacts, whether internal or external, can mask or distort the true brain signals recorded by the electrodes. Effective management of artifacts in EEG recordings is essential to ensure the quality of the collected data and its correct interpretation. Fortunately, a range of methods have been developed for artifact removal or reduction, which rely on both technical procedures prior to recording and post-recording signal processing algorithms.

Before recording, best practices include proper preparation and placement of electrodes, such as using scalp cleaning and preparation techniques to improve electrical conductance and reduce artifacts related to skin and sweat [162]. Additionally, measures can be taken to minimize subject movement during recording, as this can generate muscle and motion artifacts. For ocular artifacts, electrooculographic (EOG) recording techniques can be used to monitor eye movements and blinks, facilitating their subsequent removal from the EEG.

Regarding post-recording signal processing, various methods exist for artifact re-

moval or reduction. One of the most commonly used is filtering, which can eliminate certain frequencies that do not correspond to brain activity, but care must be taken not to remove relevant EEG information. Furthermore, more sophisticated artifact correction algorithms, such as ICA, can decompose EEG recordings into separate components and remove those representing artifacts. This is possible due to the availability of separate signals considered as "artifacts," such as those generated by eye movement or heartbeat. These signals can be identified and isolated from actual brain activity, allowing for more efficient cleaning of EEG data.

The choice of method for artifact removal or reduction should be based on the nature and magnitude of the artifact, as well as the characteristics of the task and study design. It is important to note that proper artifact removal or reduction not only improves the quality of EEG data but also contributes to a more accurate interpretation of brain activity patterns and, ultimately, a better understanding of the human brain.

3.3 EEG Preprocessing

Before beginning any computational analysis, several preprocessing steps are necessary to ensure the precision and dependability of further analysis. The preprocessing workflow starts with referencing EEG channels to a consistent point, using a central reference electrode to minimize noise and normalize signals across all channels. This process helps standardize the signals against a stable baseline, improving the reliability of subsequent analysis.

Following re-referencing, bipolar channels are created to monitor eye movements and adjust for ocular artifacts. Specifically, vertical and horizontal channels are defined, which capture vertical and horizontal eye movements, essential for identifying and correcting artifacts caused by blinks and lateral eye shifts [38].

After removing artifacts, the data undergoes filtering to refine the signals further. A notch filter removes interference from power lines, while high-pass and low-pass filters eliminate unwanted low-frequency drift and high-frequency noise, retaining only relevant frequency bands for the analysis [207].

Automatic artifact rejection is then applied, discarding segments of data with extreme amplitude fluctuations. This process identifies and excludes sections with values outside a specified amplitude range, ensuring only clean, artifact-free data continues through the workflow.

With artifact rejection completed, the data is segmented into epochs, defined within a set time window. Each segment is then stored and baseline-corrected by using a pre-determined reference window. Baseline correction adjusts the data to a common reference level, helping eliminate any remaining drifts or offsets that could impact interpretation and ensuring consistency across conditions.

The preprocessing workflow concludes with averaging, where all segments within a specified range are averaged to reduce noise. This averaging enhances signal stability and highlights consistent patterns in the EEG data, creating a clearer, more interpretable dataset for further analysis.

With preprocessing complete, the data is ready for further processing, such as decomposition into specific frequency ranges. EEG signals can be segmented into distinct

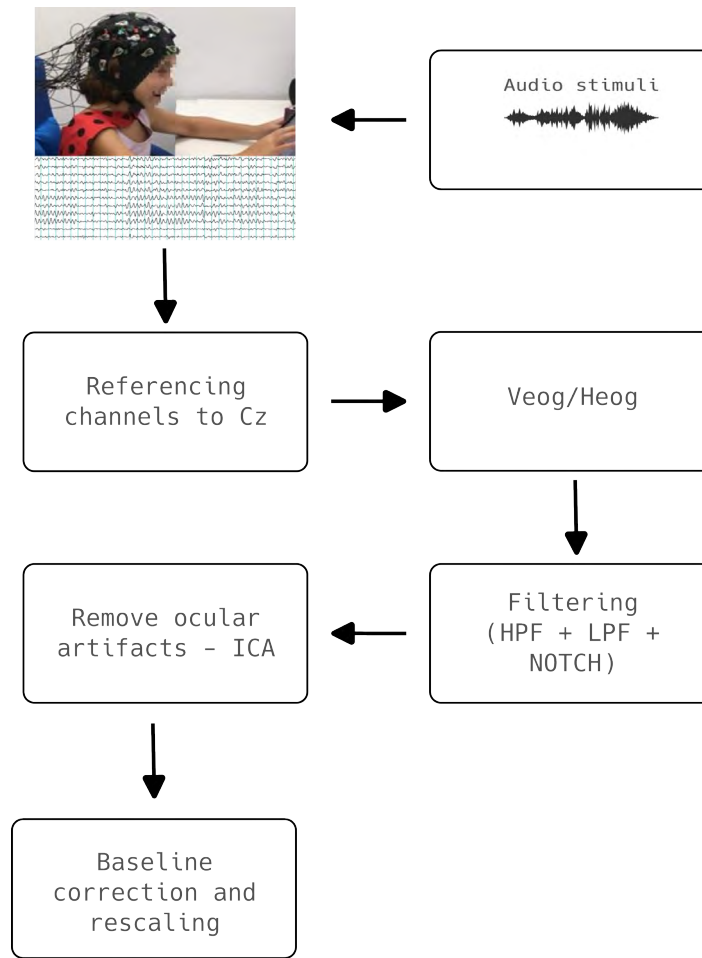


Figure 3.6: EEG Preprocessing Pipeline

frequency bands using a filter with specific characteristics. In this way, FIR filters are usually preferred due to their inherent stability and linear phase characteristics, which preserve the temporal integrity of neural signals. These filters excel in removing specific frequency components like power line interference (50/60 Hz) and high-frequency noise while maintaining the signal's morphological features. By using symmetric or antisymmetric filter designs, such as the Parks-McClellan algorithm, FIR filters can achieve zero phase distortion, effectively eliminating time-based signal warping. Conversely, IIR filters offer computational efficiency and steeper frequency response transitions, making them advantageous for bandpass filtering of specific EEG frequency bands. However, the use of IIR filters is more challenging in applications in which phase distortion is not allowable. These filters introduce non-linear phase shifts that can significantly alter the temporal characteristics of EEG signals. To mitigate this issue, several techniques can be employed, such as Forward-Backward Filtering. This technique involves applying the filter twice (first forward through the signal, then backward). By processing the signal in both directions, the cumulative phase shifts are effectively cancelled out, creating a zero-phase filter with minimal temporal distortion [185].

The selection between IIR and FIR filters depends on the specific signal processing requirements, computational constraints, and desired signal characteristics. Advanced

implementations often involve cascaded or adaptive filtering techniques that combine the strengths of both filter types. Researchers and clinicians carefully design these digital filters to optimize signal-to-noise ratio, extract meaningful neurological features, and enhance the diagnostic potential of EEG recordings in applications ranging from epilepsy detection to cognitive neuroscience research.

With these preprocessing steps—including referencing, artifact removal, filtering, ocular correction, segmentation, baseline correction, and averaging—the EEG data is effectively prepared for detailed analysis, facilitating the extraction of meaningful neural features and patterns.

3.4 Connectivity from EEG signals

The human brain is incredibly complex, with a network of billions of neurons. Central to this complexity is the idea of brain connectivity, the ways different parts of the brain communicate and work together. This connectivity is essential for all processes from basic sensory experiences to intricate abstract thinking and plays a crucial role in understanding how the brain works.

Brain connectivity can be divided into three types: structural, functional and effective connectivity. Each type offers a perspective on how the brain is organized and functions:

- **Structural connectivity** focuses on the links between different areas of the brain, primarily through white matter pathways. These pathways act as the communication highways within the brain are made up of myelinated axons that enable transmission of electrical signals between neurons. Advanced techniques like diffusion tensor imaging have transformed our ability to map these connections in human brains, revealing the intricate "connectome" that enables brain function.
- **Functional connectivity** refers to the relationships or correlations in activity, among various regions of the brain. The idea of connectivity in the brain doesn't always mean a physical link but rather shows which regions tend to activate together during different tasks or even when the brain is idle. MRI and EEG have played a role in uncovering these coordinated activity patterns leading to the identification of large scale brain networks such as the default mode network that is active when the brain is not engaged in any specific task.
- **Effective connectivity** takes it a step further by trying to understand the cause and effect relationships between brain regions. It aims to not identify which areas work together, but also how they impact each other. This kind of analysis offers insights into how information flows within the brain. It is essential for comprehending how various parts of the brain collaborate to generate complex behaviors.

Studying brain connectivity has implications for our understanding of normal brain function and neurological disorders. In healthy individuals, connectivity patterns

change throughout development, with the brain becoming more efficiently connected as we grow and learn. These changes continue throughout adulthood, reflecting the brain's remarkable plasticity.

In various neurological and psychiatric disorders, alterations in brain connectivity are often observed. For instance, schizophrenia has been associated with disrupted connectivity in networks involved in cognitive control and emotional processing. Autism spectrum disorders have been linked to atypical patterns of both over- and under-connectivity in different brain systems. Even in neurodegenerative diseases like Alzheimer's, changes in connectivity often precede observable structural changes, potentially offering early diagnostic markers.

3.4.1 Connectivity Estimation and Modeling

In this thesis, the main focus is on functional connectivity, which refers to the statistical relationships or correlations between different brain regions. This provides insights into how different parts of the brain communicate and coordinate activity during cognitive processes and behaviors [64].

One important aspect is that connectivity is always measured between two signals. Bivariate connectivity measures, which involve interactions between only two brain regions or electrodes, are the most commonly used in cognitive electrophysiology. This prevalence can be attributed to several factors. First, the current understanding of brain interactions is still relatively limited, with few widely used detailed models of multinode networks in cognitive electrophysiology. This may be due to the lack of approachable and intuitive mathematical and statistical analyses for quantifying larger and more complex networks. Additionally, multinode brain interaction models become increasingly difficult to conceptualize, making it easier to break them down into simpler bivariate cases. Bivariate connectivity analyses are also easier to implement, interpret, and test with established statistical procedures. Another possible reason for the abundance of bivariate connectivity methods is that the brain may actually function in this manner, with bivariate connections being the most relevant for many cognitive functions.

It is important to note that bivariate correlations can potentially inflate or misrepresent estimates of relationships if the network structure is actually multivariate. This is particularly relevant for brain connectivity, as the brain is a highly multivariate system. However, for task-related connectivity, this potential inflation is somewhat mitigated by condition comparisons, as inflated connectivity estimates should affect all conditions and subtract out during comparisons.

When interpreting bivariate connectivity analysis results, several key concepts should be considered. First, the phase lag between two electrodes is often not taken into consideration, and what matters is the consistency of the phase lag across time and/or trials. Second, a nonzero phase lag in connectivity does not necessarily imply a causal or directed relationship and should be interpreted cautiously. Third, phase-based and power-based measures of connectivity tend to reveal different patterns of results, likely reflecting different neurophysiological dynamics. Fourth, a distinction can be made between functional connectivity (covariation between activity fluctuations) and effective connectivity (causal influence of activity in one network over another) [65]. Finally,

many connectivity results can be confounded by volume conduction, and care must be taken to investigate and address this alternative hypothesis.

Taking into account all these problems and its advantages, we can perform connectivity measures based on several features that can be extracted from the signals:

- **Phase-based connectivity** analyses rely on the distribution of phase angle differences between two electrodes to measure functional coupling of neural populations. They are widely used, computationally fast, and have a neurophysiological interpretation. However, they are susceptible to temporal jitter and do not provide compelling evidence for directionality.
- **Power-based connectivity** analyses involve correlating time-frequency power between two electrodes across time or over trials. They provide flexible analyses for testing specific hypotheses and are relatively insensitive to temporal jitter. They are similar to connectivity measures used in fMRI.
- **Causal connectivity** tests whether variance in one signal can be predicted from variance in another signal earlier in time. It can dissociate directional connectivity and ignore simultaneous connectivity. However, it is sensitive to violations of stationarity and can be computationally time-consuming. This method has the advantage of detecting directional connectivity.
- **Information-based connectivity** is a robust method for detecting shared information between two variables. It can detect linear and nonlinear interactions but does not provide information on the type of relationship. It is sensitive to the number of histogram bins and can be computationally intensive.
- **Cross-frequency coupling** refers to a statistical relationship between activities in two different frequency bands. It can be linked across species and to computational models. However, it has a potentially huge search space, which can be time-consuming and requires many tests to control for during statistical evaluation.

All measurements based only on signal amplitude (such as power-based connectivity) rely on measuring correlations or coherence between different regions of the brain. However, a significant challenge with power-based measures is volume conduction. Volume conduction refers to the phenomenon where electrical activity from a single neural source is detected at multiple EEG electrodes. This happens because the electrical signals generated by the brain spread through the conductive tissues of the scalp and skull. Consequently, power-based measures can produce spurious results (two electrodes may appear to be correlated simply because they are detecting signals from a common source), rather than reflecting true interaction between distinct regions. As a consequence, phase-based metrics, which rely only on the signal phase, and cross-frequency coupling statistics (which measure interactions between different frequency bands) are less susceptible to spurious connectivity than power-based measures.

Connectivity methods that make use of the signal phase, requires some processing of the raw signal. An usual way to extract the instantaneous phase and amplitude is

through the analytic signal, that can be computed by means of the Hilbert Transform [Hilbert Transform \(HT\)](#). Named after the renowned German mathematician David Hilbert, the Hilbert transform is a linear operator that transforms a function of a real variable into another function of a complex variable. Its main function is to introduce a 90 degree phase shift in the signal across all frequencies. This transformation is particularly beneficial as it enables the creation of the analytic signal, which presents a complex valued representation of the initial real valued signal. This analytic signal also lacks of negative frequencies. For a signal $x(t)$, it is defined as:

$$\mathcal{H}[x(t)] = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{x(\tau)}{t - \tau} d\tau \quad (3.1)$$

and the analytic signal $z_i(t)$ for a signal $x(t)$ can be obtained as

$$z_i(t) = x_i(t) + j\mathcal{H}\{x_i(t)\} = a(t)e^{j\phi(t)} \quad (3.2)$$

One of the benefits of using the analytic signal is its ability to calculate instantaneous properties of a signal, such as amplitude, phase and frequency. These real time characteristics offer an understanding of how the signal changes over time.

From [3.2](#), it is straightforward to compute the instantaneous amplitude as

$$a(t) = \sqrt{\text{re}(z_i(t))^2 + \text{im}(z_i(t))^2} \quad (3.3)$$

and the instantaneous, unwrapped phase is

$$\phi(t) = \tan^{-1} \frac{\text{im}(z_i(t))}{\text{re}(z_i(t))} \quad (3.4)$$

Finally, the instantaneous frequency $f(t)$ is obtained by taking the time derivative of the instantaneous phase.:

$$f(t) = \frac{1}{2\pi} \frac{d\phi(t)}{dt} \quad (3.5)$$

In practice, for discrete-time signals, the derivative is often approximated using finite differences.

The instantaneous amplitude, also referred to as the signal envelope, indicates the strength of the signal at any given moment. It provides insight into the signal intensity over time. The instantaneous phase reveals details about the behavior of the signal, while the instantaneous frequency, which is derived from the instantaneous phase, shows how the frequency components of the signal evolve with time.

3.4.1.1 Cross-frequency coupling

[Cross Frequency Coupling \(CFC\)](#) refers to the statistical relationship between oscillatory activities in different frequency bands and is considered a mechanism by which the brain coordinates information across different temporal scales. This concept is important because neural networks often rely on the integration of activities occurring at different frequencies to support cognitive processes, including perception, memory,

and attention.

There are several methods for Cross-frequency coupling estimation:

1. **Phase-To-Amplitude Coupling (PAC)**: PAC occurs when the phase of a lower frequency oscillation modulates the amplitude of a higher frequency oscillation. This cross-frequency coupling is thought to facilitate communication between brain regions and support cognitive functions like memory and attention. One of the critical points in PAC calculation is the filtering of the signals to extract the band in which phase and amplitude are computed. Since phase distortion has to be avoided for the proper computation of PAC, bandpass FIR (Finite Impulse Response) filters to extract phase (low-frequency) and amplitude (high-frequency) signals for phase-amplitude coupling (PAC) analysis. Moreover, according to [22, 195], the bandwidth of the filter used to extract the phase signal has to be 2-3 times narrower than the bandwidth of the filter used to extract the amplitude signal. For FIR filter design, the following equations can be considered:

- **Filter order (N):**

$$N \propto \frac{f_s}{\Delta f} \quad (3.6)$$

where:

- f_s : Sampling frequency.
- Δf : Transition bandwidth between passband and stopband.

- **Slope (in dB/decade):**

$$\text{Slope (dB/decade)} \approx \frac{20 \cdot \log_{10}(A)}{\Delta f \cdot \log_{10}(2)} \quad (3.7)$$

where:

- A : Attenuation in the stopband (e.g., 10^{-3} for 60 dB, 10^{-4} for 80 dB).
- Δf : Transition bandwidth in Hz.

A common way to compute the bandwidth used in phase and amplitude bands is considering that the bandwidth of the high-frequency signal (amplitude) is proportional to the bandwidth of the low-frequency signal (phase), scaled by the ratio of their central frequencies. This can be expressed as:

$$\text{BW}_{\text{amplitude}} \propto \text{BW}_{\text{phase}} \cdot \frac{f_{\text{phase}}}{f_{\text{amplitude}}} \quad (3.8)$$

Given the growing interest in PAC, various methods have been proposed to quantify this phenomenon in neurophysiological data: Equation 3.8 expresses the proportional relationship between the bandwidths of the amplitude and phase signals, scaled by their respective central frequencies. In this thesis, we used the criteria in [195] for FIR filters. Thus, phase filter uses 2–4 Hz bandwidth (e.g., theta: 4–8 Hz) while amplitude filter bandwidth is typically 10–100 Hz,

depending on the central frequency. Additionally, the filter slope is steep enough to minimize spectral leakage, with attenuation of 60–80 dB.

Mean Vector Length (MVL)[22]. calculates the length of the average vector formed by the phase of the low frequency signal and amplitude of the high frequency signal. Longer vector lengths indicate stronger coupling [22]. It is particularly effective under noisy conditions and short data segments like those found in EEG. However, MVL's performance is amplitude dependent.

$$\text{MVL} = \left| \frac{1}{N} \sum_{n=1}^N A_{\text{high}}(n) e^{i\theta_{\text{low}}(n)} \right| \quad (3.9)$$

Where:

- N is the total number of data points
- A_{high} is the amplitude of the high frequency signal
- θ_{low} is the phase of the low frequency signal

The idea behind this metrics is to use the analytic signal of the low frequency signal, multiply each of this complex points with the amplitude of the high frequency signal and average it. If an important part of the points is multiplied by relatively high values in the same phase, the average length would increase, thus indicating coupling. Otherwise, each vector in the complex plane will cancel each other and the average value would be smaller or close to zero. This behavior can be seen in Fig.3.9b. In the polar plot are represented each vector of the low frequency analytic signal multiplied by the amplitude of the high frequency signal in that time point. In the left column (representing coupling), it can be seen a grouping of high values points towards 0° phase and in red the average (MVL) of all vectors in the plot. In the right column, there is no coupling as all the points are almost equally distributed in the plot, averaging down the MVL value.

Modulation Index (MI)[193]. The Modulation Index is a widely used metric for phase-amplitude coupling (PAC) and measures the deviation of the amplitude of a high-frequency signal from a uniform distribution relative to the phase of a low-frequency oscillation. The MI is robust to noise and performs well in detecting both monophasic and biphasic coupling. It is particularly effective in cases where coupling is expected to be nonlinear or noisy, though it may require longer data segments for optimal performance.

The process involves binning phase angles from -180° to 180° , typically into 18 bins of 20° each, although the number of bins can vary depending on the analysis. The average amplitude of the amplitude-providing frequency is calculated within each phase bin, and this data is normalized.

The next step involves calculating Shannon entropy, which measures the predictability of the amplitude distribution across phase bins. A uniform distribution, where amplitude values are evenly spread across bins, would indicate no phase-amplitude coupling. Shannon entropy is then used in conjunction with

the Kullback-Leibler distance to quantify the divergence of the actual amplitude distribution from this uniform baseline. Finally, the Modulation Index is calculated by normalizing the Kullback-Leibler distance, providing a scalar value that represents the degree of phase-amplitude coupling. The MI is particularly robust against noise and confounding factors, making it suitable for short and noisy data segments.

$$MI = \frac{KL(U, X)}{\log N} \quad (3.10)$$

where:

$$KL(U, X) = \log N - H(p) \quad (3.11)$$

is the Kullback-Leibler distance between the uniform distribution U and the observed distribution X .

$$H(p) = - \sum_{j=1}^N p(j) \log p(j) \quad (3.12)$$

is the Shannon entropy of the observed distribution $p(j)$, where $p(j)$ is the normalized amplitude in the j -th phase bin. N is the total number of phase bins.

2. **Phase-To-Phase Coupling (PPC)**: This type of CFC describes a consistent relationship between the phases of two oscillations at different frequencies. It is less commonly studied than PAC but can still offer insights into how different frequency bands interact.

Phase-Locking Value (PLV)[200]. This method assesses the consistency of the phase relationship between the low-frequency oscillation's phase and the high-frequency oscillation's amplitude envelope.

$$PLV = \left| \frac{1}{N} \sum_{n=1}^N e^{i(\theta_{\text{low}}(n) - \theta_{\text{high}}(n))} \right| \quad (3.13)$$

Where:

- N is the total number of data points
- θ_{low} is the phase of the low frequency signal
- θ_{high} is the phase of the envelope of the high frequency signal

The idea of this metric is the next: if two signals are coupled, the amplitude of the high frequency signal (the envelope of the signal), should oscillate at the same rate as the phase signal (see Fig.3.8. Thus, subtracting both instantaneous phases, the one of the low frequency signal and the instantaneous phase of the envelope of the high frequency signal, should return a constant phase lag if there is coupling or an inconsistent lag if there is not. This can be represented in a

polar plot like in 3.9b where each green line represents the phase lag between the signals for every x point. The left column correspond to the coupled signals and the right one the no-coupled signals. In the coupled column, it can be seen a grouping of phase lags towards 0° , whereas the other column depicts not coupling at all. The red line represents the final value obtained for PLV value, being larger for the coupled signals and shorter for the no coupled ones.

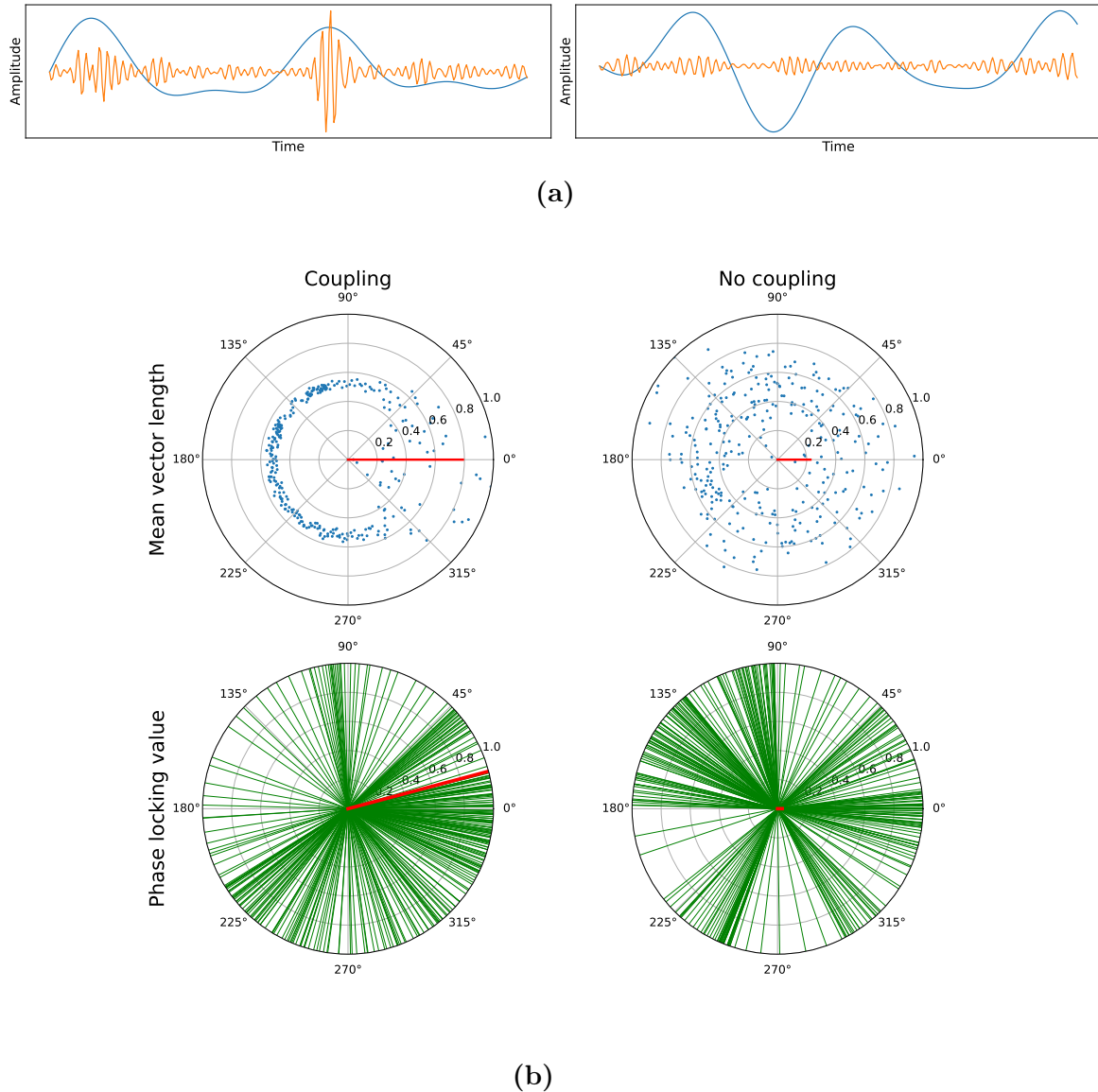


Figure 3.9: (a) Examples of Phase Amplitude Coupling (PAC) are showcased here. On the left side there is a strong coupling scenario where the faster signals amplitude (orange) is greatly influenced by the slower signals phase (blue). In contrast, the right side illustrates a scenario where there is little correlation between the slow signals phase and the fast signals amplitude, thus no coupling is happening.

(b) The left column exhibits plots for high PAC values for MVL and PLV metrics. These plots display a clustering of amplitude values at specific phase angles (0 in this case). On the right column displays polar plots, for low PAC values, where amplitude values are evenly spread across all phase angles indicating minimal or no coupling. The red line indicates the strength of the coupling for every metrics.

3.4.2 Surrogate Analysis

No matter the metric used to measure connectivity, there is always the possibility that the observed results could arise by chance. One commonly used method to ensure more

robust results is the permutation test.

Permutation testing revolves around the idea of rearranging (permuting) the sample data to explore every possible assignment of the observed values across the different groups in a study. Through this thorough reordering, a set of test statistics is generated under the null hypothesis, which assumes no real differences exist between the groups. This set of statistics forms an empirical distribution of the expected differences under the null hypothesis. By comparing the observed difference in the original data to this empirical distribution, we can assess the statistical significance of the results. Hence, the outcome of this process is a distribution of test statistics under the null hypothesis. This distribution reflects what we would expect to observe if the null hypothesis were true, that is, if there were no actual differences between the groups. It provides a reference point against which the observed test statistic can be evaluated.

Unlike permutation tests, traditional parametric tests rely on theoretical probability distributions (such as the normal, t, or chi-squared distributions) and make assumptions about the parameters of these distributions, like the mean and variance. For instance, a Student's t-test assumes that the data are drawn from populations following a normal distribution with equal variances. When these assumptions are not met, the results may lead to incorrect conclusions. Permutation tests, however, are advantageous in that they do not rely on any assumptions about the data's distribution. This makes them particularly valuable when working with small sample sizes, where parametric tests can sometimes lack accuracy. However, this does not come without drawbacks. These tests are often computationally expensive, as they require computing the same metric for every possible permutation. To mitigate this effect, normally a predefined number of permutations is set.

In particular, in bivariate connectivity analysis, this is done by splitting one signal into segments and rearranging them before calculating the connectivity metric between the two signals. By iteratively repeating this process, a null distribution is created, representing the hypothesis that the observed connectivity metric value could arise purely by chance.

Fig. 3.10 illustrates the standardization of the connectivity metric value relative to this null distribution. The formula for the standardized z-score is:

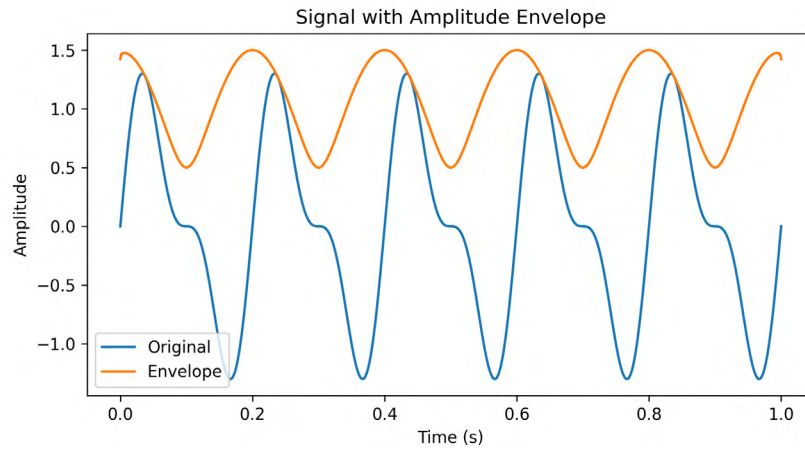
$$z_{\text{metric}} = \frac{\text{Metric} - m_{\text{metric}}^{\text{null}}}{s_{\text{metric}}^{\text{null}}} \quad (3.14)$$

where $m_{\text{metric}}^{\text{null}}$ and $s_{\text{metric}}^{\text{null}}$ represent the mean and standard deviation of the null distribution, respectively. This z-score indicates how far the observed connectivity metric value is from the mean of the null distribution in terms of standard deviations.

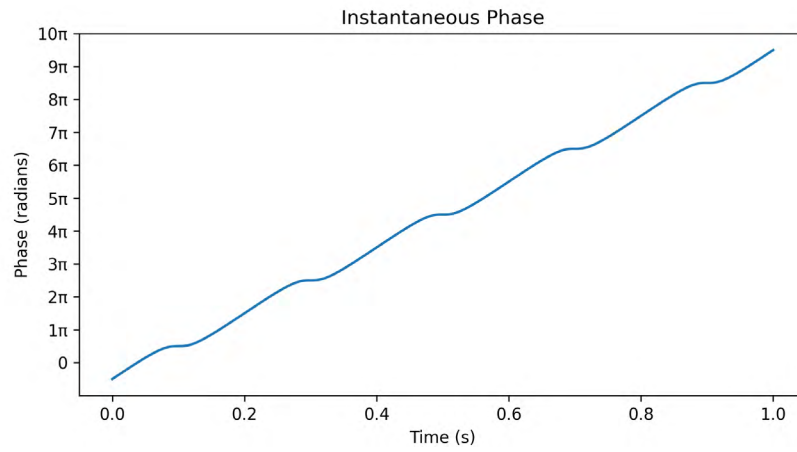
The z-score thresholds are essential for significance testing:

- In the left histogram, a low z-score (e.g., $z=0.3$) suggests the connectivity metric value is close to the null hypothesis distribution, implying no significant connectivity.
- In the right histogram, a higher z-score (e.g., $z=2.3$) suggests a significant deviation from the null distribution, indicating potential connectivity.

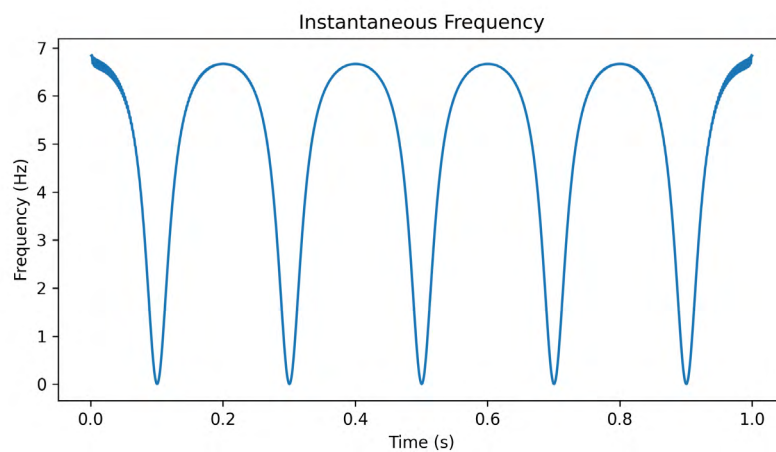
For a one-tailed test with a 95% confidence interval, a z-score of 1.65 serves as the critical threshold for significance. If the observed Z value exceeds this threshold, the connectivity metric value is statistically significant, indicating that the connectivity is likely not due to chance.



(a)



(b)



(c)

Figure 3.7: Example of an analytic signal obtained from Hilbert Transform. The signal is composed of two signals, one at 5Hz and another at 10Hz. (a) shows the original signal along with its envelope. (b) represents the unwrapped instantaneous phase, and (c) the instantaneous frequency.

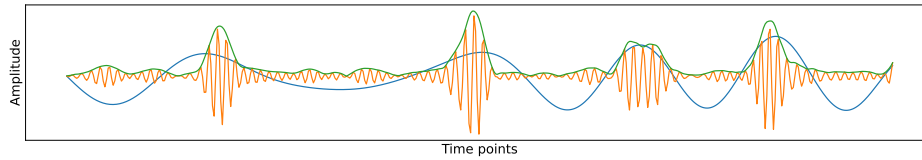


Figure 3.8: The idea behind PLV is that the amplitude (green) of the high frequency signal (orange) should oscillate at the same frequency that the low frequency signal (blue) when there is coupling between both signals, otherwise if not.

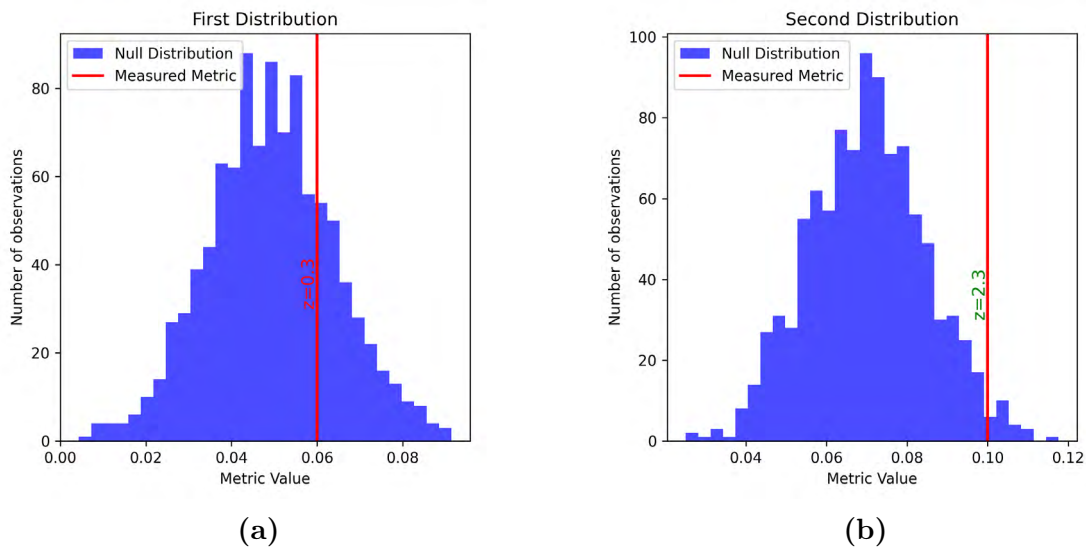


Figure 3.10: Histograms showing the null distribution of a connectivity metric with the observed metric value (red line) and corresponding z-scores. (a) represents a low z-score ($z = 0.3$), indicating no significant deviation from the null distribution, while (b) shows a higher z-score ($z = 2.3$), suggesting statistical significance. The z-scores are displayed rotated and positioned along the red lines for clarity.

Chapter 4

Artificial Intelligence

Artificial Intelligence (AI) has had an impact on various industries, particularly in the field of biomedical signal processing. The concept of AI has evolved significantly since its inception in the 1950s, moving from rule-based systems to more sophisticated learning algorithms. Today, AI encompasses a wide range of techniques and approaches, with machine learning as a particularly powerful subset. Machine learning enable systems to improve their performance through learning from data instead of relying only on direct programming. In the context of signal processing, machine learning methods have transformed how we analyze and make sense of intricate physiological data. These techniques are especially effective in handling the noisy and non-linear nature of biomedical signals, like **EEG**. Thus, in the field of biomedicine, the combination of signal processing and artificial intelligence techniques have enabled to extract complex patterns from data and to avoid the use of predefined features. As a matter of fact, one of the most important contributions of AI and, in particular, deep learning approaches, is the ability to learn features directly from data for a particular problem.

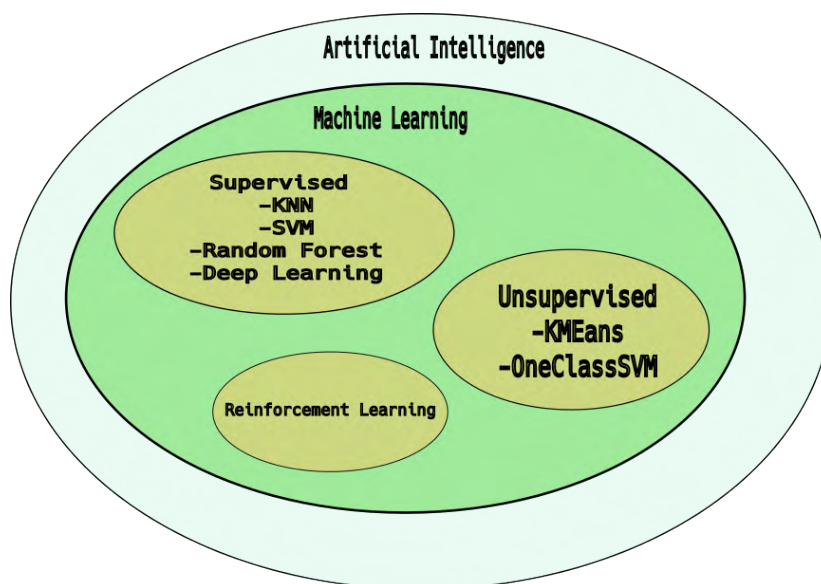


Figure 4.1: Artificial Intelligence field. Machine learning and deep learning as a subset of it.

Nevertheless, the inherent complexity of the models and the difficulty in interpreting how it arrives at a specific decision or prediction makes the AI based models to be considered black boxes. This way, the construction of interpretable models (XAI) or the use of explainability methods are essential to extract information about the rules and the process learned by the algorithm (a neural network, for example) to provide an outcome from a specific input. In other words, this is what we could call *opening* the black box (4.2).

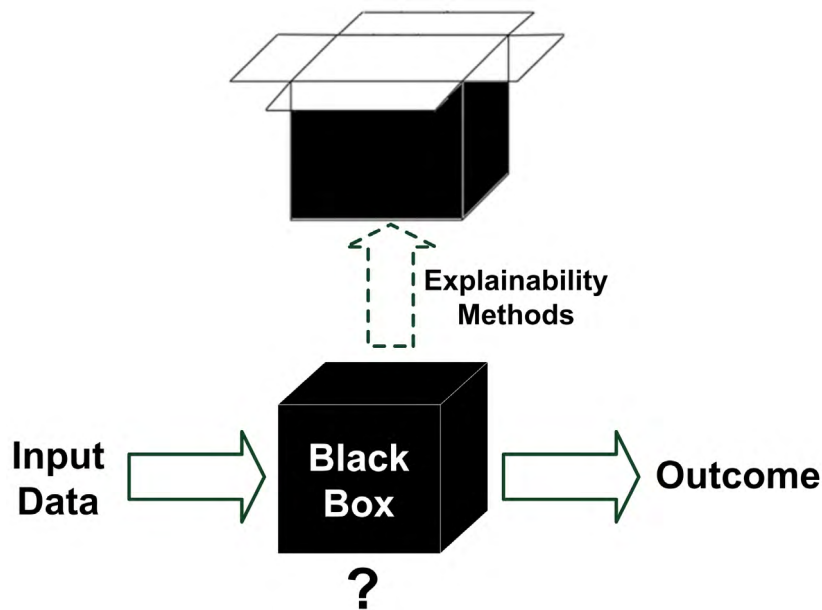


Figure 4.2: Models explainability. Openening the black box.

4.1 Machine Learning

Artificial Intelligence is a set of methods and algorithms that allow to learn from data, computing features and transforming the information into knowledge. It includes concepts such Machine Learning, Deep Learning while different learning paradigms can be applied, as shown in Fig. 4.1.

Machine learning can be further subdivided depending on the algorithm learning strategy (see Fig. 4.1). We can divide it in three main concepts:

- **Supervised learning.** In this type of learning, the data provided to the algorithm is labeled. This means that for each input in the training data, there is a corresponding correct output or target. The goal of supervised learning is to learn a function that can map new, unseen inputs to their correct outputs. Seeing multiples examples of input-output pairs, the algorithm adjusts its internal parameters to minimize the error between the predicted output and the true output. This leads to two variants; if the label is categorical it is called classification and if it is a number it is called regression. Fig. 4.3

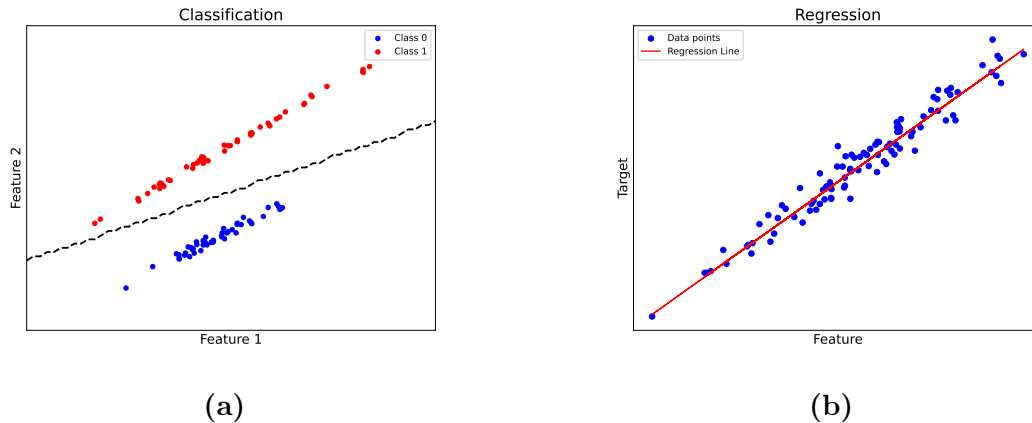


Figure 4.3: Supervised Learning. (a) represents a classification problem and (b) a regression problem. In classification, a new point will be classified as one class or another depending on where it falls at one side or another of the boundary (black line). In regression, the value returned by the algorithm for a new point will be the one best approaches the red line.

- **Unsupervised learning.** This type of learning works with unlabeled data, meaning the algorithm is given input data without any corresponding output labels. The goal is to discover hidden patterns, structures, or relationships within the data. This is particularly useful in scenarios where labeled data is scarce or expensive to obtain, or when exploring datasets to uncover hidden patterns. Fig. 4.4
- **Semi-supervised learning.** Semi-supervised learning bridges the gap between supervised and unsupervised learning by leveraging both labeled and unlabeled data. In this method, a small portion of the dataset is labeled, while the majority remains unlabeled. By propagating label information across similar data points, this learning method can significantly reduce the annotation burden and enhance model generalization. This approach is particularly valuable in scenarios with limited labeled data, such as medical imaging, speech recognition, and natural language processing, where obtaining fully labeled datasets is expensive or time-consuming.
- **Reinforcement learning.** In this approach, the parameters are updated based on a reward. The algorithm acts like an agent in an environment, making decisions and learning through trial and error. The goal is for the agent to learn a policy (a strategy for choosing actions) that maximizes the cumulative reward over time. This approach is inspired by behavioral psychology and how humans and animals learn from experience.

4.1.1 Training Process

The learning process in machine learning involves three main stages: training, validation, and testing. Each of these stages plays a crucial role in developing a model

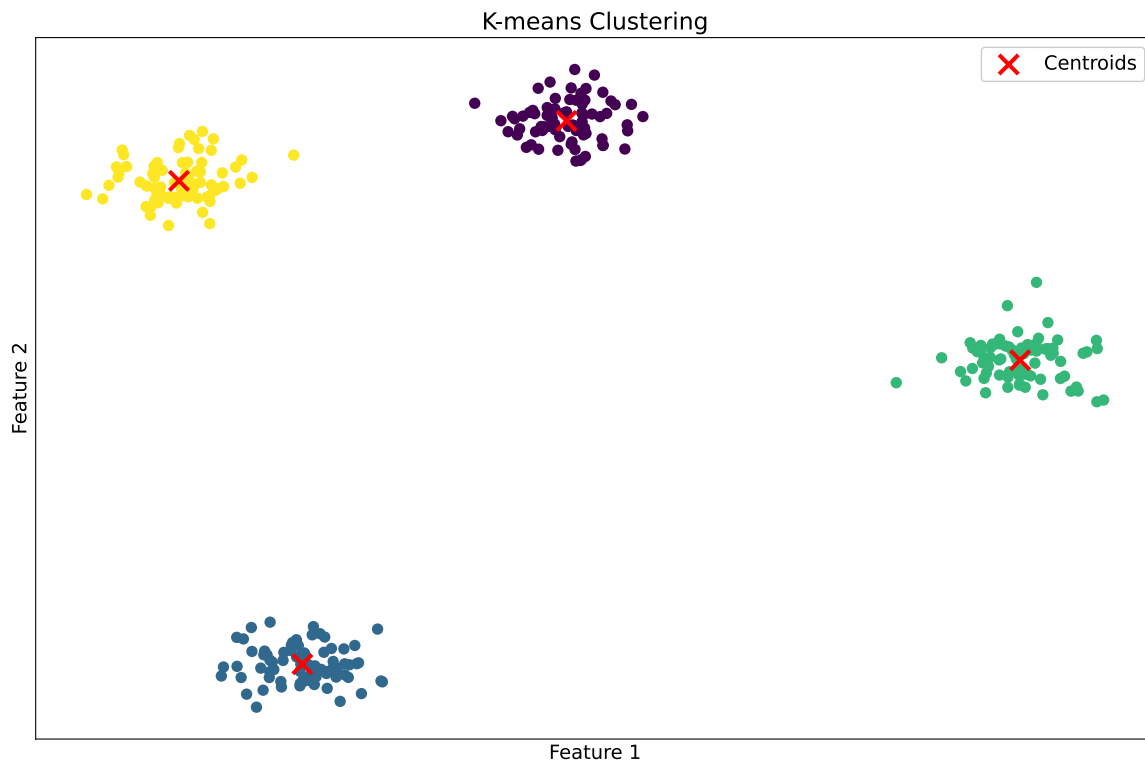


Figure 4.4: Unsupervised Learning. In unsupervised learning, the goal is to discover patterns in the data automatically. In this case, the data are fragmented into several clusters that are easily classified in four different classes.

that can effectively generalize from the data it has seen to make accurate predictions or decisions on new, unseen data and correspond to a subset of the data in which the original data is split. Every dataset contains examples of inputs and their corresponding correct outputs (in supervised learning) or just inputs (in unsupervised learning).

- **Training.** In this phase, the algorithm is exposed to the larger part of the dataset used to train. The algorithm processes this data, adjusting its parameters to minimize the difference between its predictions and the actual outcomes. This process of adjustment is typically guided by an optimization algorithm that aims to minimize a loss function, which quantifies the model's error.
- **Validation.** A separate set of data, called the validation set, is used to evaluate the model's performance during training. In this step, we try to predict the results for a subset of data on which the algorithm has not been trained. This helps in tuning the model's hyperparameters (settings that are not learned from the data but set prior to training). This is a crucial step in the learning process because we can assess the algorithm performance.
- **Testing.** The last step is to test the model on truly unseen data. Although the data in the validation set was also unseen by the algorithm (data the algorithm was not trained with), it was used to tune the hyperparameters. As a result, there

is a risk of indirect overfitting to the validation set, as information about that data is leaked into the learning process. This test set should be used only once, after the training and validation steps are complete, to provide a final evaluation of the model's generalization ability.

4.2 Overview of Machine Learning Algorithms

In this section, a selection is made of some of the huge number of algorithms in the machine learning toolbox; those used in the works for this thesis.

4.2.1 Supervised Learning

Supervised learning is a paradigmatic machine learning approach where an algorithm learns a mapping function from labeled training data, minimizing a loss function. The fundamental objective involves inferring a generalized predictive model from a finite set of observations, where each data point consists of an input vector and its corresponding target variable. By employing optimization techniques such as gradient descent and leveraging loss functions like mean squared error or cross-entropy, the algorithm iteratively adjusts its internal parameters to minimize prediction error. This process enables the model to generalize from the training dataset and make accurate predictions on unseen instances within the same statistical distribution.

4.2.1.1 XGBoost: eXtreme Gradient Boosting

XGBoost, also known as eXtreme Gradient Boosting is a version of the gradient decision tree algorithm that emphasizes speed and performance. It has become popular for handling data due to its efficiency, accuracy and versatility. Developed by Tianqi Chen and his team [28], XGBoost has become an algorithm in both data science competitions and real world applications. It has been successfully applied in various domains, including Kaggle competitions, recommendation systems, fraud detection and even natural language processing.

In essence, XGBoost works by creating an ensemble of decision trees. These trees make predictions by following a series of if-then rules that split the data based on feature values to form a tree structure. Each internal node serves as a decision point to build a branch representing an outcome of that decision, while each leaf node indicates a prediction or classification label.

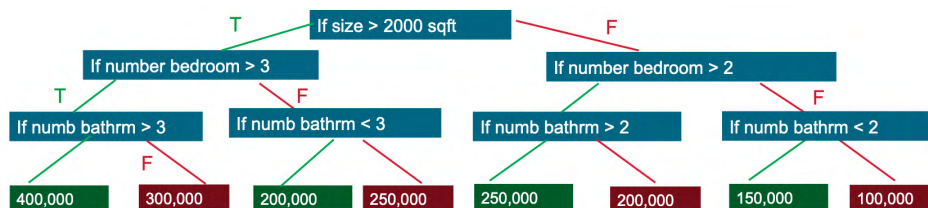


Figure 4.5: Illustration of a decision tree. Each decision split the data in several branches depending on a feature value [151]

Although individual decision trees are simple and easy to interpret, they often encounter issues with variance and potential overfitting to the training data. This is where techniques like XGBoost excel. Ensemble methods merge models to generate a robust and accurate predictive model. There are several ensemble techniques like bagging; which involves creating multiple subsets of the original dataset by randomly selecting with replacement, random forests; a variation of bagging introduces feature randomness in constructing decision trees and boosting; builds models sequentially with each new model focusing on correcting the mistakes of ones.



Figure 4.6: Illustration of a boosting ensemble. Each new classifier tries to correct the errors in the previous one [70]

XGBoost is part of the boosting methods group. It constructs trees one by one, where each new tree helps correct errors from previous trained trees. This incremental approach allows XGBoost to better understand patterns in the data, leading to better predictions.

Working Mechanism

- **Initialization:** The algorithm starts with an initial prediction, which is often the mean value of the target variable for regression tasks or a similar baseline for classification tasks.
- **Boosting Process:** In each subsequent step, the algorithm adds a new tree to the model. Rather than creating trees independently, each new tree is built to minimize the errors made by the sum of all previous trees. This is done by focusing on the instances that were poorly predicted by the previous ensemble.
- **Gradient Descent:** XGBoost utilizes gradient descent to optimize the performance. It uses the gradients of the loss function (which measures the difference

between the predicted and actual values) to guide the construction of each new tree. The gradients help in identifying the direction and magnitude of adjustment needed for each prediction.

- **Regularization:** To prevent overfitting (where the model performs well on training data but poorly on new data), XGBoost incorporates regularization techniques. These techniques penalize overly complex models. Specifically, it uses L2 regularization on the leaf weights and a complexity penalty on the number of leaves in the trees.
- **Tree Pruning:** Unlike traditional decision tree algorithms that may grow deep trees, XGBoost uses a process called tree pruning. This process helps in reducing complexity by removing branches from the tree that do not provide enough predictive power, thus aiding in the creation of simpler and more generalizable models.
- **Parallel and Distributed Computing:** One of the strengths of XGBoost is its ability to utilize parallel and distributed computing. This capability significantly speeds up the training process, especially on large datasets.

Mathematical Notation

The objective function for XGBoost can be formulated as follows:

$$\mathcal{L}(\phi) = \sum_{i=1}^n l(\hat{y}_i, y_i) + \sum_{k=1}^K \Omega(f_k) \quad (4.1)$$

where:

- $\mathcal{L}(\phi)$ is the objective function to minimize.
- l is a differentiable convex loss function that measures the difference between the prediction $\hat{y}_i$ and the target y_i .
- $\hat{y}_i$ is the predicted value for the i -th instance.
- y_i is the actual target for the i -th instance.
- $\Omega(f_k)$ is the regularization term which helps to smooth the final learnt weights to avoid overfitting.
- n is the number of instances.
- K is the number of trees.

The predicted value for each instance is:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), \quad f_k \in \mathcal{F} \quad (4.2)$$

where:

- $\mathcal{F}$ is the space of regression trees (a set of CARTs).
- K is the total number of trees.
- f_k is a function in the functional space $\mathcal{F}$ that represents a tree structure and maps an instance x_i to a predicted value.

The regularization term $\Omega(f_k)$ is defined as:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (4.3)$$

where:

- γ is the complexity parameter which penalizes the number of leaves in the tree.
- T is the number of leaves in the tree.
- λ is the L2 regularization term on the leaf weights.
- w represents the vector of scores on leaves.

XGBoost trains the model in an additive manner. At step t , the model adds a new tree f_t to minimize the following:

$$\mathcal{L}^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \quad (4.4)$$

Using the second-order Taylor approximation around the prediction $\hat{y}_i^{(t-1)}$:

$$\mathcal{L}^{(t)} \approx \sum_{i=1}^n \left[l(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \quad (4.5)$$

where:

- $g_i = \frac{\partial l(\hat{y}_i, y_i)}{\partial \hat{y}_i^{(t-1)}}$ is the first-order gradient.
- $h_i = \frac{\partial^2 l(\hat{y}_i, y_i)}{\partial (\hat{y}_i^{(t-1)})^2}$ is the second-order gradient (Hessian).

To optimize the tree structure, the objective function reduces to:

$$\mathcal{L}^{(t)} = \sum_{j=1}^T \left[\left(\sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \gamma T \quad (4.6)$$

where w_j is the score on the j -th leaf, and I_j is the instance set of leaf j . The optimal weight for each leaf w_j is given by:

$$w_j^* = - \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \quad (4.7)$$

Substituting w_j back into the objective function gives the optimal value:

$$\mathcal{L}^{(t)} = -\frac{1}{2} \sum_{j=1}^T \frac{\left(\sum_{i \in I_j} g_i \right)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T \quad (4.8)$$

4.2.1.2 KNearest Neighbors

KNN is known for its simplicity and effectiveness. It is a non-parametric method used for both, classification and regression tasks. It works by classifying a new data point looking at the most k closest neighbors in the feature space. The "k", in KNN indicates how many neighbors are taken into account when deciding. This method is based on the concept that comparable data points often cluster together closely.

For the algorithm to work, it needs a set of data that includes examples with labels, where each sample contains characteristics and a corresponding category or numerical value. When a new data point without a label is introduced, KNN measures the distance between this point and all other k points in the dataset using a distance measurement. In tasks involving classification KNN assigns the most frequent category among these neighbors to the new point. In regression scenarios, it typically computes the average of the k closest neighbors.

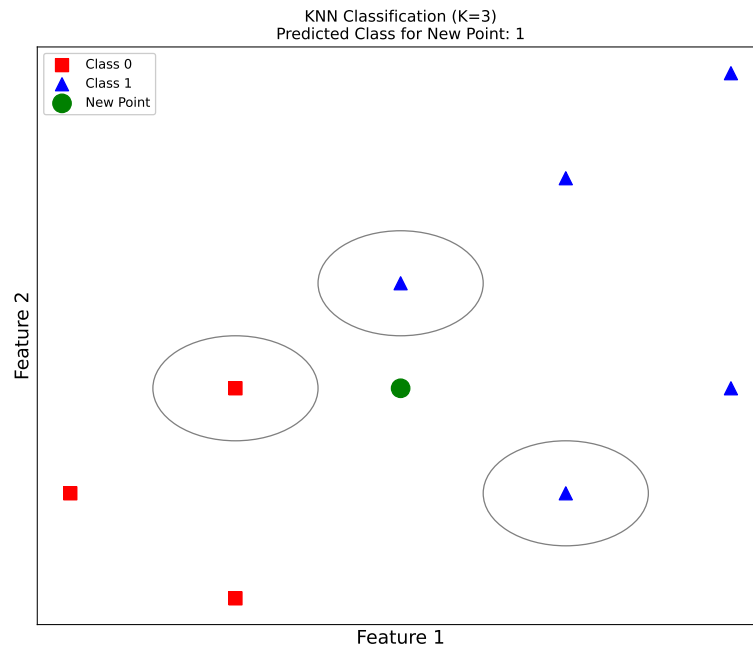


Figure 4.7: KNN classification task. Knowing the K nearest neighbors points, the new point will be classified as the same class as the majority of those K nearest neighbors.

One of the key advantages of KNN is its simplicity and interpretability. KNN is easy to understand and implement, making it an excellent starting point to establish a baseline prior to try any more complex algorithm or solution. Another point to take

into account is that it is a lazy learning algorithm. This means that it doesn't require a training phase. Instead, it memorizes the training dataset and performs computations only when making a new prediction.

KNN is particularly effective in scenarios where the decision boundary is irregular. It can capture complex patterns in the data without making strong assumptions about the underlying distribution. This makes it versatile and applicable to a wide range of problems. However, it has its drawbacks. The entire training dataset must be searched for each prediction, become computationally expensive and memory-intensive for large datasets. It is also sensitive to the scale of features and can be affected by irrelevant or redundant features.

As seen before, the most important hyperparameters to tune are the metric to use which will define the space in which the points are and the number of neighbors when classifying. Usually, the chosen metric is Minkowski distance in which a parameter p defines the type of vectorial space, being 2 for a Euclidean space or 1 for Manhattan distance. Additionally, choosing the optimal value of k can be challenging and often requires experimentation or cross-validation.

The *curse of dimensionality* is another significant challenge for KNN; as the number of features increases, the concept of distance becomes less meaningful, potentially leading to decreased performance. This issue underscores the importance of feature selection and dimensionality reduction techniques when working with high-dimensional data.

4.2.1.3 Support Vector Machines

Support Vector Machines (SVM) are a powerful and versatile class of supervised machine learning algorithms used for both classification and regression tasks. Developed in the 1990s by Vladimir Vapnik [37], SVM have become popular due to their ability to handle complex, high-dimensional data and their effectiveness.

The idea behind SVMs is to find the hyperplane that separates different classes of data points in a high-dimensional space. A hyperplane is a decision boundary that separates different classes in the feature space. The hyperplane is defined by:

$$\mathbf{w}^T \mathbf{x} + b = 0 \quad (4.9)$$

where:

- $\mathbf{w}$ is the weight vector.
- b is the bias term.
- $\mathbf{x}$ is the input feature vector.

This hyperplane is chosen to maximize the margin between the closest data points of different classes, known as **support vectors**. The larger the margin, the better the generalization of the model to unseen data. It divides the space into two halves; one side where the points are classified as one class (e.g., (+1)) and the other side where the points are classified as the other class (e.g., (-1)).

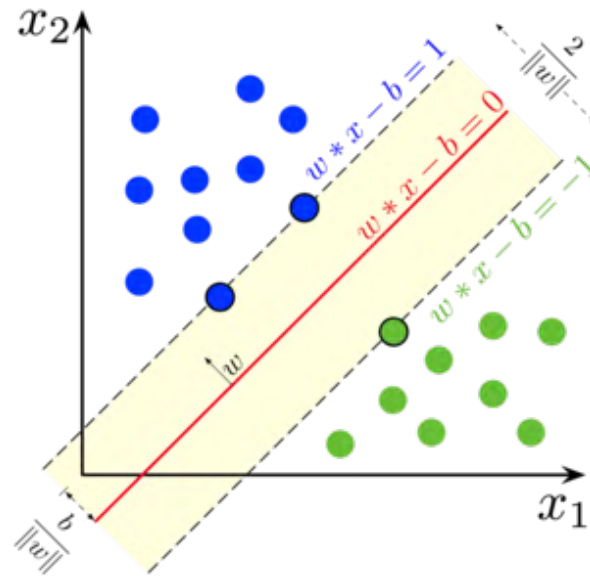


Figure 4.8: Hyperplane (red) solution for a SVM. The margins are maximized between the plan and the closest datapoints. The datapoints bolded are the **support vectors** [118]

Sometimes, the data is not linear separable and the input space is transformed into a higher dimensional one using what is called the *kernel trick*, where the data can be linearly separable. Common kernel functions include linear, polynomial, and radial basis function (RBF) kernels. The choice of kernel depends on the nature of the data and the problem at hand. This flexibility allows SVMs to adapt to various types of data distributions and capture complex relationships between features.

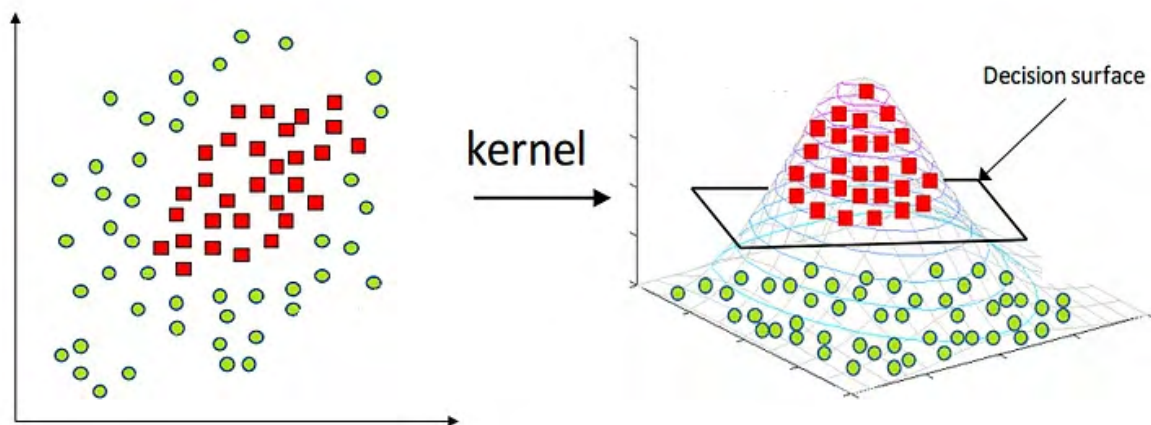


Figure 4.9: In the kernel trick, the input data is transformed into a higher dimensional space where a hyperplane can separate the data. Different types of kernels result in different types of spaces [214]

SVMs also offer good resistance to overfitting, especially in high-dimensional spaces. This is because the optimization objective of SVMs focuses on maximizing the margin

rather than minimizing the training error. As a result, SVMs can often generalize well to unseen data, even when the number of features is large compared to the number of training samples. This property makes SVMs particularly useful in domains where the number of features greatly exceeds the number of observations, like the majority present in medicine field.

Some of its drawbacks are that they can be computationally expensive, especially for large datasets, as the training time scales poorly with the number of samples. Another challenge is the selection of appropriate hyperparameters, such as the choice of kernel and regularization parameter, which can significantly impact the model performance.

Mathematical Notation

The optimization problem for a Support Vector Machine (SVM) can be formulated as follows:

$$\min_{\mathbf{w}, b, \xi_i} \frac{1}{2} \mathbf{w}^\top \mathbf{w} + C \sum_{i=1}^n \xi_i, \quad (4.10)$$

subject to

$$y_i(\mathbf{w}^\top \phi(\mathbf{x}_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall i = 1, \dots, n, \quad (4.11)$$

where:

- $\mathbf{w}$ is the weight vector,
- b is the bias term,
- ξ_i are the slack variables,
- C is the regularization parameter,
- $\mathbf{x}_i$ represents the input vector for instance i ,
- y_i represents the class label for instance i ,
- $\phi(\mathbf{x}_i)$ is a feature map that transforms the input vector into a higher-dimensional space.

The decision function of the SVM is given by:

$$f(\mathbf{x}) = \mathbf{w}^\top \phi(\mathbf{x}) + b. \quad (4.12)$$

To facilitate non-linear classification, the kernel trick can be applied. The dual form of the SVM optimization problem is given by:

$$\max_{\alpha_i} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j), \quad (4.13)$$

subject to

$$\sum_{i=1}^n \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C, \quad \forall i = 1, \dots, n, \quad (4.14)$$

where:

- α_i are Lagrange multipliers,
- $K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^\top \phi(\mathbf{x}_j)$ is the kernel function.

The decision function in terms of the Lagrange multipliers becomes:

$$f(\mathbf{x}) = \sum_{i=1}^n \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b. \quad (4.15)$$

Common choices for the kernel function $K(\mathbf{x}_i, \mathbf{x}_j)$ include:

- Linear kernel: $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^\top \mathbf{x}_j$,
- Polynomial kernel: $K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^\top \mathbf{x}_j + c)^d$,
- Radial basis function (RBF) kernel (Gaussian): $K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right)$,
- Sigmoid kernel: $K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\kappa \mathbf{x}_i^\top \mathbf{x}_j + \theta)$.

4.2.2 Unsupervised Learning

Unsupervised learning constitutes a machine learning paradigm focused on extracting patterns from unlabeled datasets through statistical inference and dimensional reduction techniques. The core computational objective involves discovering latent representations, identifying inherent data structure, or transforming high-dimensional feature spaces without prescribed target variables. Clustering algorithms and principal component analysis are examples that leverage probabilistic and geometric approaches to uncover hidden statistical dependencies, manifold structures, and intrinsic data geometries. By employing optimization criteria such as minimizing within-cluster variance or unsupervised learning methods enable feature extraction, dimensionality reduction, and knowledge discovery from complex data distributions. Alternatively, autoencoders are neural networks that also learn in an unsupervised manner, obtaining a latent representation in a lower dimensional space. A representative example of unsupervised learning is the Self-Organizing Map (SOM), which are competitive neural networks that perform vector quantization by the projection of the high-dimensional input space into a low-dimensional one, preserving the topological representations and intrinsic data relationships. This enables nonlinear visualization and clustering of complex, multidimensional datasets through adaptive spatial mapping. SOM algorithm is explained in Chapter 6 as it is used to develop an anomaly detection method.

4.2.2.1 Clustering

4.2.2.2 One Class SVM

OneClassSVM is an extension of the traditional Support Vector Machine (SVM) algorithm. OneClassSVM is trained on a single class of data, making it particularly useful in scenarios where normal instances are abundant, but anomalies or outliers are rare or difficult to obtain. Like vanilla SVMs, the objective of OneClassSVM is to learn a decision boundary that encompasses the majority of the normal data points in the feature space. This boundary is then used to classify new, unseen data points as either normal or anomalous.

The train procedure is the same as in two-class SVMs, but, since only one class is available at the moment of training, the origin of coordinates is selected as an example of the negative class. At the inference stage, anything that falls from the boundary decision is classified as an outlier.

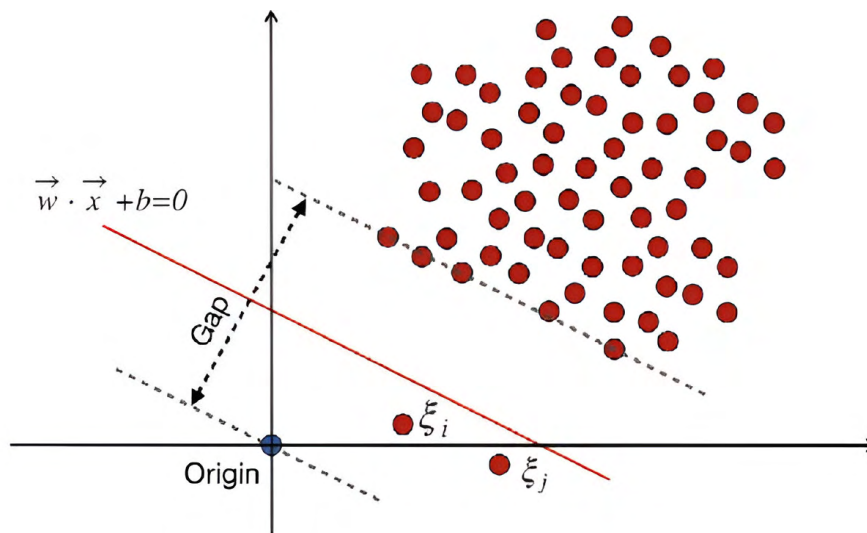


Figure 4.10: In one class SVM the origin is taken as a negative class sample. Anything that falls outside the boundary of decision of the positive class is classified as an outlier (ξ_i, ξ_j) [113]

4.3 Deep Learning

Deep learning, has revolutionized the field in recent years. This powerful approach to AI has led to breakthrough advancements in various domains, including computer vision, natural language processing, speech recognition, and robotics.

The basic element of deep learning is inspired by the biological neurons. The perceptron, a simple binary classifier, laid the groundwork for future neural network architectures. The perceptron was designed to perform basic pattern recognition tasks and could learn from examples.

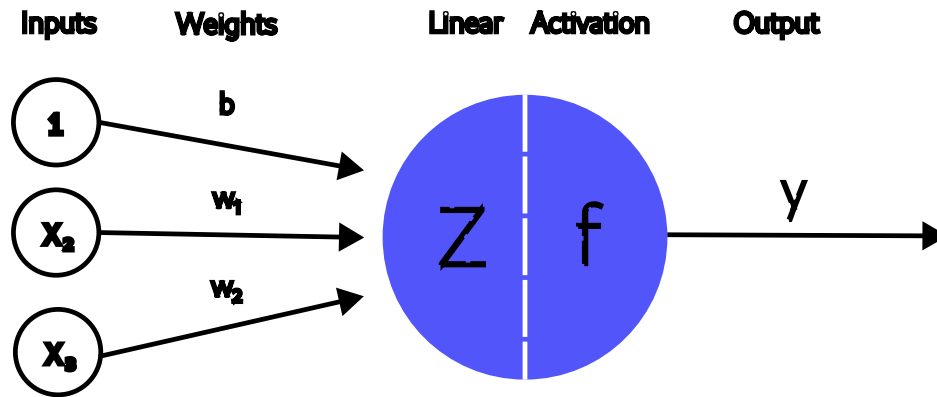


Figure 4.11: Perceptron, the basic element of the neural network. The inputs are multiplied by the weights (the learnable parameters) and summed together. Then, the previous result is passed through an activation function, giving the final result.

The working process is simple, and summed up in equation 4.16. First, each input is multiplied by its corresponding weight. This weight is a learnable parameter that changes in every train iteration in order to adjust the desired output. Then, they are summed together and passed through an activation function, giving the output. There is a special weight that can be added to the perceptron called bias term. As its input, its value is always 1, although the weight value varies. It represents the constant offset or shift in the activation of neurons, allowing the model to capture patterns that cannot be represented solely by the input features.

One of the main problems of this simple algorithm is that it is not able to solve non-linearly separable problems, such as the XOR function. To overcome this limitation, researchers developed multi-layer perceptrons (MLPs), also known as feedforward neural networks. MLPs consist of an input layer, one or more hidden layers, and an output layer. The addition of hidden layers allowed these networks to learn more complex, non-linear relationships between inputs and outputs. The basic structure of an MLP can be represented mathematically as:

$$\text{Single perceptron output: } y = f\left(\sum_{i=1}^n w_i x_i + b\right) \quad (4.16)$$

$$y_1 = f(W_1 \cdot x + b_1) \quad (4.17)$$

$$y_2 = f(W_2 \cdot y_1 + b_2) \quad (4.18)$$

Where x is the input vector, n is the number of weights, W_1 and W_2 are weight matrices, b_1 and b_2 are bias vectors, f is a non-linear activation function (Section 4.3.3), and y_1 , y_2 are the output vectors.

Note that in equation 4.16, we are dealing with a single perceptron, but in equations 4.17 and 4.18 matrices of weights are created where each row corresponds to the weights of one single perceptron in that particular layer.

This tool had been in the toolkit since the 1950s, but it was not until the 2000s that its great potential was realized. The AI field experienced a resurgence in the

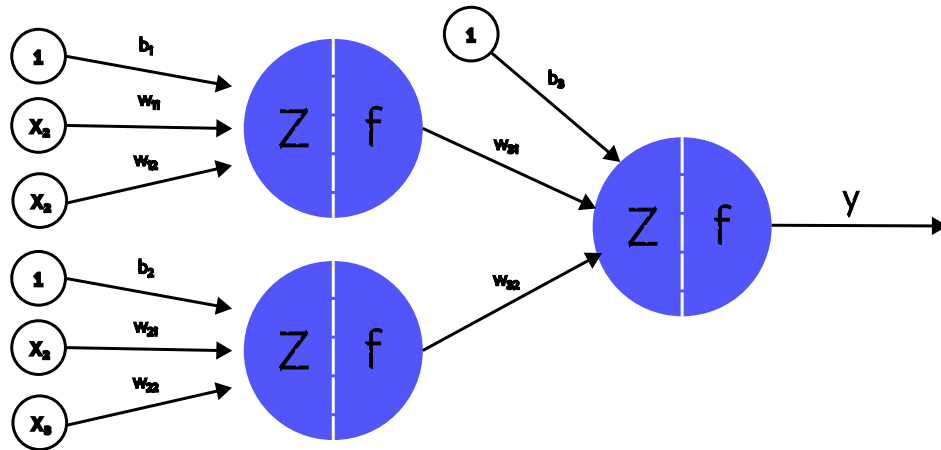


Figure 4.12: The multilayer perceptron is an evolution of a perceptron capable to deal with non-linearities. There is no inherent limit to how many layers or neurons per layer can be included in the network structure.

2000s, thanks to several key developments. The development of [GPU](#) and distributed computing systems gave the possibility to train larger and more complex neural networks efficiently. In addition, the expansion of the Internet has made it possible to collect vast amounts of data never before experienced. All this together has led to deep learning becoming the best-known branch of AI.

4.3.1 Convolutional Neural Networks

Along the way, new ways of bundling artificial neurons and creating new structures have emerged. Some of these are biologically based, others are not. One of the most influential architectures is the [CNN](#) by Lecun et al. [121]. CNNs, inspired by the visual cortex of animals, are particularly well-suited for image-related tasks, although they have been successfully applied in time series analysis aswell. They use convolutional layers to automatically learn hierarchical features from input data. The basic structure of a CNN can be represented as:

$$y = f_{FC}(f_{Pool}(f_{Conv}(x))) \quad (4.19)$$

Where f_{Conv} represents convolutional layers, f_{Pool} represents pooling layers, and f_{FC} represents fully connected layers. CNNs have achieved state-of-the-art results in image classification, object detection, and image segmentation tasks.

4.3.2 Recurrent Neural Networks

Another important architecture in deep learning is the [Recurrent Neural Network \(RNN\)](#). RNNs are designed to process sequential data by maintaining an internal state or *memory*. This makes them particularly useful for tasks involving time series or natural language. The basic RNN structure can be described as:

$$h_t = f(W_h \cdot h_{t-1} + W_x \cdot x_t + b) \quad (4.20)$$

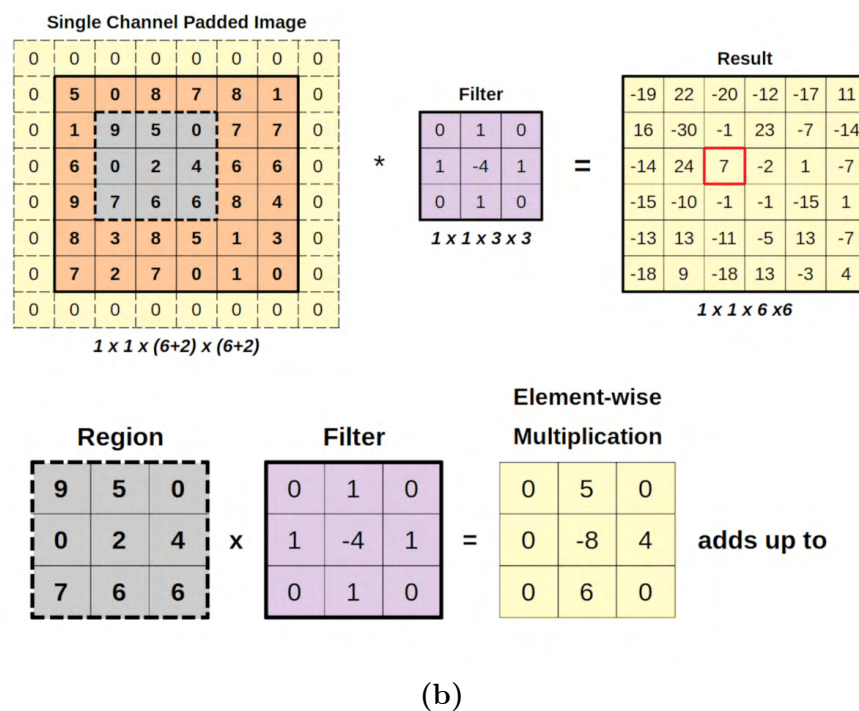
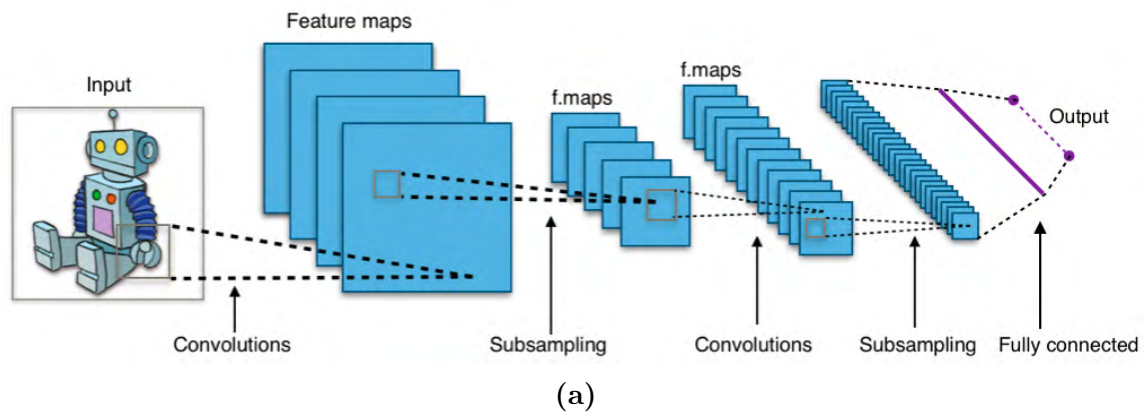


Figure 4.13: (a) A typical CNN network. In each layer, the network learns to extract a hierarchical feature of the training images. (b) The convolution operation used in CNNs. The filter parameters are what the network try to learn from the data [76]

Where h_t is the hidden state at time step t , x_t is the input at time step t , W_h and W_x are weight matrices, b is a bias vector, and f is a non-linear activation function.

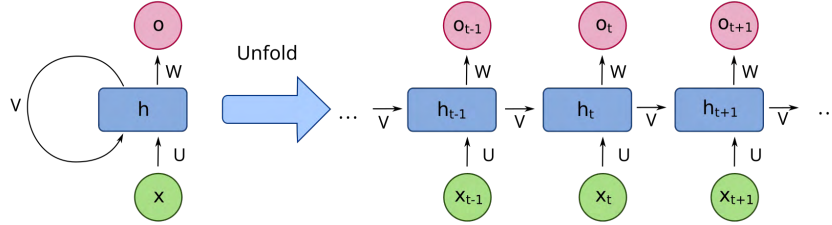


Figure 4.14: The recurrent network is a network block in which the input vector x is feed iteratively to the architecture. A hidden state h is maintained in every step, working as *memory*. [105]

Several developments to increase the power in RNNs have been proposed. Hochreiter et al. [94] introduced the **Long Short-Term Memory (LSTM)** architecture. LSTMs use gating mechanisms to control the flow of information, allowing them to capture long-term dependencies more effectively. A simplified version of the LSTM, called the Gated Recurrent Unit (GRU), was developed by Chung et al. [31], maintaining comparable performance with fewer parameters.

4.3.2.1 Long Short-Term Memory Cell

LSTMs are particularly effective in addressing the problem of gradient vanishing faced by traditional recurrent neural networks. Using a series of memory cells, LSTM layers can selectively remember or forget information over time, controlled by three gates: input, forgetting, and output. Mathematically, we can define the state of cell C_t , the hidden state (output) h_t , and the f_t forgotten, i_t entrance, and o_t exit gates, as follows:

Forget Gate:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (4.21)$$

Input Gate:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (4.22)$$

Output Gate:

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (4.23)$$

Cell Input Activation:

$$\tilde{C}_t = \tanh(W_C x_t + U_C h_{t-1} + b_C) \quad (4.24)$$

Cell State Update:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (4.25)$$

Hidden (Output) State Update:

$$h_t = o_t * \tanh(C_t) \quad (4.26)$$

where σ is the activation function of the sigmoid, $\tanh$ is the activation function of the hyperbolic tangent. W , U , and b are trainable weight matrices and bias vectors for each gate operation. Finally, $*$ represents the multiplication by element and x_t is the input at the time step t .

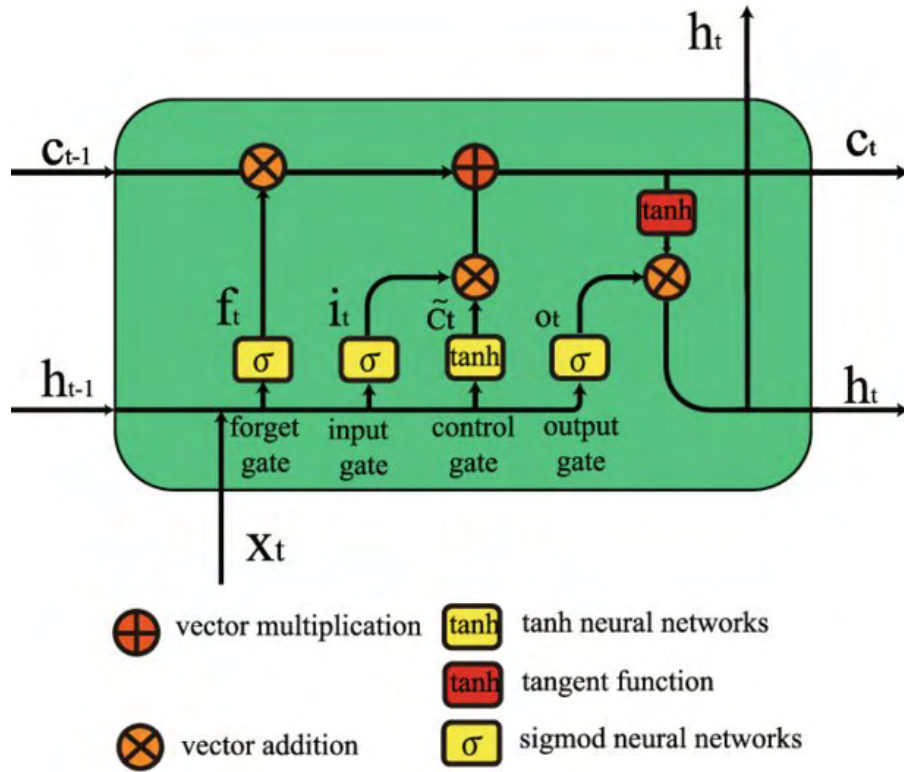


Figure 4.15: The LSTM block architecture showing input, forget, and output gates, along with the cell state [132]

4.3.3 Activation functions

Activation functions are components of neural networks. They introduce the non-linearity that allows these networks to model complex patterns and make accurate predictions. They are also essential for the backpropagation of gradients, which is the core mechanism behind the training of neural networks. The choice of activation function can significantly affect the ability of the network to converge during training, its accuracy, and its overall performance. Some of the most used functions are:

- Sigmoid. Sigmoid particularly useful for binary classification tasks, where outputs can be interpreted as probabilities.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (4.27)$$

- Tanh. It has an output range between -1 and 1, offering a zero-centered alternative to the Sigmoid function. This zero-centered nature can be advantageous

because it helps balance the gradients during training, potentially leading to a more efficient learning process.

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (4.28)$$

- ReLU. It introduces a simple yet effective non-linearity by setting all negative input values to zero while leaving positive values unchanged. One of the key advantages of ReLU is its computational efficiency, as it involves straightforward thresholding operations at the same time that helps with the vanishing gradient problem.

$$\text{ReLU}(x) = \max(0, x) \quad (4.29)$$

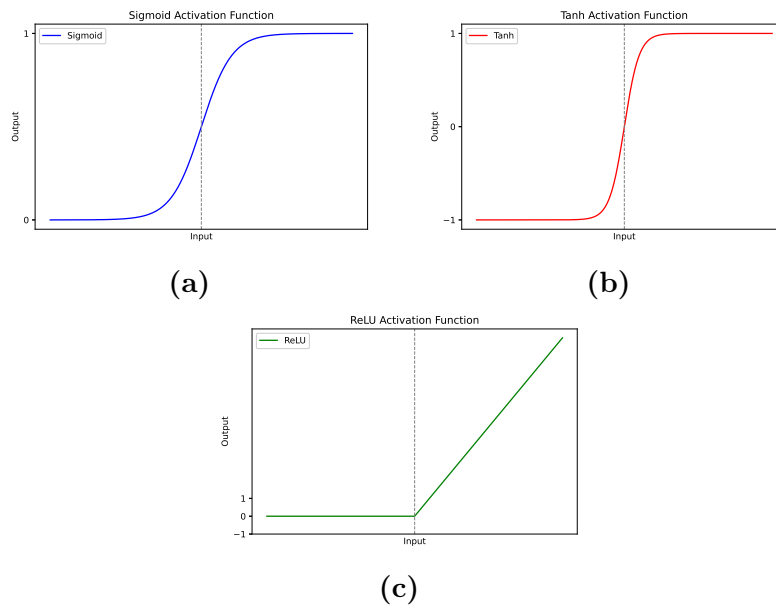


Figure 4.16: Three of the most used activation functions: **a)** Sigmoid, **b)** Tanh, **c)** ReLU. These functions introduce the non-linearity part that allows the networks to learn complex patterns.

4.3.4 Training

The objective of training a neural network is to find the optimal set of weights and biases that minimize a loss function. This function gives the error between the predicted output and the desired output. This is performed iteratively adjusting the weight of the network. The most widely used method to achieve it is the *backpropagation* algorithm.

Backpropagation is based on the principle of the chain rule of calculus, which allows us to compute the gradient of the loss function with respect to each weight and bias in the network. It involves two main phases: the forward and the backward pass.

1. In the forward pass, the input data is propagated through the network layer by layer, applying weights, biases, and activation functions at each layer to produce

the output. This phase computes the activations and weighted sums that are required for the backward pass.

For the input layer (layer 0), we have:

$$\mathbf{a}^{(0)} = \mathbf{x}$$

For each layer ($l = 1, 2, \dots, L$):

$$\mathbf{z}^{(l)} = \mathbf{W}^{(l)} \mathbf{a}^{(l-1)} + \mathbf{b}^{(l)} \quad (4.30)$$

$$\mathbf{a}^{(l)} = \sigma(\mathbf{z}^{(l)}) \quad (4.31)$$

Loss Calculation Suppose that we use a loss function $\mathcal{L}$, such as the mean squared error (MSE). For simplicity, let us denote the target output as $\mathbf{y}$ and the output from the network as $\mathbf{a}^{(L)}$. The loss can be written as:

$$\mathcal{L} = \mathcal{L}(\mathbf{a}^{(L)}, \mathbf{y}) \quad (4.32)$$

2. In the backward pass, the error is propagated back through the network from the output layer to the input layer. During this phase, the gradients of the loss function with respect to each weight and bias are computed, enabling us to update the parameters using an optimization algorithm such as *Gradient Descent*.

Compute the gradient of the loss with respect to the output of the last layer:

$$\delta^{(L)} = \nabla_{\mathbf{a}^{(L)}} \mathcal{L} \quad (4.33)$$

For instance, if using MSE as the loss function:

$$\delta^{(L)} = \frac{\partial \mathcal{L}}{\partial \mathbf{a}^{(L)}} = \mathbf{a}^{(L)} - \mathbf{y} \quad (4.34)$$

Backpropagate the gradient through each layer: For $l = L, L-1, \dots, 1$,

$$\delta^{(l)} = (\mathbf{W}^{(l+1)})^T \delta^{(l+1)} \circ \sigma'(\mathbf{z}^{(l)}) \quad (4.35)$$

where $\circ$ denotes the element-wise (Hadamard) product and σ' represents the derivative of the activation function.

Compute the gradients of the loss with respect to weights and biases:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{W}^{(l)}} = \delta^{(l)} (\mathbf{a}^{(l-1)})^T \quad (4.36)$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{b}^{(l)}} = \delta^{(l)} \quad (4.37)$$

Finally, update the weights and biases using a suitable optimization method, such as Gradient Descent:

$$\mathbf{W}^{(l)} := \mathbf{W}^{(l)} - \eta \frac{\partial \mathcal{L}}{\partial \mathbf{W}^{(l)}} \quad (4.38)$$

$$\mathbf{b}^{(l)} := \mathbf{b}^{(l)} - \eta \frac{\partial \mathcal{L}}{\partial \mathbf{b}^{(l)}} \quad (4.39)$$

One of the most important parameter that will control the training process is the learning rate (η). The selection of this parameter could cause the training process to advance too slow or even not train at all. A simplified example of how this parameter could affect the training is depicted in Fig. 4.17. Suppose that the landscape of the problem is the line draw, and through gradient descent, we want to minimize (reach the bottom) the function. If the learning rate is too small, the training process will eventually reach the bottom point but at a time cost. Setting the learning rate at an optimal value will get it to the bottom point but in less time. However, setting the learning rate too large, will cause the training process to oscillate around or even diverge from the minimum point. When the learning rate is too large, the updates to the model's parameters are so substantial that they overshoot the minimum point, causing the training process to bounce back and forth without ever settling down.

Several techniques can be used to alleviate this issue. Techniques like learning rate schedules or adaptive learning rate methods (e.g., Adam, RMSprop) that adjust the learning rate dynamically during training can help to prevent the worsts scenarios.

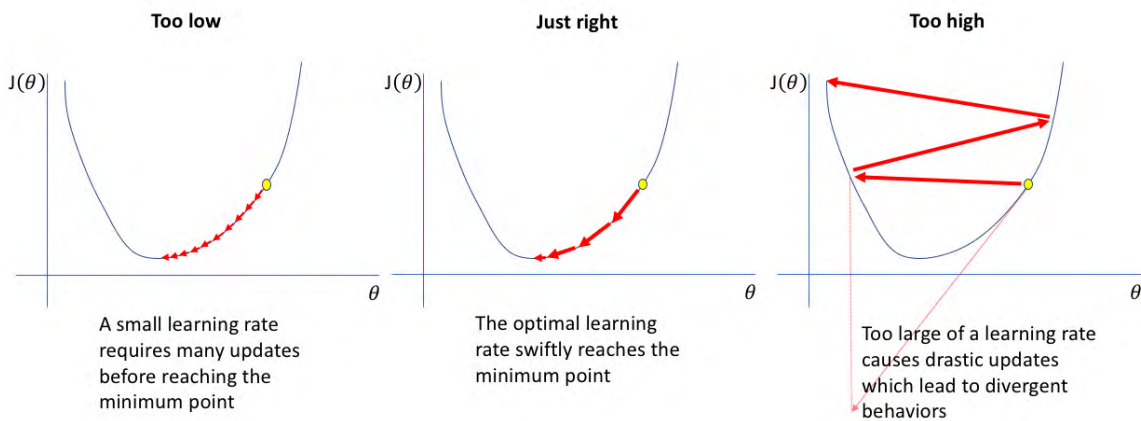


Figure 4.17: Learning rate examples and its outcomes [111].

Depending on the sequence length or the number of layers of the network, one problem could appear, called *vanishing gradient*. This problem is common to all deep learning architectures and is related to the way these networks are trained. In backpropagation, during each training iteration, the weights of the neural network are updated in proportion to the partial derivative of the error function with respect to the current weight. However, as the sequence length grows, or we go deeper in the architecture, the gradient magnitude is likely to either diminish or increase uncontrollably (*exploding gradient*), which can decelerate the training process, even stop it at all.

Several solutions have been proposed to mitigate this problem:

- **Activation Functions.** Select activation functions like ReLu can help alleviate the vanishing gradient problem by allowing gradients to flow through the network when the output is positive, while Tanh and Sigmoid can exacerbate the problem.
- **Weight Initialization.** In the first iteration, the weights of the network are normally random numbers. But they can be select in a way that ensures that the weights are initialized maintaining the variance of gradients throughout the network layers mitigating the problem.
- **Normalization Techniques.** Normalize the input on each layer stabilize and accelerate the training process, making it less sensitive to the initial weights.
- **Architectural Changes.** The use of *residuals* (shortcut connections between networks) allows gradients to flow more easily through the network.

4.4 Model Evaluation

Various metrics are used to evaluate the performance of a model. For a binary problem, two of the most important validation metrics are the Confusion Matrix and the [Area Under the Receiver Operating Characteristic Curve \(AUC\)](#). These metrics provide valuable insights into the model performance and help in assessing its effectiveness in classification tasks.

4.4.1 Confusion Matrix

The Confusion Matrix is a tabular representation of the performance of a classification model. It provides a comprehensive view of the model's predictions compared to the actual outcomes. For a binary classification problem, the confusion matrix is typically a 2x2 table with the following components:

- **True Positives (TP):** Correctly predicted positive instances
- **True Negatives (TN):** Correctly predicted negative instances
- **False Positives (FP):** Incorrectly predicted positive instances (Type I error)
- **False Negatives (FN):** Incorrectly predicted negative instances (Type II error)

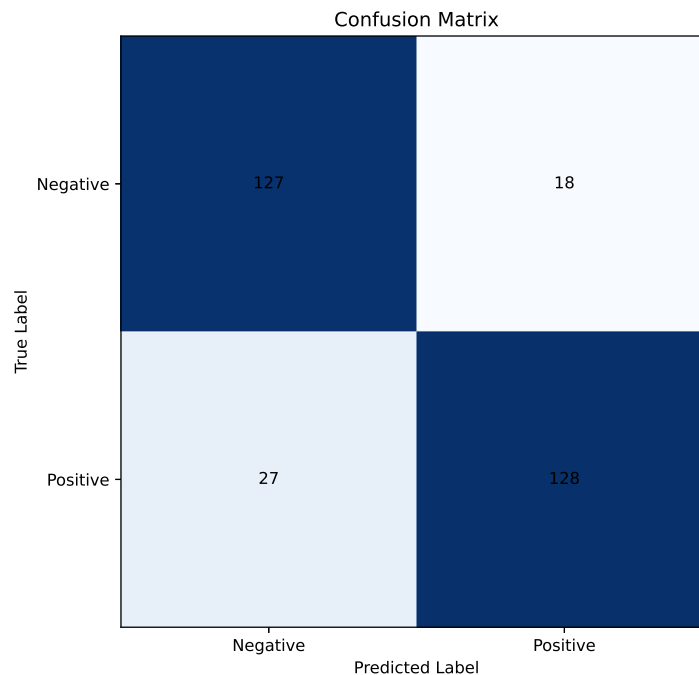


Figure 4.18: Example of a confusion matrix for a binary example

The confusion matrix allows for the calculation of various performance metrics, including:

- Accuracy: $(TP + TN) / (TP + TN + FP + FN)$

It represents the overall correctness of the model. However, accuracy can be misleading in cases of imbalanced datasets, where one class significantly outnumbers the other.

- Precision: $TP / (TP + FP)$

It answers the question: "Of all the instances the model labeled as positive, how many were actually positive?" Precision is particularly important when the cost of false positives is high.

- Recall (Sensitivity): $TP / (TP + FN)$

It answers the question: "Of all the actual positive instances, how many did the model correctly identify?" Recall is crucial when the cost of false negatives is high.

- Specificity: $TN / (TN + FP)$

It measures the model's ability to correctly identify negative instances.

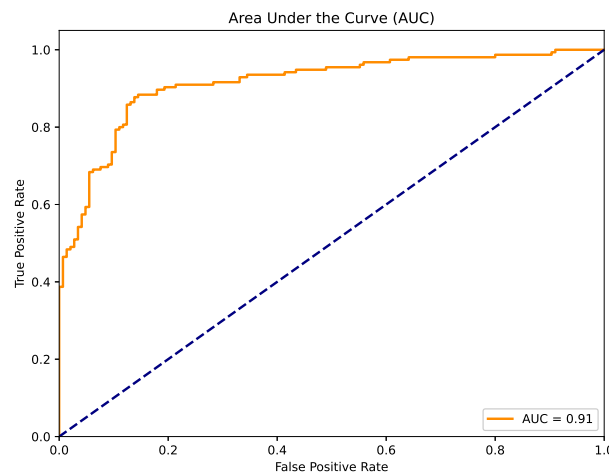
- F1 Score: $2 * (Precision * Recall) / (Precision + Recall)$

It provides a single score that balances both precision and recall. The F1 score is particularly useful when you have an uneven class distribution.

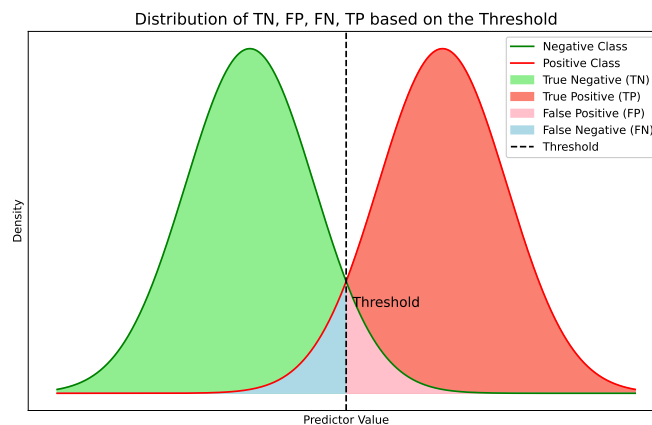
These metrics provide a comprehensive understanding of the model performance across different aspects of classification. Usually, there is always a tradeoff between several of these metrics. For example, in the medical field, it is common to encounter problems where increased sensitivity leads to decreased specificity and vice versa, being necessary to choose which of the two is more important to keep high.

4.4.2 Area Under ROC Curve

The Area Under the Receiver Operating Characteristic Curve, or simply AUC, is a performance metric used to evaluate the quality of a binary classifier predictions. The ROC curve is created by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) at various classification thresholds.



(a)



(b)

Figure 4.19: (a) AUC example. (b) Distribution of TN, FP, FN, TP based on the chosen threshold. Every point in the AUC represents the True Positive Rate against the False Positive Rate for a certain threshold.

Key points about the AUC:

1. The AUC ranges from 0 to 1, where 1 represents a perfect classifier and 0.5 represents a random classifier.
2. It provides a single scalar value that summarizes the classifier's performance across all possible thresholds.
3. AUC is threshold-invariant, making it useful for comparing different models or classifiers.
4. It is particularly valuable when dealing with imbalanced datasets, as it is insensitive to class distribution.

AUC provides a robust measure of model performance, especially when comparing different algorithms or tuning model parameters. Both the Confusion Matrix and the **AUC** are essential validation metrics in evaluating classification models. While the Confusion Matrix offers a detailed breakdown of the model predictions, the **AUC** provides a single, threshold-independent measure of the model discriminative power. Together, these metrics offer a comprehensive assessment of a classifier performance. This is what guides the hyperparameters tuning of the model as well as the behavior in the final stage. In summary, The AUC indicates the probability that the model will correctly classify two random samples, assigning a higher classification score to the positive sample than to the negative sample.

4.5 Model Validation

In cases where the simple train validation test split may not give an estimate of the models performance due to limited data, several techniques have been developed in order help maximize the use of data and offer a more dependable assessment of how well the model generalizes.

When validating a model, various scenarios are possible: underfitting, overfitting and good fit. The main way to detect which one is, is plotting the loss versus epochs for both the training and validation sets, like in Fig. 4.20b.

- *Underfitting.* This occurs when the algorithm performs poorly both in the data it was trained and the validation data. In this case, the model is too simple to capture the underlying patterns in the data. Both the training and validation loss will be high, and the curves will be close together, showing little improvement as training progresses.
- *Overfitting.* In this case, the model learns the training data too well, including its noise and peculiarities. The training loss will be very low, but the validation loss will be higher. There will be a gap between training and validation loss curves.
- *Good Fit.* This is the ideal scenario where the model has learned the underlying patterns, and it is good at generalization. Both training and validation loss will be low, and the curves will be close together, showing consistent improvement and then stabilizing.

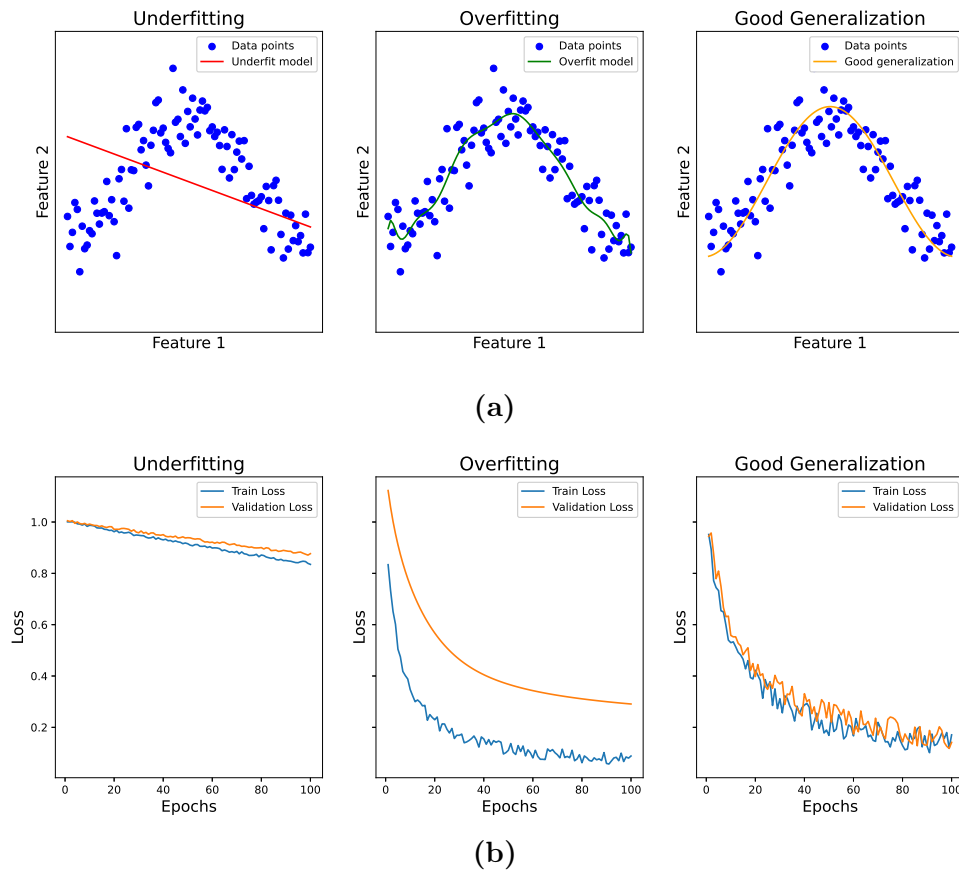


Figure 4.20: Fitting Validation. (a) Example of a model incurring in underfitting, overfitting and good fit. In *underfitting* the model is not able to make a good generalization and failed to predict samples even in the train set. In *overfitting*, the model adjusts too well to the train data, giving good predictions when using that data, but failing when trying to predict an outcome for a new sample point. When the model is *good in generalization*, its fit is ideal for predicting the train set within an admissible error, as well as for predicting new data points. (b) Example of curves showing train and validation loss for several scenarios: underfitting, overfitting and good fit.

Also, different fittings correspond to some explanation in bias/variance decomposition:

- *Bias*. It refers to the error introduced by approximating a real-world problem, which may be complex, by a simplified model. High bias indicates that the model makes strong assumptions about the data, potentially leading to underfitting. This results in poor performance on both the training and test datasets.
- *Variance*. It measures the model's sensitivity to fluctuations in the training data. High variance means that the model pays too much attention to the training data, capturing noise along with the underlying patterns, which can lead to overfitting. This happens because the model becomes overly complex, modeling the random noise in the training data rather than the intended outputs.

There is always a tradeoff between these metrics. Trying to adjust the balance between both is an iterative process: as model complexity increases, bias tends to decrease because the model can better fit the training data. However, variance tends to increase with complexity, as the model becomes more sensitive to the specific data points in the training set. Conversely, a simpler model may have higher bias but lower variance. The goal is to find a model complexity that minimizes the total error. This is summarized in Fig. 4.21.

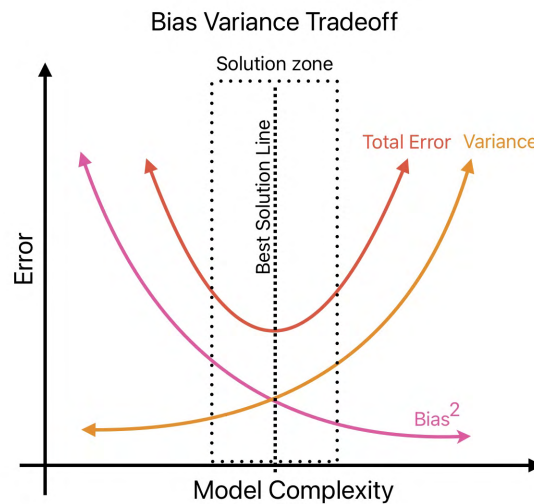


Figure 4.21: Bias/Variance tradeoff. When bias is high and variance is low, we are probably underfitting the data, whereas the variance is high but bias is low, we are probably overfitting the data. The goal is to find a model that sits between these two extremes, achieving a balance where the bias and variance are minimized enough to ensure the model generalizes well to new, unseen data, resulting in optimal performance. Source: [141]

4.5.1 Cross Validation

Cross-validation is possibly the most used resampling technique used to evaluate a model's performance and generalizability. It consists in by dividing the data into subsets, training the model on one subset (the training set) and validating it on another subset (the validation set). There are multiples ways of doing it but the most widely used is *K-Fold Cross Validation*. This method involves splitting the data into k sized subsets or *folds*. The model is validated k times, each time using a different fold as the validation set and the remaining folds for training. The final performance measure is an average of values computed during each iteration. Sometimes, due to an unbalance dataset, it is necessary that each fold maintain the same class distribution as the original, resulting in a *Stratified K-Fold*. An extreme example of cross validation is *Leave-One-Out Cross Validation*. This method involves using only one observation as the validation set, leaving the rest of the dataset for training.

K-fold and Leave-One-Out validation techniques provide insights into model bias and variance estimation. In Leave-One-Out each fold contains a single observation, minimizing bias but potentially introducing high variance in performance estimates. Cross-validation of k folds offers a computational compromise, balancing the trade-offs between bias and variance by partitioning the data into k approximately equal subsets. As k decreases, bias increases due to fewer training instances, while variance decreases. Thus, the choice of the number of folds (k) is crucial. While larger k reduces bias but may introduce variance, whereas a smaller k is less computationally intensive but could lead to a less accurate performance. Typically, 5 or 10 folds provide an optimal balance, providing robust performance estimates with reasonable computational complexity and statistically reliable variance characterization of the generalization capabilities of the predictive model.

Cross validation helps prevent overfitting by ensuring that each data point is included in the validation set once and in the training set k once. It provides a way to evaluate how well models perform, especially when dealing with limited datasets. However, cross validation can be demanding for datasets or complex models due, to the repeated training of models needed.

4.5.2 Permutation Test

The permutation test is a non-parametric statistical method used to assess the significance of a model's performance by comparing it to the performance expected under the null hypothesis that there is no association between the input features and the target variable. In the context of model validation, the permutation test serves as a resampling technique to estimate the generalization error and evaluate whether the observed model performance is better than what would be expected by chance.

Procedure of the Permutation Test:

1. Train the Model on Original Data:

- Train the model using the original dataset with true labels.
- Evaluate the model's performance on a validation set or using cross-validation to obtain the observed performance metric (AUC, accuracy, ...).

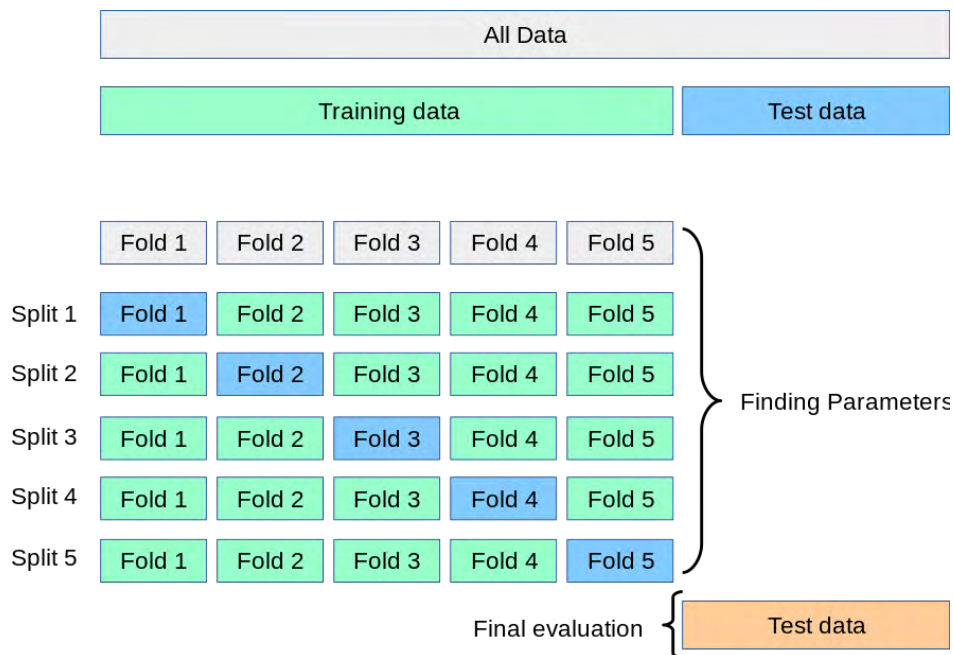


Figure 4.22: Illustration of k-fold cross-validation with $k=5$. The data is split into 5 folds, and the model is trained and validated 5 times, each time using a different fold as the validation set. Source: [180]

2. Permute the Labels:

- Randomly shuffle the target variable (labels) across the dataset while keeping the input features fixed. This breaks any existing associations between features and labels.

3. Train on Permuted Data:

- Train the same model architecture on the dataset with permuted labels.
- Evaluate the model's performance on the validation set or using cross-validation.

4. Repeat the Permutation:

- Repeat steps 2 and 3 multiple times (e.g., 1000 permutations) to build a distribution of performance metrics under the null hypothesis.

5. Compute p-value:

- Compare the observed performance from the original data to the distribution of performance metrics from the permuted data.
- The p-value is calculated as the proportion of permuted performance metrics that are equal to or better than the observed performance.
- A low p-value indicates that the observed performance is unlikely to have occurred by chance, suggesting that the model has captured meaningful patterns.

4.5.3 Permutation Test vs Cross Validation

Cross-validation and the permutation test are both essential techniques in model validation, each offering unique advantages. While k-fold cross-validation and leave-one-out (LOO) cross-validation provide estimates of a model's generalization error, they do not directly assess the statistical significance of the model's performance. In contrast, the permutation test quantifies the likelihood of achieving the observed performance under the null hypothesis, providing a p-value for statistical inference. This means that the permutation test can determine whether the model's performance is significantly better than what would be expected by chance.

Moreover, the permutation test is robust to *overfitting*. It can help detect overfitting by revealing if the model's performance is not significantly better than chance when the labels are permuted. This property makes the permutation test a valuable tool for identifying overfitting that might not be apparent through cross-validation alone.

Another advantage of the permutation test is that it makes no assumptions about the underlying data distribution. Being a non-parametric method, it does not rely on any specific data distribution, making it suitable for complex or unconventional datasets where traditional parametric assumptions do not hold.

The permutation test also provides insights into the *bias-variance trade-off* of the model. By comparing the model's performance to that under the null hypothesis, it offers an understanding of the bias component (systematic error) and helps in assessing the variance due to random chance in the data. This deeper insight into the model's error components can inform better model selection and tuning.

Also, visualization plays a crucial role in interpreting the results. A distribution plot obtained from the permutation test, such as a histogram or density plot, shows the distribution of performance metrics from the permuted data. Overlaid with a vertical line indicating the observed performance on the original data (see Fig. 4.23), this visual representation helps in understanding where the observed performance lies within the null distribution. It allows to visually assess the significance of the model's performance.

However, the permutation test can be computationally intensive, especially with a large number of permutations and complex models. In comparison, k-fold cross-validation may be less demanding computationally, depending on the number of folds and repetitions. Additionally, k-fold cross-validation ensures that each data point is used for both training and validation, maximizing data utilization. While the permutation test uses the entire dataset for each permutation, it may not provide as detailed an estimate of generalization error as cross-validation does.

In conclusion, the permutation test offers a valuable approach to model validation by directly testing the significance of a model's performance against the null hypothesis. It complements traditional methods like k-fold and LOO cross-validation by providing statistical inference about the model's ability to generalize. Rather than being two separate procedures, the permutation test and cross-validation are complementary. By combining both methods, we can gain deeper insights into the model's generalization capabilities, detect overfitting or underfitting, and make more informed decisions in model selection and evaluation.

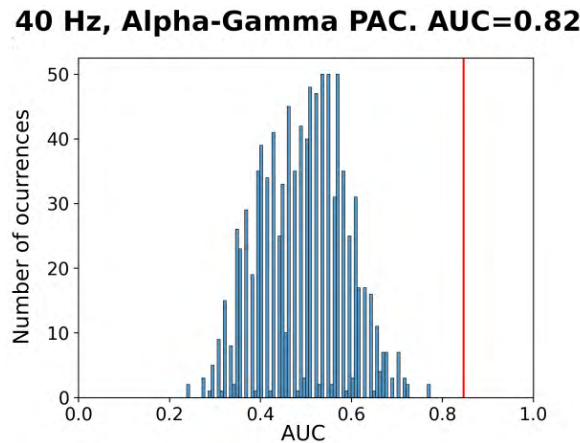


Figure 4.23: Permutation Test Distribution. The histogram represents the distribution of performance metrics (AUC) obtained from models trained on permuted data. The vertical red line indicates the observed performance from the model trained on the original data. A p-value is calculated as the proportion of permuted performance metrics equal to or exceeding the observed performance.

4.6 Explainability

Explainable Artificial Intelligence (XAI) refers to methods and techniques that make the decision-making processes of AI systems transparent and understandable to humans. As AI models, especially complex ones like deep neural networks, become integral to various sectors, ensuring their decisions are interpretable is crucial for trust, accountability, and compliance with ethical standards.

Understanding how AI models make decisions fosters trust among users and stakeholders. When the rationale behind AI outputs is clear, individuals are more likely to accept and rely on these systems. XAI helps identify and address biases within AI models, promoting fairness and equity. For instance, if a hiring algorithm disproportionately favors certain demographic groups, explainability techniques can uncover this bias and guide corrective measures. Many industries are subject to regulations that require transparency in automated decision-making. XAI ensures that AI systems comply with such legal standards, reducing the risk of penalties and enhancing accountability. By providing insights into model behavior, XAI enables data scientists to debug and refine AI systems, leading to enhanced performance and reliability.

Techniques in Explainable AI

- **Model-Specific Methods:** These are tailored to specific types of models. For example, decision trees are inherently interpretable due to their straightforward structure.
- **Model-Agnostic Methods:** Applicable across various models, these techniques treat the AI system as a black box and analyze input-output relationships to provide explanations. Tools like LIME (Local Interpretable Model-agnostic Explanations) and **SHapley Additive exPlanations (SHAP)** fall into this category.

- **Techniques for Deep Learning Models:**

- **Saliency Maps:** These highlight regions in input data (e.g., parts of an image) that significantly influence the model's output, aiding in understanding which features the model deems important.
- **Layer-wise Relevance Propagation (LRP):** This method backpropagates the prediction through the network to determine the contribution of each input feature to the final decision.
- **Attention Mechanisms:** In models like transformers, attention weights can be visualized to interpret which parts of the input data the model focuses on when making decisions.
- **Feature Visualization:** Techniques such as DeepDream generate images that strongly activate specific neurons, providing insights into the patterns recognized by the model.

SHapley Additive exPlanations (SHAP)

SHAP is a unified approach to interpreting machine learning models by assigning each feature an importance value for a particular prediction. It leverages concepts from cooperative game theory, specifically Shapley values, to provide consistent and locally accurate explanations for model outputs [182].

In cooperative game theory, Shapley value distributes the total gains to players based on their individual contributions. For a set of players $N = \{1, 2, \dots, n\}$ and a characteristic function $v : 2^N \rightarrow \mathbb{R}$ that assigns a value to each subset of players, the Shapley value ϕ_i for player i is defined as:

$$\phi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|! (n - |S| - 1)!}{n!} [v(S \cup \{i\}) - v(S)] \quad (4.40)$$

where:

- N is the set of all players.
- S is a subset of N not containing player i .
- $v(S)$ is the value or worth of coalition S .
- n is the total number of players in N .

This formula ensures a fair distribution of the total value by considering all possible coalitions.

In the context of machine learning, features are treated as players, and the model's prediction is the value to be distributed. The contribution of each feature to the prediction is assessed by computing the Shapley value, which considers the effect of including the feature in all possible subsets of features.

SHAP explanations satisfy several desirable properties:

- **Local Accuracy:** The sum of feature attributions equals the output of the model for a given input.
- **Missingness:** Features that are not present receive an attribution of zero.
- **Consistency:** If a model changes so that a feature's contribution increases or stays the same regardless of other features, the Shapley value for that feature should not decrease.

However, calculating exact Shapley values requires evaluating the model on all possible subsets of features, which is computationally infeasible for models with many features. To address this, SHAP employs approximation methods and leverages model-specific structures to efficiently estimate Shapley values. For instance, Tree SHAP provides a polynomial-time algorithm for tree-based models [130].

The SHAP values can be plotted for a better and simpler understanding. Several plots can be drawn, each one providing information about the problem in hands.

- **Summary Plot:** Displays the distribution of SHAP values for each feature across the dataset, indicating their overall impact.
- **Dependence Plot:** Shows the relationship between a feature's value and its SHAP value, highlighting interactions with other features.
- **Force Plot:** Visualizes how individual feature contributions combine to influence a specific prediction.



Figure 4.24: Example of SHAP summary plot.

Fig. 4.24 shows an example of a summary plot. It depicts how every variable affect the outcome of the model. In this example, the target value is the price of a house given certain features. Looking at the figure, we can observe that the primary feature

that drives the price of the house is the median income (MedInc) of the neighborhood, following by the latitude and longitude of the house. Is this capacity

SHAP has become a powerful tool widely used for model interpretability, elucidating how features influence individual predictions. Thus, it allows computing the feature importance identifying and ranking the most impactful features within a model along with fairness analysis, detecting and mitigating biases by examining feature attributions.

Challenges in Explainable AI

There is often a trade-off between a model's complexity and its interpretability. Complex models, such as deep neural networks, tend to offer higher accuracy because they can capture intricate patterns in the data. However, this complexity comes at the cost of being harder to explain. In contrast, simpler models like linear regression are more transparent and easier to interpret, but they might not perform as well in tasks that require modeling highly non-linear relationships.

The field of Explainable Artificial Intelligence (XAI) currently lacks standardized metrics to assess the degree of explainability. This makes it difficult to evaluate and compare different methods objectively. Without a consistent framework, determining the effectiveness of various explainability techniques becomes a significant challenge, hindering progress and the adoption of XAI solutions in critical applications.

Explanations provided by XAI methods must be tailored to the expertise of the audience. What might be clear and insightful to a data scientist could be confusing or overly technical for a layperson. This necessitates the development of adaptable explanation approaches that can effectively communicate insights to diverse stakeholders, ensuring the utility and accessibility of AI systems across different user groups.

In neuroscience, linking model outputs and explanations to real-time physiological processes in the human body demands a comprehensive understanding of both the model's internal workings and the biological systems it seeks to describe. Advanced models, such as deep neural networks, excel at capturing the complex and non-linear dynamics of neural processes, achieving high predictive accuracy. However, their inherent opacity poses challenges in connecting their outputs to specific physiological events. To address this, methods like incorporating descriptive layers such as Attention (see 8.4.2), leveraging SHAP values as previously defined, or utilizing custom solutions like the Lambert Projection (see 7.2) are invaluable. These approaches significantly enhance the ability to derive meaningful insights from complex models.

In summary, Explainable AI is essential for building trust, ensuring fairness, and complying with regulations in AI applications. While challenges exist, ongoing research and development in XAI aim to create AI systems that are both powerful and transparent.

Chapter 5

Database

To evaluate the potential of the proposed methodology, it is tested using an EEG database of individuals with [DD](#). Collecting neuropsychological data is crucial for investigating [Speech and Language Impairment \(SLI\)](#). The data, gathered through a rigorous protocol by the Leeduca Study Group at the University of Malaga, has established itself as a benchmark in researching and assessing reading and writing difficulties and other neurodevelopmental disorders.

The Leeduca Study Group has implemented a response to intervention system, used for two decades in the United States and ten years in Finland. This system proposes dynamic evaluations at three points throughout the year for large population samples, ranging from 4 to 8 years old. As a result, the Leeduca platform has become a robust resource for obtaining dynamic longitudinal evaluations of the subjects.

In addition to the data collected through this evaluation system, the platform has incorporated other indicators of relevance for the diagnosis and understanding of dyslexia. These include the ATLAS reports, a self-report questionnaire on reading and writing difficulties for adults, which offers information on the family risk of presenting these difficulties, as well as a detailed description of the conditions of the subjects at the age of 7 years, including standardized assessment tasks.

The Leeduca platform has provided a rich set of neuropsychological data and ensured a framework that adheres to recognized standards in researching dyslexia and other neurodevelopmental disorders. It has also incorporated a comprehensive and multidimensional view of the subject, considering their specific abilities, difficulties, and the environment in which they develop. The data collection procedure established by the Leeduca platform has been crucial in advancing our understanding of DD and its impact on child development. Consequently, this platform has become an essential instrument for researching and studying this disorder, with the ultimate goal of improving the detection, diagnosis, and intervention in dyslexia.

5.1 Study participants

The selection process for recruiting participants, including both those with dyslexia and the control group, involved choosing individuals from a group of 700 children. These participants were selected from 20 schools within the educational system of Junta de

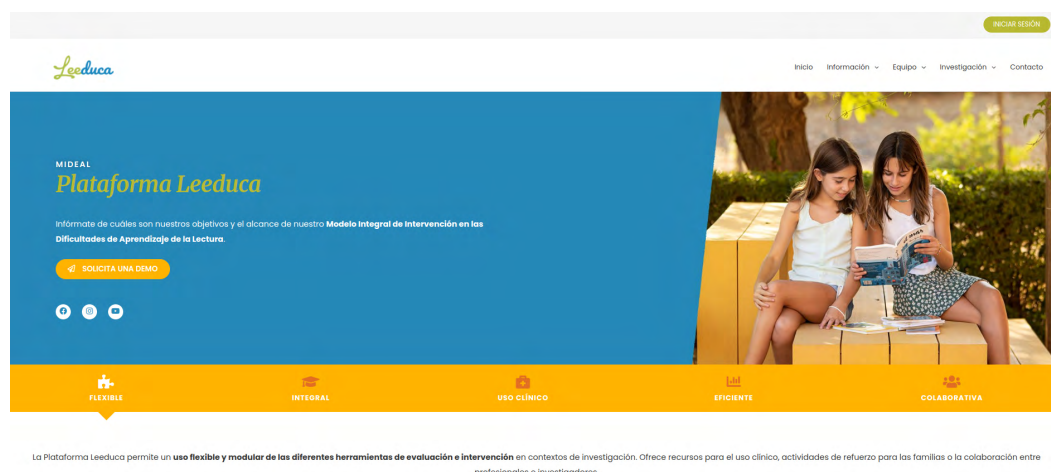


Figure 5.1: Screenshot of the Leeduca platform login page. <https://leeduca.es/>

Andalucía and were observed from the age of 4 until the second assessment at 7 years old. This long term study considered factors such as socioeconomic status, immigration background, bilingualism and other conditions like language impairment, speech sound disorder, attention deficit hyperactivity disorder and autism.

In the field of behavioral sciences, the selection and classification of experimental and control groups is essential for the establishment of comparative studies and the inference of results. The rigor of these procedures largely determines the validity and reliability of the conclusions obtained. To achieve an adequate pairing of the groups, use has been made of the longitudinal dynamic evaluation offered by the Leeduca Project, an instrument of great relevance in the field of reading and writing difficulties. This evaluation, carried out three times a year on large population samples, has allowed the identification of subjects with and without reading difficulties, which has favored the formation of experimental and control groups. The comprehensive evaluation provided by the Leeduca platform (see Fig. 5.1) has included not only aspects related to reading and writing, but also information on comorbidities and other conditions that could affect reading performance.

The formation of the experimental and control groups has been carried out based on the dynamic evaluation by the Leeduca project, the family risk information obtained through ATLAS and a complete report of the subjects at the age of 7 years. This has allowed a longitudinal view of the subjects, which has been revealed as an aspect of great value for research.

A total of 48 individuals took part in the research; 32 readers (including 17 males) and 16 individuals with dyslexia (including 7 males). The average age was 94.1 ± 3.3 months for the control group and 95.6 ± 2.9 months for the group.

Both groups are right-handed, speak Spanish as their first language, and have no hearing or vision problems to participate. In addition to these criteria, members of the dyslexia group have a formal diagnosis of dyslexia from their respective schools. The participants in the control group do not report any difficulties with reading or spelling and have not been formally diagnosed with dyslexia before. The diagnostic procedures



Figure 5.2: Laboratory view. actiCAP headset and NIRS acquisition systems

involved assessment tasks that were part of a detailed evaluation report completed when the children were 7 years old. These diagnostic criteria ensured reliable classification.

The research was carried out with approval from the Medical Ethical Committee of Málaga University (reference 16 2020 H). The study followed guidelines outlined in the World Medical Association Declaration of Helsinki. Legal guardians of the participating children provided consent, and were present during data collection sessions. It is important to note that this selection and classification process has been carried out in accordance with the provisions of the Declaration of Helsinki of the World Medical Association, thus ensuring compliance with ethical principles in research with human subjects.

5.2 EEG Recording Protocol

The EEG recording was done using a Brainvision actiCHamp Plus amplifier, equipped with a 32-channel amplifier that offers a sampling rate of up to 100 kHz. This advanced equipment enabled the collection of high-quality brain activity data, using active electrodes (actiCAP, Brain Products GmbH, Germany) which allow for a greater range of impedances, improving the signal-to-noise ratio and offering greater resistance to head movement. This feature is especially relevant when working with child populations, who may have difficulty keeping their heads still during recording sessions. The laboratory setup is shown in Fig. 5.2.

To ensure consistency and reproducibility, the electrodes were positioned according to the 10-20 system, a standardized montage that allows for consistent electrode positioning across subjects and studies. The Cz electrode, located at the vertex of the scalp, served as the reference point for all other electrodes.

The equipment was powered with lithium-ion batteries to ensure isolation from the power line and reduce acquisition noise. During the recording sessions, participants were exposed to auditory stimuli comprising white noise modulated at three different

frequencies: 4.8 Hz, 16 Hz, and 40 Hz. These frequency choices aimed to explore aspects of auditory processing and potential resonance effects in the brain. The stimuli were presented through headphones in organized sessions lasting 2.5 minutes each, with a total recording time per participant of up to 15 minutes. The sequence of stimulus presentation was carefully planned to balance out any order effects: first 4.8 Hz, then 16 Hz, followed by 40 Hz, and then a repeat of 40 Hz, 16 Hz, and back to 4.8 Hz.

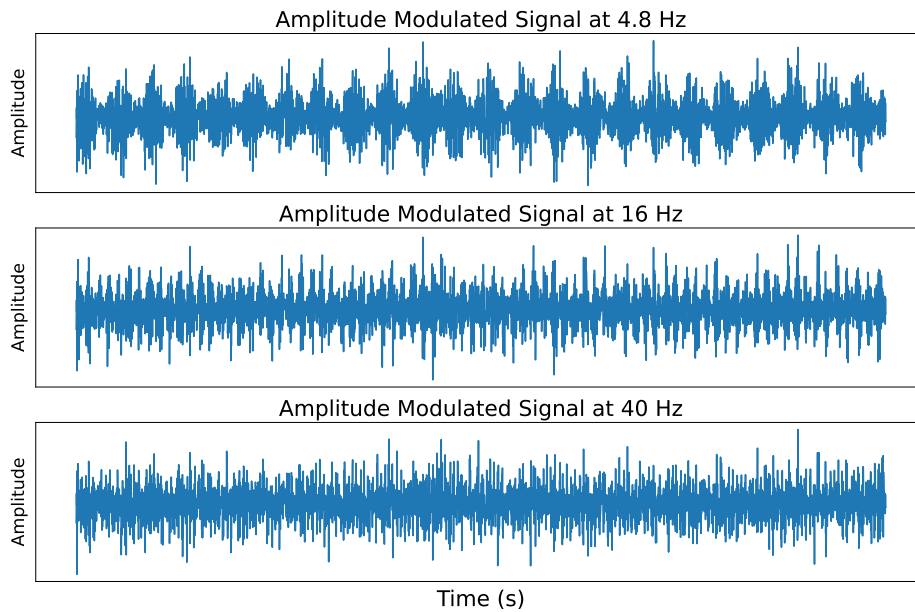


Figure 5.3: White noise modulated at different frequencies as used in the experiment

During the recording sessions, participants were instructed to remain still and concentrate on the sounds they heard to reduce any movements that could interfere with the data collection and ensure high-quality data.

The acquired EEG data were digitized at a sampling rate of 500 Hz, striking a balance between capturing details and managing data efficiently. This sampling rate is adequate for capturing frequency ranges in cognitive and sensory studies.

5.3 EEG Preprocessing

The EEG preprocessing pipeline is already described in 3.3. It includes essential steps for refining raw data, ensuring consistency, and preparing it for in-depth analysis. Each step in the pipeline addresses specific types of artifacts or unwanted noise and applies filters to isolate relevant signals. Below is a summary of the pipeline steps, along with specific measurement parameters applied at each stage to optimize the data quality for further analysis.

Table 5.1: EEG Preprocessing Pipeline

Step	Specific Measurements
Referencing	Central electrode (Cz)
Bipolar Channel Creation	VEOG (difference top - bottom), HEOG (left - right)
Notch Filtering	60 Hz
High-Pass Filtering	0.5 Hz
Low-Pass Filtering	120 Hz
Slope	24dB/oct
Artifact Rejection	Outside -200 to 200 μV
Segmentation	-100 ms to 1000 ms (overlapping)
Baseline Correction	-100 ms to 0 ms

5.4 Stimuli and Language Processing

5.4.1 Relevant EEG Bands in language disorders

The role of brain oscillations in language processing has been a source of intense study and controversy in dyslexia research. Up to five types of oscillations have been identified that could be relevant to explain DD and other language disorders [108]. However, in this study, we have focused on Delta, Theta and Beta oscillations, as they have been shown to be especially relevant in previous studies[45, 46]

Dyslexia, a reading disorder frequently characterized by reduced awareness of speech units, is among the most persistent learning difficulties faced by children [144]. This specific learning disorder, of neurological origin, generates difficulties with accuracy and fluency in word recognition, and problems in spelling and decoding. By better understanding the neural oscillations that have been identified as relevant for DD and other language disorders, we can advance in the creation of more effective interventions for those affected by these conditions.

Delta oscillations (1-4 Hz; around 2 Hz), located in the right hemisphere, allow for processing of the prosodic aspects of speech, that is, the rhythm and cadence of speech. Theta oscillations (4-7 Hz; around 5 Hz), preferentially shown in the right hemisphere, allow for the segmentation of syllables. In the Beta frequency band (around 20 Hz), the internal segmentation of the syllable occurs, that is, the distinction between the initial segment (attack) and its rhyme (for example: s-ol, pl-an, etc.).

Other models of speech coding suggest that dyslexia originates from the unusual neuronal dominance in the right hemisphere of slow-rhythmic prosodic oscillations (Delta band, 0.5-1 Hz), syllabic (Theta band, 4- Hz) or phoneme (Beta-Gamma band) modulations of speech, which are defined by the rise time in amplitude (i.e., the envelope) generated by the rhythm of speech [48, 57].

5.4.1.1 Creation and use of stimuli for the evaluation of language processing and Dyslexia

In this study, white noise amplitude modulated at fixed rates was used as auditory stimulus to activate populations of neurons responsible for the segmentation of different language units. The stimuli used were obtained by rhythmically modulating the amplitude of white noise sound in the Delta, Theta or Gamma bands. This methodology, far from the works that use stimuli based on speech or word processing that require subject interaction during the experiment, allowed us to generate different auditory stimuli aimed at activating the functional brain networks involved in language processing at different levels [61]. It is expected that these neurophysiological responses will explain the manifestations of the temporal processing deficits described in DD. Within the framework of our study on dyslexia reading disorder, we have directed special attention towards the creation and use of stimuli that allow us to evaluate more precisely and in detail the processing of language. This process is critical since dyslexia is characterized, among other things, by a difficulty in processing and segmenting speech units, as has been detailed in previous studies [144].

For the generation of these stimuli, we resorted to the modulation of the stimulus at the frequencies of production of words, syllables and phonemes in the Spanish language. The choice of these frequency bands is not arbitrary, but responds to previous findings that identify Delta, Theta and Beta oscillations as particularly relevant to explain dyslexia and other language disorders [108]. This type of stimuli, due to their specificity, allows us to selectively activate the brain networks involved in language processing [61]. It is important to highlight that this approach departs from those works that use stimuli based on speech or word processing that require subject interaction during the experiment. On the contrary, by only requiring the subject to listen to the stimulus in a state similar to rest, we minimize the interactions that could interfere with the precise evaluation of the neurophysiological responses we seek to measure.

By focusing on language processing and its relationship with dyslexia, we cannot lose sight of the fact that language processing, in general, is a complex task that involves the interaction between different brain areas in charge of auditory, visual and linguistic processing [96]. Thus, DD, although characterized by a particular set of symptoms and manifestations, must be understood as part of a broader system of cognitive and neurological functions. This perspective is what has led us to use an experimental protocol that, through the use of specific stimuli and the minimization of unnecessary interactions, aims to shed light on the oscillatory neural processes in dyslexia. Through the observation of these responses, we hope to contribute to the understanding of temporal processing deficits in dyslexia, with a view to developing more effective interventions for those affected by this condition.

5.4.1.2 Evaluation Methodology

Within the evaluation methodology used in our study, Auditory steady-state response (ASSR) play a fundamental role in investigating oscillatory neural processes in the construction of predictions of incoming auditory information, particularly in the context of dyslexia. ASSR is an evoked auditory response that occurs when an auditory stimulus

modulated in amplitude or frequency is presented at a constant rate. Because ASSR can be recorded while the individual is at rest or even asleep, they offer a significant advantage for audiological evaluation of individuals who cannot actively participate in traditional audiometric testing. The decision to use amplitude modulated white noise lies in the ability of this type of stimuli to generate a selective and precise activation of the brain networks related to language processing [46]

This method of stimulus generation differs from approaches that use stimuli based on speech or word processing, which require active interaction of the subject during the experiment. On the contrary, our approach allows the subject to simply listen to the stimulus in a state similar to rest, which minimizes possible interference and allows us to focus on the neurophysiological responses to incoming auditory information. The central objective of this methodology is to evaluate the quality of the oscillatory neural processes measured through ASSR. These processes are essential for the construction of predictions of incoming auditory information, such as linguistic sequences or their simplification. In the context of DD, these types of predictions may be affected, which could explain some of the temporal processing deficits described in this disorder.

Chapter 6

Phase-synchrony Connectivity Anomalies in EEG signals

Variability and unbalanced datasets are usual problems in biomedicine, that pose significant challenges for statistical and machine learning methods. Subject data often exhibits high variability due to differences in demographics, genetics, or environmental factors, making it difficult to build generalizable models. Additionally, unbalanced datasets, where rare conditions are underrepresented, can skew results and lead to biased predictions. These issues compromise model performance and diagnostic accuracy. Anomaly detection methods offer a solution by identifying patterns or instances that deviate from the normal behavior. This way, it is possible to handle these problems by generating a one-class model (corresponding to the majority class) that can highlight underrepresented classes, improve the robustness and provide insights into rare but clinically significant conditions.

This chapter introduces the detection of anomalies in EEG signals related to phase-synchrony connectivity. Functional connectivity refers to the synchronized activation of different brain regions during specific tasks or resting states. Moreover, the Pearson's Circular Correlation is used as a method for estimating phase synchrony between EEG channels. The proposed method accounts for the circular nature of phase data and mitigates the effects of volume conduction.

For outlier detection in small and imbalanced datasets, we propose a method based on the [Self-organizing map \(SOM\)](#). SOM is trained using data from control subjects to create a reference map of typical EEG connectivity patterns. The quantization error (Q_e) between the SOM prototypes and new data samples is used as a feature to identify anomalies. Dyslexic subjects are expected to produce higher Q_e due to deviations from the normal patterns. A Bayesian anomaly detection approach is then applied, utilizing the Q_e distributions to classify subjects as controls or dyslexic. The proposed methodology allowed identifying synchronization anomalies in dyslexic subjects, particularly in phase synchronization between EEG channels. Notable differences were observed in connections involving occipital and frontal electrodes, which are associated with visual processing and language networks.

6.1 Functional Connectivity from EEG signals

As explained in Chapter 3, functional connectivity pertains to the synchronized activation of brain areas during specific tasks or resting states. Several studies have developed methods to estimate functional connectivity by identifying patterns associated with neurological diseases such as Alzheimer's or Parkinson's Diseases, using functional imaging techniques [14, 149, 154, 164], EEG [25, 87], or for classifying sleep stages [98]. EEG-based functional connectivity research has potential due to EEG's high temporal resolution, which measures synchronization between brain areas. This synchronization, reflecting cooperation in functional connectivity, can be inferred from phase differences between EEG channels, statistical relationships across EEG frequency bands (Cross-Frequency Coupling), or techniques identifying inter-channel causality [34].

In this context, previous studies have used EEG and Magnetoencephalography (MEG) to explore DD, assuming it is characterized by reduced speech unit awareness [144]. Atypical neural entrainment at various rates may occur in affected individuals. Some studies investigating speech coding foundations suggest DD stems from atypical right hemisphere synchronization in lower EEG bands [48, 57]. In [165], differences were found in the Delta band entrainment phase between dyslexic and control groups after presenting visual and auditory stimuli to children.

Examples of EEG-based classification include work by Gibbon et al. [72], where 8-week-old infants exposed to drumbeats or syllables showed distinguishable EEG patterns, achieving AUCs of 0.875 and 0.86 for convolutional neural networks and support vector machines, respectively. Perera et al. [160] recorded EEGs during writing tasks to detect dyslexic children using a support vector machine, achieving varied classification results across different brain areas. Ortiz et al. [155] used stimuli at various frequencies to differentiate EEG signals, yielding AUCs between 0.69 and 0.89. Martinez et al. [137] used periodogram-based features and Principal Component Analysis, obtaining AUC values up to 0.75. Another study [139] utilized power-based connectivity features for training a neural network to represent dyslexic and control groups in a reduced dimensional space.

Feature extraction must yield discriminative descriptors for modeling dataset categories, using time, frequency [172], or time-frequency [13] features. These features can denote *functional connectivity* through frequency and phase variations and power spectrum. Connectivity measures include:

- Statistical relationships between electrodes within the same frequency band, such as **Spectral Coherence (SC)**, indicating *functional connections* between neuron clusters [198], or phase angle differences between channels [34].
- Statistical relationships between activities in two channels at different frequency bands, referred to as **CFC**. Phase-Amplitude Coupling exemplifies CFC with neural and physical implications [148].

In our study, we use **SC** descriptors to estimate connectivity between channels across five EEG bands (Delta, Theta, Alpha, Beta, Gamma) for differential DD diagnosis. Then, an anomaly detection approach combines unsupervised learning via

vector quantization and a Bayesian classifier, addressing small databases, imbalances, and overfitting issues.

The work has two main objectives: firstly, to demonstrate that low-level auditory processing produces distinct patterns in brain networks, studied through EEG phase synchronization. This methodology advances diagnostic capabilities and understanding of auditory processing in dyslexic vs. control children, providing insights into dyslexia's biological basis. Secondly, the classification pipeline offers an objective method for detecting dyslexic subjects using EEG signals, serving as a valuable clinical tool. The main contributions are:

1. Using low-level auditory stimuli to investigate brain processes involved in language processing, unlike previous studies relying on speech-based stimuli [144].
2. Searching for brain area connectivity through phase synchronization between EEG channels, computed using Circular Correlation.
3. Implementing an anomaly detection approach with a combination of unsupervised learning by vector quantization and a Bayesian classifier, effective for small databases, imbalanced data, and reducing overfitting.

6.2 Pearson's Circular Correlation

One of the main problems using some of the connectivity metrics already seen in 3.4.1 is the volume conduction problem, leading to several electrodes capturing the same underlying brain activity and resulting in spurious connectivity estimates. To address this challenge, we employed Circular Correlation to estimate phase synchrony. However, as highlighted by [117], coherence measures have certain drawbacks. Firstly, coherence is a measure of linear covariance between two spectra and is not suitable for non-stationary signals. In contrast, phase-locking based measures do not impose stationarity requirements on the signals. Secondly, coherence is influenced by amplitude covariance, which may introduce bias in the estimation of interaction between channels.

Burgess et al. [17] explains that Circular Correlation assesses the covariance of phase variance between two channels. It determines whether the phase of one channel is slightly advanced relative to its expected phase at a given time point, and if so, whether the phase of the other channel is also advanced (referred to as positive correlation in this example). When the channels are unrelated, their phase variances will not co-vary, resulting in a Circular Correlation of zero.

As phase information is circular (ranging from $[0, 2\pi]$), circular statistics are employed [177]. Specifically, the circular equivalent of Pearson's correlation coefficient [106] is defined as:

$$r_{circular} = \frac{\sum_{t=1}^n [\sin(\phi_{xt} - \bar{\phi}_x) \sin(\phi_{yt} - \bar{\phi}_y)]}{\sqrt{\sum_{t=1}^n \sin(\phi_{xt} - \bar{\phi}_x)^2} \sqrt{\sum_{t=1}^n \sin(\phi_{yt} - \bar{\phi}_y)^2}} \quad (6.1)$$

where ϕ_{xt} , ϕ_{yt} is the phase value of the channels x and y at time point t , $\overline{\phi_x}$, $\overline{\phi_y}$ the circular mean of x and y channels, and n the number of samples of the signal. The circular mean of a channel i can be computed as:

$$\overline{\phi_i} = \arctan \left(\frac{\sum_{t=1}^n \sin(\phi_{it})}{\sum_{t=1}^n \cos(\phi_{it})} \right) \quad (6.2)$$

According to the above definitions, $r_{circular}$ measures the circular covariance of differences between the observed phase and the expected ($\overline{\phi_i}$) phase. As a result, relationship (i.e., synchronization) between channels is detected as a co-variation between the phase variance of the channels. On the other hand, out-of-sync channels will not co-vary and the circular covariance will take values close to zero. In the case of Pearson's correlation, $r_{circular}$ will take values in the range $[0,1]$.

6.3 Outlier Detection

When the data is very unbalanced, like in the case of medical diagnosis where the vast majority of patients are healthy and only a small fraction have a specific condition, it can be challenging to build accurate predictive models. In such scenarios, outlier detection techniques can be particularly useful for identifying rare cases that deviate significantly from the norm.

One approach to outlier detection in this context is to use unsupervised learning algorithms, such as clustering or density-based methods. These techniques can help identify data points that are far removed from the main clusters or have a low density of neighboring points. For example, if we have data on reading speed and accuracy for a large population, we might expect most individuals to fall within a certain range. Those who have significantly lower scores on these measures could potentially be flagged as outliers and considered for further evaluation. This approach can be done with what it is called [Self-organizing map \(SOM\)](#).

6.3.1 Self-Organizing Maps

[SOMs](#) are a type of artificial neural network that can learn to map high-dimensional data onto a lower-dimensional space, typically a two-dimensional grid. They preserve the topological relationships between data points while reducing the dimensionality of the feature space. The algorithm works by iteratively adjusting the weights of the neurons in the map to minimize the distance between the input data and the corresponding neurons, effectively organizing the data into clusters based on their similarity. This ability, to self-organize and map high-dimensional data onto a lower-dimensional space, makes them a valuable tool for data visualization, clustering, and anomaly detection across a wide range of applications. By preserving the topological relationships between data points, SOMs can provide insights into the underlying structure of complex datasets, facilitating the identification of patterns and clusters that may not be readily apparent in the original high-dimensional space.

SOMs have been applied in various fields, including medicine, computer vision, fault prediction, and network security. In medicine, Golub et al. [77] used SOM to group molecular leukaemia samples using genes as features, successfully clustering two types of cancer cells. In the computer vision field, Lawrence et al. [120] applied SOM to reduce the dimensionality of an image as the first step in a face recognition pipeline. Betti et al. [11] designed a fault prediction system for large power plants based on SOM, while Cai et al. [20] employed an improved SOM algorithm to classify different sitting postures. Vesanto and Alhoniemi [204] conducted more general experiments to demonstrate the effectiveness of using SOM for dimensionality reduction and clustering, showing that training a SOM before attempting to cluster the data is beneficial. In the field of network security, De La Hoz et al. [44] developed a SOM-based method that uses Bayesian activation of SOM units to detect anomalies in computer networks.

More in detail, the process of train a SOM is at follows:

Let $X \in \mathbb{R}^n$ be an n -dimensional dataset consisting of m data instances, i.e., $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$, where each $\mathbf{x}_i$ is an n -dimensional vector. The SOM map is composed of d units, arranged in a 2-dimensional lattice and each represented by an n -dimensional model vector $\boldsymbol{\omega}_i$, where $i = 1, 2, \dots, d$.

1. Initialization: Initialize the model vectors $\boldsymbol{\omega}_i$ randomly or using a specific initialization technique, such as linear initialization based on the principal components of the training data.

2. Competitive Learning: For each input data instance $\mathbf{x}_k$, find the **Best Matching Unit (BMU)** among the model vectors. The **BMU** is defined as the unit $\boldsymbol{\omega}_i$ that is closest to $\mathbf{x}_k$ in terms of the Euclidean distance $\|\cdot\|$:

$$\|\boldsymbol{\omega}_i - \mathbf{x}_k\| \leq \|\boldsymbol{\omega}_j - \mathbf{x}_k\|, \forall \mathbf{x}_k \in X, i \neq j \quad (6.3)$$

3. Cooperative Learning: Update the model vectors in a neighborhood of the BMU according to the Euclidean distance to the input data sample. The update rule for each model vector $\boldsymbol{\omega}_i$ is given by:

$$\boldsymbol{\omega}_i(t+1) = \boldsymbol{\omega}_i(t) + \alpha(t)h_i(t)(\mathbf{x}_k - \boldsymbol{\omega}_i(t)) \quad (6.4)$$

where t represents the current iteration, $\alpha(t)$ is the learning rate at iteration t , and $h_i(t)$ is a neighborhood function that defines the influence of the BMU on the neighboring units. The learning rate $\alpha(t)$ typically decreases over time, often following an exponential decay rule. The neighborhood function $h_i(t)$ is usually a Gaussian function centered on the BMU, and its width shrinks as the iterations progress.

4. Iteration: Repeat steps 2 and 3 for a predefined number of iterations or until convergence.

6.3.2 Band relevance using quantization error distribution

After the **SOM** training is completed with controls data, each input data instance $\mathbf{x}_k$ is represented by the model vector of its corresponding BMU. The representation error, known as **Quantization Error (Qe)**, can be computed for each BMU as:

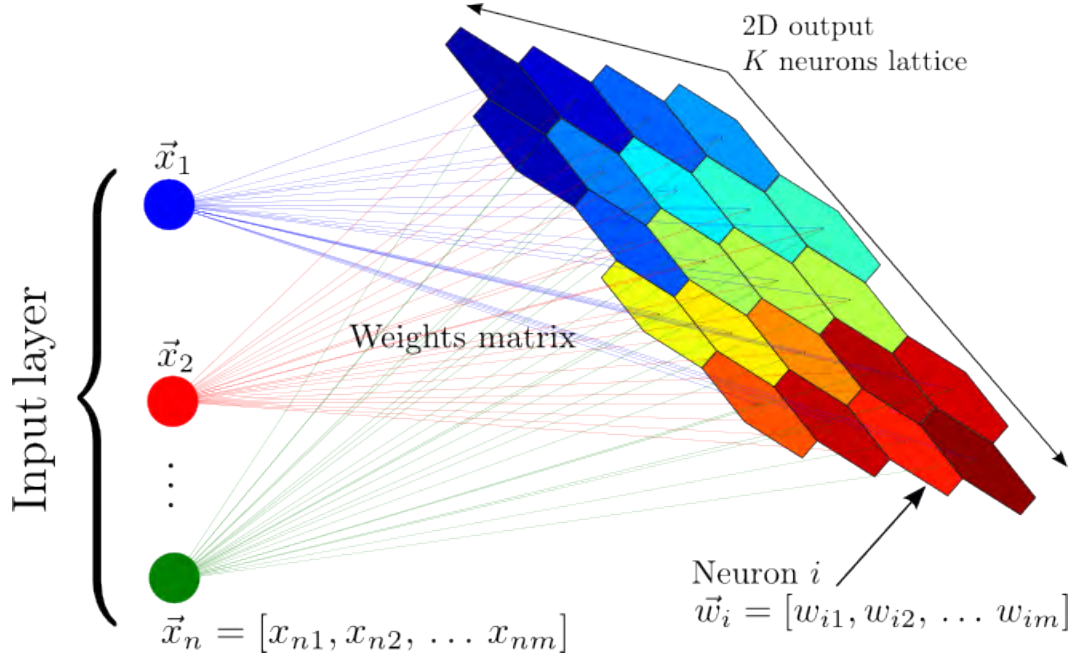


Figure 6.1: SOM showing different classes. Each input data instance is represented by the model vector of its corresponding BMU [210].

$$qe = \frac{1}{n} \sum_{k=1}^n \|\omega_i - \mathbf{x}_k\| \quad \forall \mathbf{x}_k \in RF_i \quad (6.5)$$

where RF_i is the receptive field of the BMU ω_i , defined as:

$$RF_i = \{\mathbf{x}_k \in X : \|\mathbf{x}_k - \omega_i\| \leq \|\mathbf{x}_k - \omega_j\| \forall i \neq j\} \quad (6.6)$$

After computing the Qe for every BMU some interesting insights can be gained. If we show the distribution of this errors like in Fig. 6.2 for control and dyslexia subjects, we can clearly see that there are bands that can separate both groups, making it possible to find a threshold in Qe to separate Qe values provided by samples of different classes. Even more, we can choose the most discriminant band this way. Having two distributions, we can compute the dissimilarity grade between them using what is called the Kullback-Leibler divergence (D_{KL}) [116], defined as:

$$D_{KL}(P|Q) = \sum_i P(i) \frac{P(i)}{Q(i)} \quad (6.7)$$

This metric is used to quantify the overlapping between P and Q distributions and eventually, the discriminative capability of the corresponding error quantization for each data sample. Although D_{KL} is not a symmetric measure and cannot be interpreted as a distance, it is possible to symmetrize by computing a two-ways measure as $\hat{D}_{KL} = 0.5 * (D_{KL}(P|Q) + D_{KL}(Q|P))$.

According to this measure, $KL_{gamma} > KL_{beta} > KL_{theta} > KL_{alpha} > KL_{delta}$ for 4.8 Hz stimulus and $KL_{alpha} > KL_{beta} > KL_{theta} > KL_{gamma} > KL_{delta}$ for 16 HZ

stimulus.

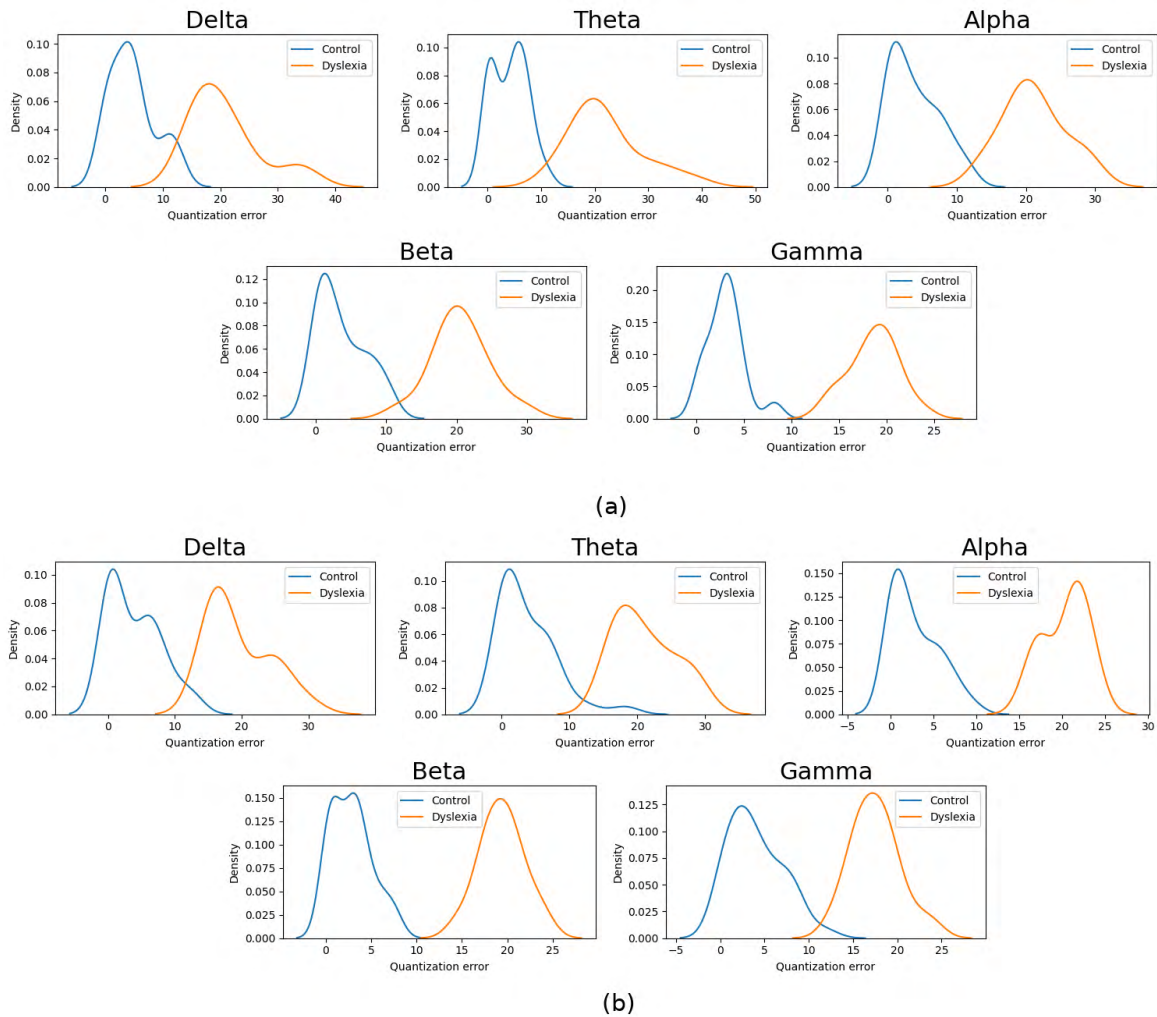


Figure 6.2: Quantization error distribution for Controls and Dyslexia subjects. The (a) 4.8 Hz and (b) 16 Hz stimulus.

From the perspective of the **BMU**, the **Qe** provides valuable insights into how well each data sample is represented by its corresponding **BMU**. Regardless of the similarity between the sample and the map prototypes, every data point is assigned to a **BMU**. However, when a particular sample exhibits a high **Qe**, it indicates that the **BMU** is struggling to accurately represent that sample, suggesting that it may be an outlier or an anomaly in the dataset.

In our case, the model was trained solely on controls data, thus a higher **Qe** for the dyslexic samples is expected if the difference between the two groups are enough. In Fig. 6.3 the activation of **SOM** units for all the available samples is shown, according to the corresponding mean quantization error. The figure shows the activations of the **SOM** units along with the corresponding **Qe** for each one and for all the available samples, being higher for the dyslexic meaning that are differences between the two groups.

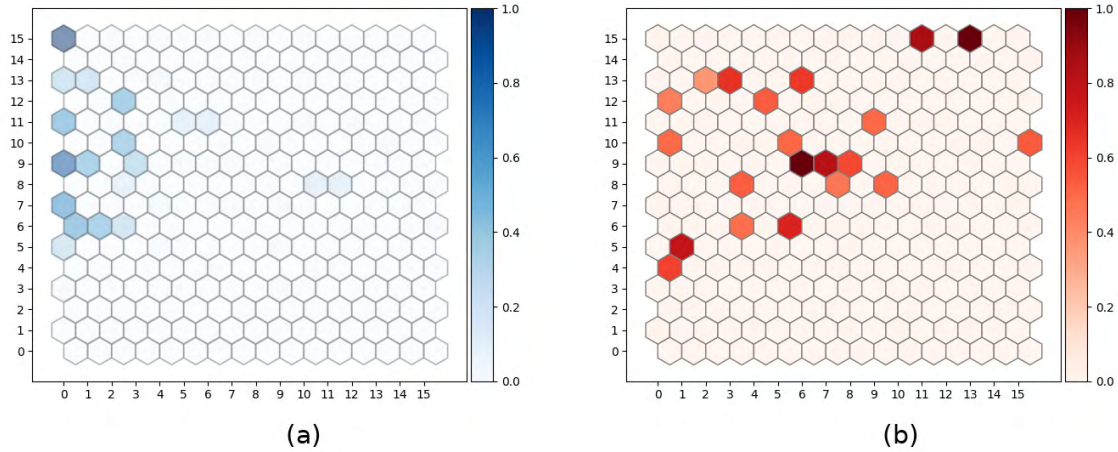


Figure 6.3: SOM map activation for each data sample, according to the corresponding mean quantization error. Example for EEG Delta band using 4.8 Hz stimulus. (a) Control (b) Dyslexia.

6.3.3 Bayesian Anomaly Detection for SOM

Naïve Bayes classifiers [110] utilize Bayes' Theorem to separate data into distinct classes, assuming independence among predictors. Although this assumption of independence is not always strictly met, studies have shown that Naïve Bayes can still outperform other classifiers, even when features are not entirely independent [12, 40, 184]. In fact, several works have demonstrated the benefits of Naïve Bayes classifiers for EEG signal classification. For instance, [40] shows that increasing the number of principal components (PC) used to project the EEG data does not reduce the classification accuracy or AUC value, compared to using only one PC, where feature independence is inherently assumed.

Moreover, Naïve Bayes classifiers offer good performance with relatively small datasets, efficiently avoiding issues related to high dimensionality and overfitting. This is particularly important when working with limited data, where supervised classifiers tend to overfit, reducing their generalization capabilities [53].

As discussed in the previous section, the anomaly detection method used here relies on the quantization error (Qe) produced by abnormal patterns, compared to the Qe generated by control patterns used to train the map. Since the prototypes represent only control samples, a higher quantization error is expected when an abnormal pattern is represented by the BMU of the map. Therefore, a threshold for the quantization error is required to differentiate normal samples from anomalies. This threshold can be determined using a Naïve Bayes approach as follows.

Let CN represent the *Normal class* and DD the *Anomaly class*. According to Bayes' theorem, the probability of a new sample x belonging to class c can be expressed as:

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \quad (6.8)$$

where $P(c)$ is the prior probability of class c , $P(x|c)$ is the likelihood of x given

Table 6.1: Summary of the best classification results obtained for the different EEG bands. Bands providing the best accuracy values are highlighted in bold.

Stimulus	Band	Accuracy	Sensitivity	Specificity	AUC
4.8 Hz	Delta	0.74 ± 0.03	0.81 ± 0.03	0.65 ± 0.06	0.81 ± 0.05
	Theta	0.85 ± 0.04	0.91 ± 0.03	0.80 ± 0.08	0.88 ± 0.05
	Alpha	0.76 ± 0.04	0.82 ± 0.02	0.70 ± 0.09	0.85 ± 0.08
	Beta	0.86 ± 0.03	0.92 ± 0.02	0.78 ± 0.06	0.92 ± 0.08
	Gamma	0.83 ± 0.02	0.87 ± 0.02	0.76 ± 0.04	0.83 ± 0.05
16 Hz	Delta	0.73 ± 0.04	0.78 ± 0.03	0.67 ± 0.08	0.86 ± 0.07
	Theta	0.76 ± 0.04	0.79 ± 0.02	0.71 ± 0.08	0.91 ± 0.05
	Alpha	0.90 ± 0.02	0.93 ± 0.01	0.82 ± 0.04	0.93 ± 0.09
	Beta	0.90 ± 0.02	0.93 ± 0.02	0.86 ± 0.03	0.95 ± 0.09
	Gamma	0.79 ± 0.01	0.82 ± 0.02	0.74 ± 0.03	0.93 ± 0.10

class c , and $P(x)$ is the marginal probability, which can be calculated as:

$$P(x) = \frac{\#nearest\ samples\ to\ x\ in\ terms\ of\ qe}{\#total\ number\ of\ samples} \quad (6.9)$$

Using this approach, $P(c|x)$ can be determined, and the optimal threshold can be computed as the value of Q_e where $P(c|x) = 0$.

6.4 Results

We address the search for differential patterns from a functional connectivity perspective, using the circular correlation as a phase synchrony measure. To highlight the most significant differences in connectivity, a Mann–Whitney Wilcoxon (distribution agnostic) hypothesis test was applied. The methodology described in Section 6.3 for feature extraction and classification is applied separately for each EEG band and each stimulus, yielding the results summarized in Table 6.1. It is important to note that this classification methodology is assessed using stratified k-fold cross-validation ($k = 5$) to estimate the generalization errors, and a test was performed for each band. Fig. 6.4 illustrates the most significant connections as those with $p < 0.01$ in the hypothesis test. Furthermore, to remove potential spurious connections and identify significant connections more effectively, we applied the FDR (False Discovery Rate) correction ($\alpha = 0.05$) to the p-values, resulting in sparser resultant matrices.

Additionally, results in Table 6.1 are graphically depicted in Fig. 6.5, where the error bars corresponds to the standard deviation obtained in the cross-validation process. Specifically, Fig. 6.5 shows the classification results for the 4.8 and 16 Hz stimuli, respectively.

It is well known that the overall performance of a binary classifier can be assessed by exploring the Receiver Operating Curve (ROC) space, as it provides the cut-off point of the sensitivity-specificity trade-off. In other words, it shows the performance of the classifier in a compact form, regarding its capability of correctly identifying positive and negative samples. Relatedly, AUC is a very useful metric that represents the probability

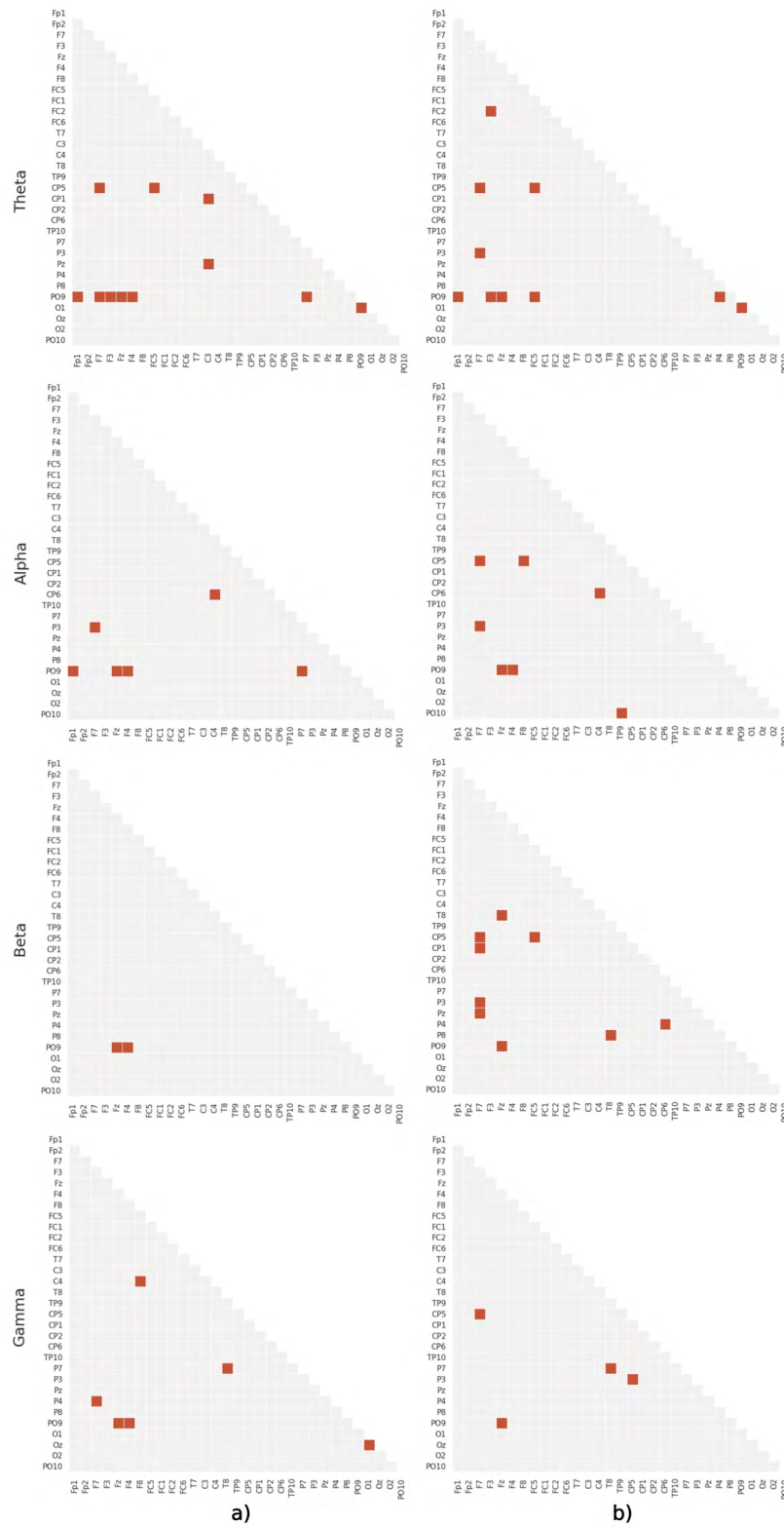


Figure 6.4: Statistically significant connections for meaningful bands. The axes represent the channels names according to the extended 10–20 EEG montage. Only connections with FDR-corrected $p < 0.01$ are shown. (a) 4.8Hz and (b) 16 Hz stimulus.

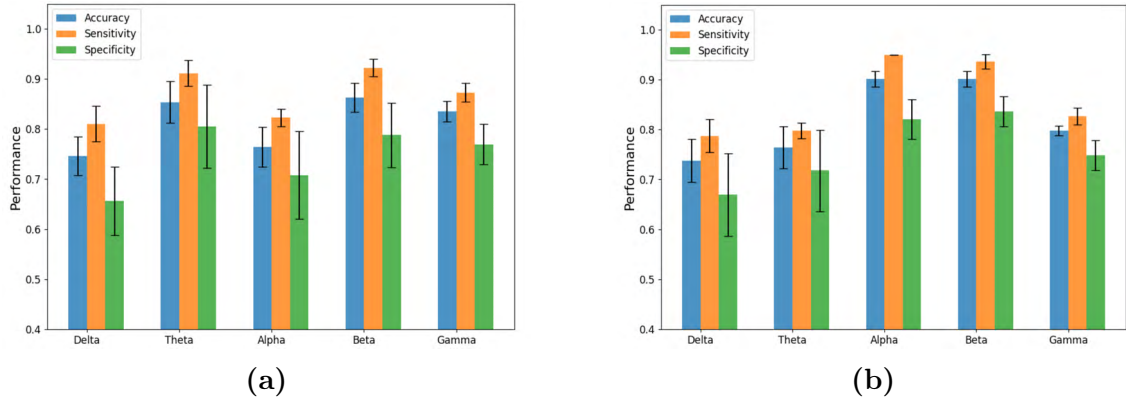


Figure 6.5: Classification results for a) 4.8 Hz and b) 16 Hz.

of a random positive to have a more extreme value of a negative one (and vice versa). Indeed, Fig. 6.6 show the ROC curves obtained for each band for the 4.8 Hz and 16 Hz stimulus, respectively, along with the corresponding AUC values computed for each curve.

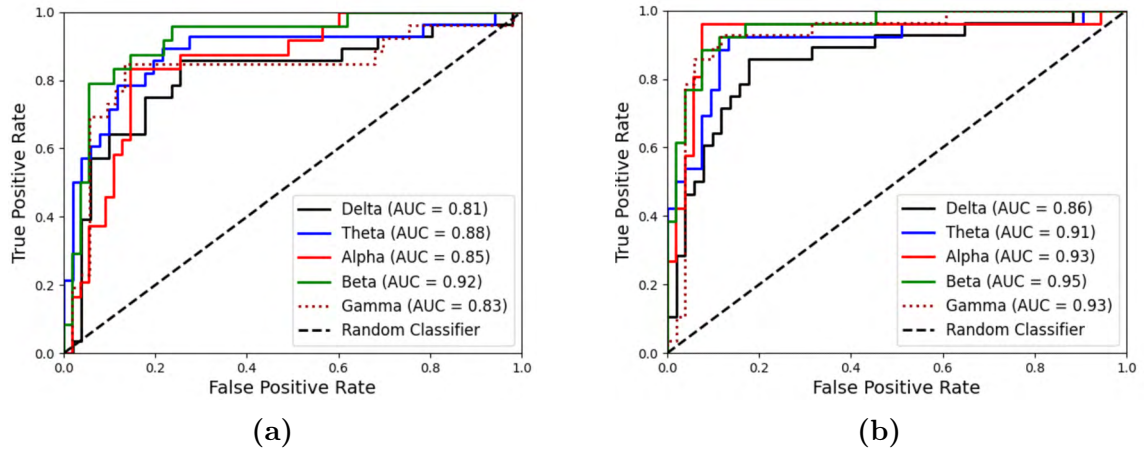


Figure 6.6: ROC Curves corresponding to the classification results for (a) 4.8 Hz and (b) 16 Hz stimulus.

Statistical Significance

The test used consists on generating a null distribution by calculating the accuracy of the classifier on 1000 permutations of the labels. This provides the distribution for the null hypothesis, which states the independence between features and labels and allows calculating the probability of reproducing the classification results with shuffled labels. This empirical p -value is then computed as:

$$p - value = \frac{\#permutations\ with\ accuracy\ higher\ than\ baseline}{\#Number\ of\ permutations} \quad (6.10)$$

The results of the permutation test are depicted in Fig. 6.7, where the null distribution obtained by permuting the labels as indicated above is shown in blue, and the accuracy obtained for the non-permuted case is indicated as a vertical red line. A 5-fold stratified cross-validation is carried out at each permutation iteration. Then, the average of the results obtained at these 5 folds is taken for the corresponding permutation iteration. Thus, Fig. 6.7 shows the probability density for the classification. This is the most computing demanding task as the classifier should be trained 1000 times. In our equipment (Dual Intel(R) Xeon(R) CPU E5-2640 v4 (10 cores)@ 2.40 GHz, 128GB of RAM) it took about one day to complete the permutations.

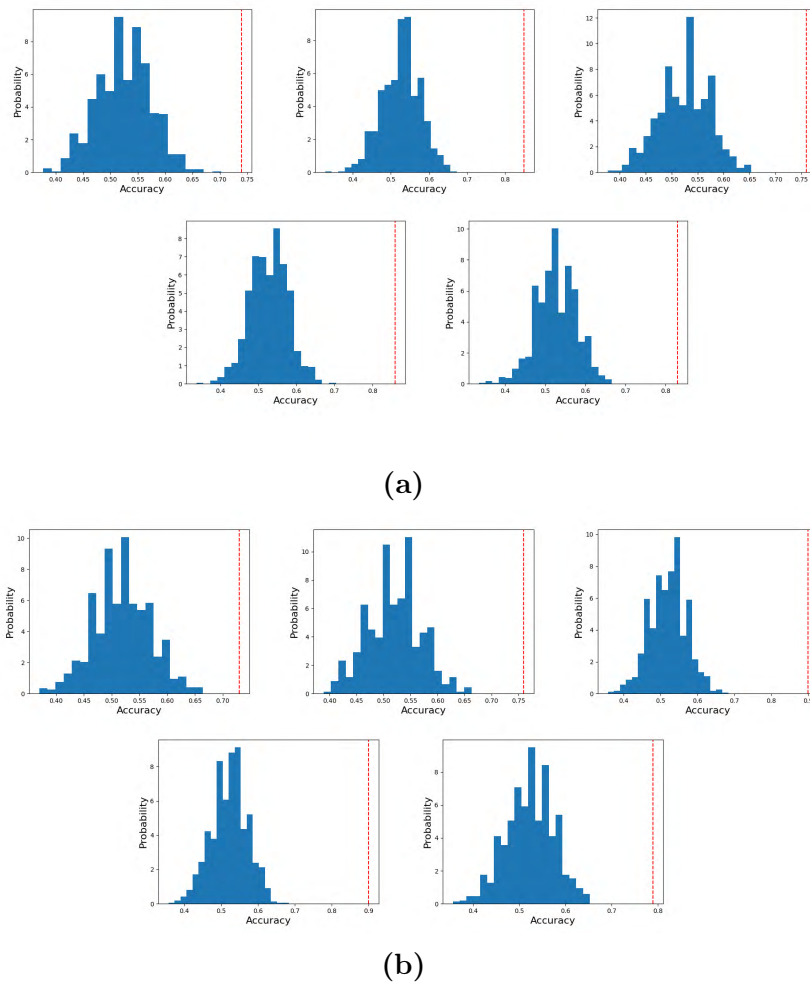


Figure 6.7: Permutation test results to assess the statistical significance of the results. Blue distribution corresponds to the null distribution and red line corresponds to the accuracy obtained with true labels. (a) 4.8 Hz stimulus and (b) 16 Hz stimulus.

6.5 Discussion

This section presents a discussion of the methodology, the results obtained, and the validation strategy, along with a comparison to previous studies addressing dyslexia

classification using biomedical signals.

The proposed approach leverages the quantization errors associated with SOM prototypes as features, which are then utilized to train a Naïve Bayes classifier that determines the optimal separation threshold. The SOM was trained using all connections, and the effectiveness of the classification algorithm was evaluated through 5-fold cross-validation and a permutation test involving 1,000 permutations. This ensures that the classification outcomes are not due to random chance and that genuine differences exist between the patterns of control and dyslexic subjects.

This validation strategy estimates the generalization error and constrains the probability that the results are obtained by chance. Using this methodology, Fig. 6.5 illustrates the discriminative capabilities of each EEG band. In other words, it highlights the influence of each stimulus on neuronal oscillations within specific frequency bands. Specifically, the beta and gamma bands yielded the best results for the 4.8 Hz stimulus, while the Alpha and Beta bands achieved the highest performance for the 16 Hz stimulus. Detailed results in Table 6.1 demonstrate an accuracy of 0.82 and an AUC of 0.92 for the 4.8 Hz stimulus. The AUC value indicates the probability of a positive sample receiving a higher score than a negative one, reflecting the misclassification probability. For the 16 Hz stimulus, the AUC increased to 0.95, indicating that the most discriminative patterns were identified in the Beta band.

Furthermore, connectivity matrices that include synchronization measures between each pair of electrodes form a connection graph, aiding in the identification of neuroanatomical regions involved in each case. Due to the significance level applied and the additional FDR correction, the resulting matrices are sparse. This method serves as a tool to identify synchronized channels related to connections between brain regions. Consequently, the circular correlation used to estimate phase-synchrony among channels effectively detects out-of-synchrony electrodes, which contribute to defining outlier (or abnormal) samples.

The highest classification performance was observed in the alpha, beta, and gamma bands. The results, as demonstrated subsequently, are highly significant, leading to the conclusion that connectivity anomalies exist in the dyslexic brain that the Bayesian classifier can detect. High-frequency patterns (beta, gamma) are crucial for interpreting segment-level representations in the brain [91]; thus, anomalies in this segmentation could result in cascade effects that lead to incorrect phoneme representation in the brain, as suggested by Virtala et al. [205], and subsequent inaccurate predictive processes.

A deeper analysis reveals that Fig. 6.4 displays the significant inter-channel connections for each stimulus and band. Notably, there is a recurrent interaction among occipital electrodes such as PO9, PO10, O1, O2, and Oz across various bands. These electrodes are highly specific to visual processing [73] and may reflect differences in visual activity during auditory stimuli. Many other electrodes are relevant to current language models [91], such as the connection between F7/FC5 and P3/CP5, corresponding to the "dorsal stream" in the Hickok and Poeppel model, which maps acoustic speech signals to frontal lobe articulatory networks. This pattern is present in the Theta and Alpha bands for both stimuli, and in the Beta/Gamma bands for the 16 Hz stimulus. The additional differences observed in the Beta/Gamma bands for the 16 Hz stimulus may indicate that phonemic-rate stimuli trigger abnormal functioning of

Table 6.2: Comparison of the proposed method with state-of-the-art methods for ADHD classification. The * denotes that the value was not reported in the original paper. The best results are highlighted in bold.

Method	Channels	Acq.Time	Accuracy	Sensitivity	Specificity	AUC
MRI + SVC [191]	T1-MRI	*	$0.8 \pm *$	$0.82 \pm *$	$0.78 \pm *$	*
MEG + SVC + GC [50]	253	3 min	0.63 ± 4.13	0.64 ± 4.01	0.65 ± 4.15	*
MEG + SVC + GE [50]	253	3 min	0.94 ± 1.78	0.93 ± 1.39	0.93 ± 2.32	*
MEG + SVC + CI [50]	253	3 min	0.80 ± 1.14	0.80 ± 1.41	0.79 ± 2.17	*
MEG + SVC + wIFCG [50]	253	3 min	0.97 ± 1.89	0.96 ± 1.89	0.95 ± 1.98	*
EEG + SVC [160] (Writing Task)	32	1 min	$0.59 \pm *$	$0.64 \pm *$	$0.53 \pm *$	*
EEG + SVC [160] (Typing Task)	32	1 min	$0.78 \pm *$	$0.88 \pm *$	$0.66 \pm *$	*
EEG + OCSVC [155]	32	5 min	$0.71 \pm *$	$0.53 \pm *$	$0.78 \pm *$	$0.79 \pm *$
EEG + DAE [139]	32	5 min	$0.56 \pm *$	$0.76 \pm *$	$0.66 \pm *$	$0.74 \pm *$
Proposed	32	5 min	0.90 ± 0.02	0.93 ± 0.02	0.86 ± 0.03	0.95 ± 0.09

higher cognitive tasks.

The P3 and CP5 reception fields are located near the temporo-parietal junction, and differences in patterns may result from variations in electrode placement. However, it is reasonable to assume that they represent the sensorimotor interface in the Hickok and Poeppel model. Similarly, F7 and FC5 may capture activity in Broca's area, considered an articulatory region in the Hickok and Poeppel model. Occipital-frontal connections, such as PO9 to Fz and F4, might be attributed to hypothesized conceptual networks distributed throughout the cortex that connect lexical and articulatory regions.

Conversely, the right hemisphere exhibits fewer but still significant anomalies detected by the analysis. Notably, there are significant anomalies at CP6-C4 in the alpha band for both stimuli, potentially corresponding to connections between the right auditory cortex and the right sensorimotor cortex. This deficit has not been identified elsewhere but may align with recent research indicating atypical rhythmic entrainment in auditory and sensorimotor coupling at low rates [36]. Additionally, two notable differences are observed in bilateral connections within the gamma band: the P4-F7 anomaly, corresponding to Molinaro's main finding [144] for dyslexia, and the P7-T8 deficit, involving an anomaly in the connection between the right phonological network (T8) and the right lexical interface in the Hickok and Poeppel model [91].

The discriminative capability of the proposed method was evaluated using Receiver Operating Characteristic (ROC) curves, which illustrate the trade-off between sensitivity and specificity. Therefore, ROC curves for all bands are presented in Fig. 6.6 for (a) the 4.8 Hz stimulus and (b) the 16 Hz stimulus. In both cases, a favorable balance between sensitivity and specificity is achieved.

The number of EEG studies utilizing auditory stimuli for dyslexia diagnosis is limited and typically focused on exploratory analysis [48, 144], in the search for the biological basis of dyslexia, which remains an open research question. Dyslexia classification using biomedical signals is also a current research trend, contrasting with traditional methods based solely on behavioral tests. Studies employing structural imaging [191], MEG [50], and EEG [137, 155, 160] signals are summarized in Table 6.2. Similar to [50], classification with biomedical signals often utilizes various features extracted from MEG signals, achieving accuracy values up to 0.93. However, this approach requires a 255-channel MEG acquisition system, and MEG data is generally more challenging to obtain and more computationally intensive to process than EEG data. Conversely,

EEG-based dyslexia classification studies have achieved accuracy values up to 0.80 [160]. However, these studies typically rely on reading and writing tasks, which depend on participants' acquired skills and limit the diagnosis age. Our methodology, based on analyzing responses to simple auditory stimuli, overcomes these limitations by simplifying the test and avoiding potential biases introduced by interactive stimuli or specific task designs. Furthermore, the proposed method enhances the understanding of the neural basis of dyslexia by revealing differential functional connectivity patterns when comparing control and dyslexic subjects, as clearly illustrated in the connectivity matrices highlighting the most significant connections.

Additionally, our classification results are comparable to those obtained using MEG signals but can be achieved with EEG equipment, which can be portable, facilitating task development and application to young children. This is particularly important for our future work, which will continue this research with pre-readers.



UNIVERSIDAD
DE MÁLAGA

Chapter 7

EEG Entrainment to Speech Rhythms

This chapter investigates the synchronization between auditory stimuli and responses in the context of dyslexia. Unlike works that study the synchronization between two EEG bands or electrodes, here we propose the exploration of the entrainment between different EEG bands and the stimulus. This also enables interpreting the response to different stimuli in the TSF framework, using Cross Frequency Coupling methods to estimate the connectivity between brain areas by means of synchronization. This has been accomplished by different Phase-Amplitude Coupling statistics such as Heights Ratio (HR), Mutual Information (MI), Phase-Locking Value (PLV), Gaussian Copula PAC (gcPAC), and Mean Vector Length (MVL). Moreover, feature selection techniques and non-linear classifiers have been used to assess entrainment patterns with the signals acquired at different EEG electrodes. As explained in Chapters 4 and 3, the complexity of the models makes difficult to extract conclusions about the patterns computed. Hence, Shapley Additive Explanations (SHAP) is used to estimate feature importance, providing insights into how different EEG channels and frequency bands contribute to the classification. Overall, the experiments carried out showed that controls exhibit stronger gamma-band entrainment in left-hemisphere regions associated with language processing (e.g., F3, FC1, P3), while dyslexic subjects show increased activity in right-hemisphere regions (e.g., CP2, FC6), suggesting potential compensatory mechanisms. The combination of the HR metric and the 40 Hz gamma stimulus achieves up to 90% accuracy in distinguishing dyslexic individuals from controls, indicating that differences in phoneme gamma entrainment may serve as a biomarker for dyslexia.

7.1 EEG Entrainment and Auditory Steady-State Response

Dyslexia research has often utilized pure speech or spectrally-simple stimuli, such as Amplitude Modulated (AM) carriers where the amplitude envelope is modulated by the patterns of multi-band temporal excitation formed in speech. Nevertheless, recent research shows that phonological awareness disabilities may actually arise from more basic processing problems with sound. This cascade effect of bottom-up nature implies that these higher-level phonological deficits may arise from more basic difficulties in sound processing [74, 205]. This would result in the idea of providing less complex

non-speech auditory stimuli presented at more typical speech rates thereby identifying these core auditory processing deficits underpinning phono-attentional disabilities. There are many really interesting studies that have shown promising results with EEG recordings from dyslexic children listening to non-speech sounds [58, 139, 156]. An analysis of both temporal and spectral features of such recordings has revealed the presence of significant differences between children with dyslexia and typical readers. And interestingly, some of these neural response measures seem to relate directly to tests of measures of phonological awareness, where the connection between basic auditory processing and reading ability could not be any clearer. Here we conducted experiments to explore the synchronization between the EEG responses and the auditory stimuli. Thus, we focused on creating Amplitude Modulated (AM) white-noise at frequencies that correspond to key speech components in Spanish at specific frequency bands: a 4.8 Hz theta rhythm matching syllabic rates, a 16 Hz beta frequency corresponding to intrasyllabic patterns (which capture transitions like /p/ to /a/ in /pa/, or more complex sequences like /pl/ to /æn/ in /plæn/), and a faster 40 Hz gamma frequency aligned with individual phoneme processing.

7.2 CFC-based synchronization between EEG and auditory stimuli

This section describes the [Phase-To-Amplitude Coupling \(PAC\)](#) metrics to analyze EEG responses to auditory stimuli in dyslexic and control subjects, focusing on the interaction between EEG signals and speech rhythms. Metrics like Heights Ratio (HR), Mutual Information (MI), and Phase-Locking Value (PLV) are computed across frequency bands.

As stated before, this study applies several [PAC](#) metrics, previously introduced in Section 3.4. The Phase-Locking Value (PLV) [145] is calculated using the phase of the low-frequency signal, $\phi_1(t)$, and the phase of the high-frequency envelope signal, $\phi_{a_2}(t)$, as follows:

$$PLV = \frac{1}{N} \sum_{t=0}^{N-1} |e^{-j\phi_{a_2}(t)} e^{-j\phi_1(t)}| \quad (7.1)$$

where N represents the signal length for $t \in 1, \dots, N$.

Heights Ratio (HR) and Mutual Information (MI) [194] utilize vector quantization of $a_2(t)$ relative to the phase signal $\phi_1(t)$. The phase range $[-\pi, \pi]$ is divided into J quantiles, creating a counting vector p_j for each j^{th} quantile of the instantaneous phase vector $\phi_1(t)$:

$$p_j = \frac{1}{M} \sum_{t \in T_j} a_2(t) \quad \text{and} \quad T_j = \{t | \phi_1(t) \in j^{\text{th}} \text{ quantile}\} \quad (7.2)$$

where M is the number of elements in the set T_j .

In early approaches to computing MI [193], the quantized vector $\mathbf{p} = [p_0, \dots, p_{J-1}]$ is normalized to sum to 1, allowing elements P_j to be interpreted as probabilities of

phase bins. MI is then derived as the Kullback-Leibler divergence of P from a uniform distribution:

$$MI = 1 + \frac{1}{\log(J)} \sum_{j=0}^{J-1} P_j \log(P_j) \quad (7.3)$$

The HR computation is more straightforward, as defined in [195]:

$$HR = \frac{\max(\mathbf{p}) - \min(\mathbf{p})}{\max(\mathbf{p})} \quad (7.4)$$

Gaussian Copula PAC (gcPAC) [103] employs a continuous MI computation approach for EEG signals. The method, underpinned by statistical copula theory, assumes that marginal distributions are standard Gaussian. According to Sklar's theorem, the joint distribution entropy $H(X, Y)$ can be expressed as the sum of marginal entropies and copula entropy:

$$H(C) = H(X, Y) - H(X) - H(Y) \quad (7.5)$$

MVL [21], suitable for high-quality data, measures the magnitude of a signal with amplitude $a_2(t)$ and phase $\phi_1(t)$:

$$MVL = \frac{1}{N} \sum_{t=0}^{N-1} |a_2(t) e^{j\phi_1(t)}| \quad (7.6)$$

7.3 Significance Testing via Permutation

Already described in 3.4.2, once a PAC value has been computed for two signals, it might be possible that high PAC was produced by chance. In order to discard this possibility we use permutation testing to generate Z-scores, which give us an idea of the statistical significance of the observed PAC measure PAC_o as in [195] and [7].

In order to do so, we generate surrogate versions of the amplitude signal by barrel-switching the amplitude signal $a_s(t)$ at different levels. Surrogate PAC measures are computed for each $a_s(t)$, generating a surrogate PAC distribution from which Z-scores for the observed PAC (PAC_o) can be computed as:

$$Z = \frac{PAC_o - \mu_{PAC_s}}{\sigma_{PAC_s}} \quad (7.7)$$

with μ_{PAC_s} and σ_{PAC_s} the mean and standard deviation of the surrogate PAC distribution. In this work, 200 surrogates are generated for computing the PAC Z-score between the phase of the stimuli and the amplitude of EEG signals at different frequencies. This Z-score will be considered henceforth as the PAC measure.

7.4 Feature Aggregation and Classification

We computed a single PAC Z -score to quantify the interaction between a synchronized stimulus $\phi(t)$ and a narrowband-filtered EEG signal, $a(t)$, at each electrode. The signal $a(t)$ was filtered at 49 frequency points spanning logarithmically from ω_ϕ (the stimulus frequency) to 200 Hz. This process produced an array of 31×49 PAC values per EEG segment, corresponding to every combination of EEG channel and the center frequency of the bandpass filter applied to extract $a(t)$. Spectral features are often grouped into frequency bands [48, 57, 122, 129, 144]. For this study, we adopted the following frequency band definitions: Delta (0.5–4 Hz), Theta (4–8 Hz), Alpha (8–12 Hz), Beta (12–25 Hz) [10], and two subdivisions of the Gamma band: Gamma (25–80 Hz) and Epsilon (80–200 Hz), also referred to as High-Gamma frequencies [129, 144].

The aggregation of PAC values across frequency bands was conducted as follows. For each band, a set of PAC values in Z -score form, $PAC_{band} = \{PAC_{f_i}, \dots\}$, was defined for all f_i within the band's frequency range (f_{ini}, f_{end}). The final PAC value for a given band B was determined by counting the PAC values within PAC_{band} exceeding a threshold $Z_{th} = 1.65$, corresponding to the 95th percentile of the permutation set. This count was then normalized by the total number of elements in PAC_{band} as shown in Fig. 7.1 The process was repeated for each 5-second EEG segment, and the final PAC for each subject was obtained by averaging across all segments.

Support Vector Machine (SVM) Analysis

To evaluate the diagnostic utility of the aggregated PAC-band features, we employed a linear Support Vector Machine classifier (SVC) [203]. SVCs are widely used in machine learning [138], particularly for their linear (basic) and radial basis function (RBF) kernel implementations. These classifiers offer a robust balance between accuracy and generalizability.

The linear SVC allows for interpretability by providing classification weights, enabling insights into the relevance of input features, even in probabilistic terms. Two regularization metrics—L1 and L2 norms—were tested. While the L2 norm generally delivers higher performance, the L1 norm promotes sparser matrices, emphasizing fewer features. Both approaches were evaluated using the Python implementation of LIBSVM [24].

7.5 Evaluation and Interpretability

Our hypothesis states that control and dyslexic subjects differ in the way they process auditory stimuli, consistent with the phonological hypothesis for the origins of Dyslexia and the Temporal ampling Framework (TSF) described in [80, 136]. Consequently, there may be differences in PAC coupling between the stimuli presented to and the EEG signals acquired from experimental subjects.

The classification is performed as follows. First, the features are standardized and divided into training and testing sets. Next, a Support Vector Classifier (SVC) (see Section 7.4) is trained. To evaluate performance, repeated stratified 5-fold cross-

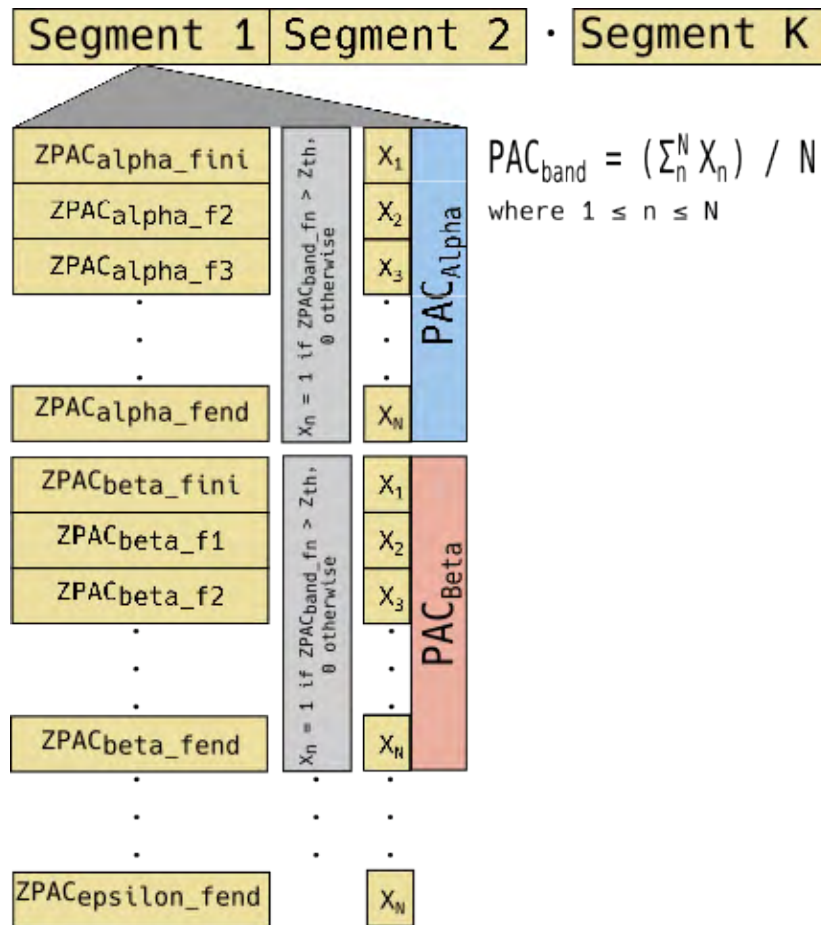


Figure 7.1: PAC aggregation to compute the final score for a subject. For each segment and every band a ZPAC score is computed and then averaged across all segments.

validation is employed. This involves randomly splitting the dataset into 5 subsets, each maintaining the same class proportions (stratified). Four subsets are combined to form a training set, while the remaining subset serves as the test set. This process is repeated 50 times to ensure robustness by testing across different training and testing splits.

SVC coefficients are used to assess interpretability. The SVM linear discrimination model learns a function $\mathbf{y} = \mathbf{w}\mathbf{x}$, yielding a coefficient vector $\mathbf{w}$ for the input data. The input data consist of the aggregated PAC Z-measures B for each electrode and band, as previously described. Then, a positive coefficient indicates that the contribution of B favors predicting a dyslexic (DLX) subject, while a negative coefficient suggests the contribution favors predicting a control (CTL) subject.

Azimuthal Lambert Projection

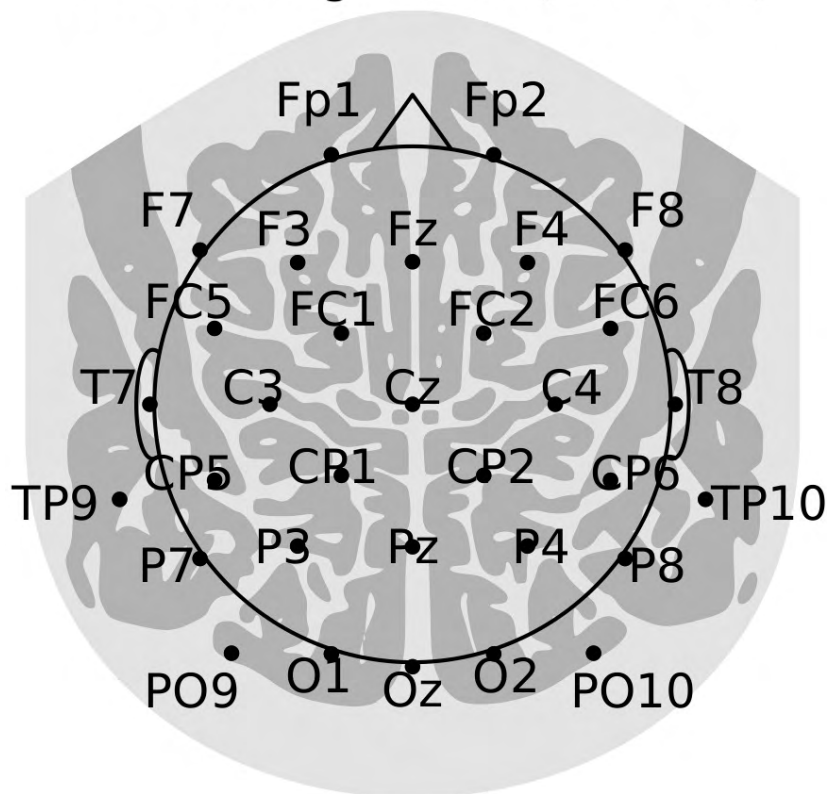
The Lambert azimuthal equal-area projection is a map projection that accurately represents area across the entire map, ensuring that all regions are depicted in their true relative sizes. However, it does not preserve shapes or angles, leading to distortions, especially near the edges of the map. This projection is particularly useful for thematic mapping where area comparison is essential.

In this work, the Lambert azimuthal projection is employed to project the three-dimensional positions of EEG electrodes onto a two-dimensional plane. This method facilitates the visualization and analysis of electrode placements and corresponding brain regions. By preserving the area relationships, this projection ensures that the spatial distribution of electrodes reflects their true positions on the scalp, aiding in accurate interpretation of EEG data. As the analysis of EEG entrainment to auditory stimuli often involves comparing activation across specific regions (like hemispheric differences in dyslexia), this projection ensures that these comparisons are not biased by distortions in electrode placement visualization. To improve and complement this visualization method, an MRI template has been projected and depicted in the background, while adjusting the scale to overlap the electrode positions. This aims to facilitate the correspondence between electrodes and anatomical positions.

Technically, the mapping process involves aligning the projection center with the top of the scalp, minimizing edge distortions and preserving the relative positions of electrodes in the standard 10-20 system (Fig. 7.2). This alignment ensures compatibility with subsequent analysis steps, including PAC metric computations and SVM-based classification, where spatial interpretations are vital.

The projection is particularly effective when analyzing differences in entrainment across electrodes, as shown in the experiments presented in this chapter. For example, the observed increased gamma-band entrainment in left-hemisphere regions for controls (e.g., F3, FC1, P3) and in right-hemisphere regions for dyslexic individuals (e.g., CP2, FC6) can be visually highlighted using this projection (Fig. 7.5). By preserving the spatial relationships of electrodes, the Lambert azimuthal map aids in identifying patterns that correspond to language processing areas or compensatory mechanisms in dyslexia.

10-20 configuration (Lambert)



Lambert azimuthal projection

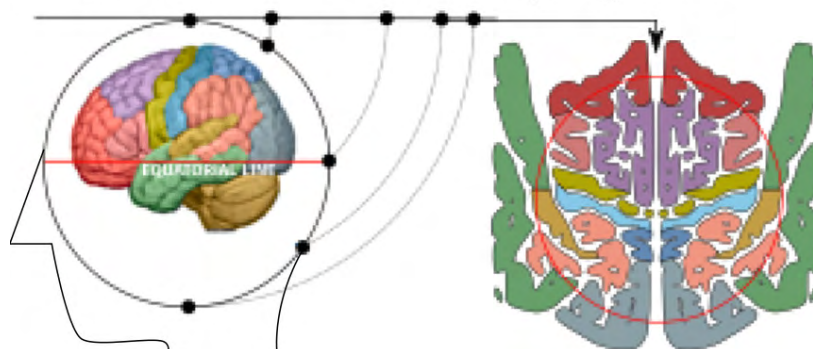


Figure 7.2: 10-20 placement of the EEG electrodes in the Lambert azimuthal representation. Below: brief explanation of the azimuthal projection of the region areas.

7.6 Results

Our classification results reveals that HR is by far the most effective PAC measure for the stimuli and segments, achieving up to 90% accuracy for the HR and 40 Hz stimulus. Fig. 7.3 illustrates the distribution of performance achieved by each PAC measure, as evaluated by balanced accuracy (trade-off between sensitivity and specificity) and sensitivity. Each point represents a performance estimation using a different combination of EEG band, the parameter C of the SVC, and the SVC penalty across all 31 channels. These parameters were chosen via grid search. This search is a systematic method to optimize hyperparameters by exhaustively evaluating combinations of parameters. This enables the identification of the best-performing measures and stimuli for distinguishing subjects with and without dyslexia.

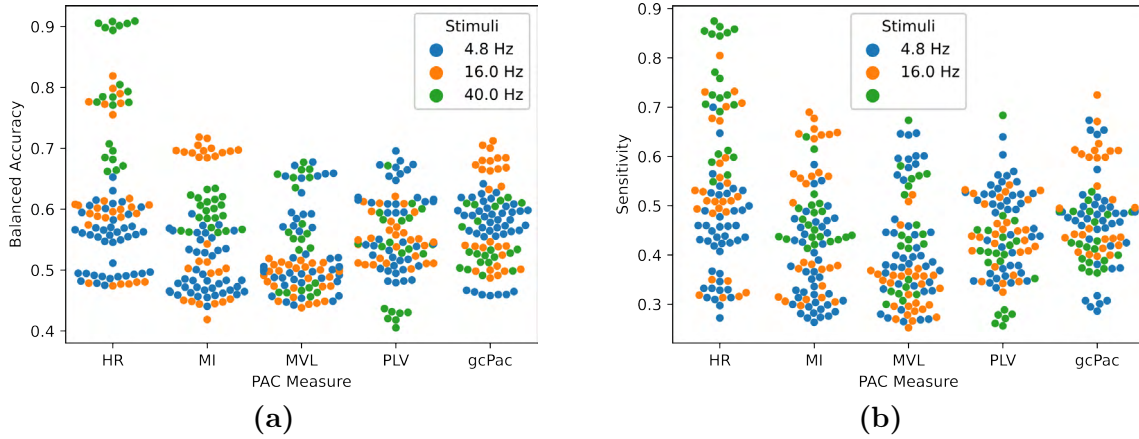


Figure 7.3: Comparison of performance measures, **a)** balanced accuracy and **b)** sensitivity, in experiments with different hyperparameters (C , penalty) for each stimulus and PAC estimation method.

While the γ -stim (40 Hz) performs best when measured using the Heights Ratio (HR), the 16 Hz β -stim also achieves high performance with HR (around 80%) and slightly lower performance with MI (approximately 70% sensitivity and balanced accuracy). For the 4.8 Hz θ -stim, all PAC measures perform similarly with poorer sensitivity, though specificity is slightly higher for PLV and MVL.

All combinations of stimuli, PAC measures, and classifier parameters are depicted in the parallel plot shown in Fig. 7.4. Each line in the plot represents a combination of parameters and performance, offering insights into individual parameter configurations. The color represents the PAC measure. From this, we observe that the SVC's performance is relatively independent of the parameter C (used for regularization). However, the choice of penalty (L2 vs. L1) moderately impacts performance, with the L2 norm yielding higher accuracy and sensitivity. The L2 norm is particularly advantageous for the β -stim.

Regarding the combination of stimulus and band, the γ -stim and the gamma band with HR demonstrate the highest discriminative power. Significant separation is also achieved by the β -stim with gamma band, as well as γ -stim with Epsilon and 16 Hz with Epsilon, all measured using HR. To delve deeper into this performance, we examine the

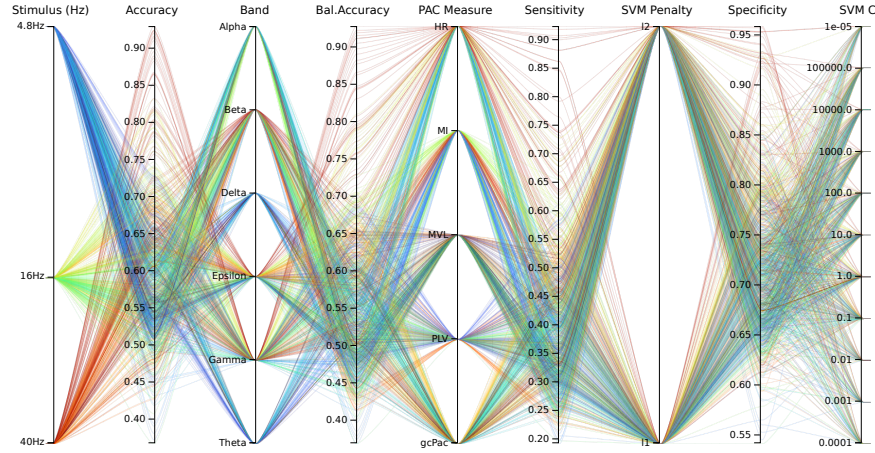


Figure 7.4: Parallel plot tracing the performance of the different combinations of stimuli, PAC measures, and SVC parameters: penalty and C .

SVC coefficients (weights) for these stimuli and bands, measured by HR, in Fig. 7.5.

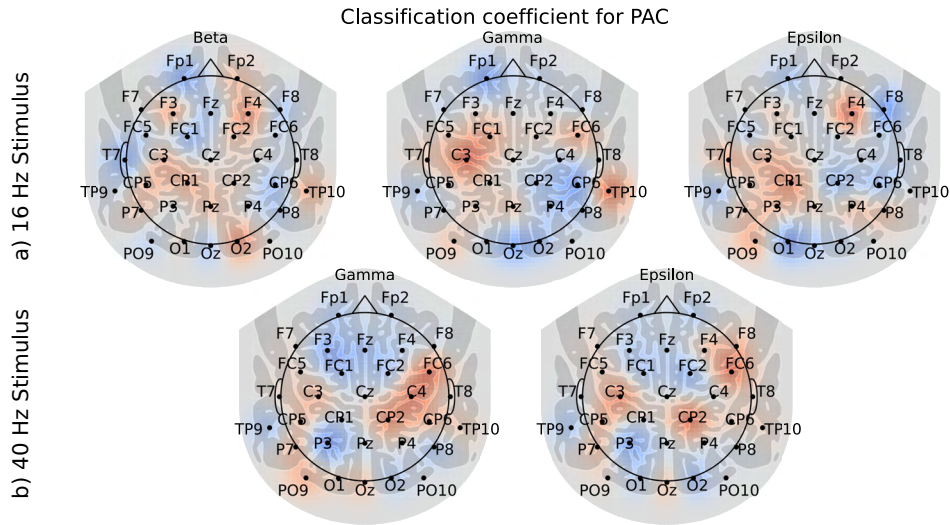


Figure 7.5: Classification weights of the linear SVC trained to discriminate between DLX and CTL subjects using the HR entrainment between the stimuli (16 and 40 Hz) and different bands.

The contribution of electrodes in the two hemispheres is asymmetrical for the β -stim and γ -stim entrainment with Beta, Gamma, and Epsilon bands. For the best-performing combinations (β -stim and γ -stim with Gamma), there are notable differences and similarities. For the β -stim (16 Hz), the largest positive coefficients are observed at electrodes C3 and TP10, with negative coefficients around CP5, P4, and C4. For the γ -stim (40 Hz), the largest contributions are along the line formed by CP2-C4-FC6, with large negative coefficients near key language areas: P3, FC1, and F3. This pattern is consistent across Beta, Gamma, and Epsilon bands. The distribution of the aggregated PAC values is depicted in Fig. 7.6 in a SHAP summary plot (see section 4.6), highlighting these differences.

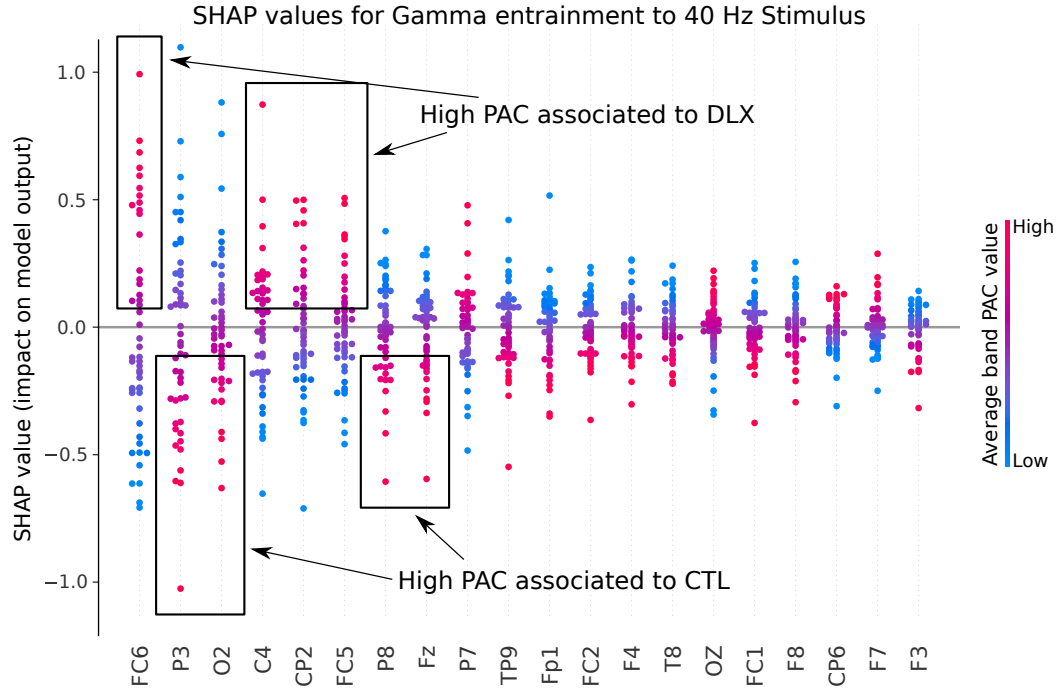


Figure 7.6: SHAP values for the DLX class using Gamma-average HR PAC for the 40 Hz stimulus. Color indicates the actual PAC value.

7.7 Discussion

Our results provide solid evidence of a link between neural oscillations and abnormal reading in children. We investigated the brain’s entrainment to three different stimuli, as measured by five distinct PAC strategies. Entrainment was then aggregated by EEG frequency band, and the resulting features were used as input to a classifier tasked with differentiating between two groups: dyslexic (DLX) and control (CTL) subjects. Consistently, we achieved accuracy above 90% for the entrainment between the γ -stim (40 Hz) and the Gamma band, as measured by Heights Ratio (HR). Furthermore, HR entrainment between the β -stim (16 Hz) and Gamma band achieved accuracy above 80%, regardless of the classifier hyperparameters used.

These results suggest that differences in phoneme gamma entrainment are a potential biomarker for detecting dyslexia. Giraud and Poeppel [74] proposed that fast-rate information in lower gamma is crucial for distinguishing features in predictive speech processing. As a result, increased sensitivity to phonemes in dyslexic individuals may explain difficulties in encoding these into language-level phoneme categories [75, 115]. This may also explain why the γ -stim, which mimics the phoneme rate in Spanish, performs best as a stimulus.

HR is the best-performing PAC measure, followed by MI. This is not coincidental, as both measures depend on the quantization of the instantaneous phase vector, as defined in Eq. 7.2. HR is particularly effective at detecting large differences in the quantized amplitude vector, as shown in Eq. 7.4. Larger differences imply a higher modulation index, which in our case is better captured by the simpler HR measure.

Table 7.1: Comparison of the performance of this and other state-of-the-art methodologies for the differential diagnosis of DLX.

Work	Accuracy	Sensitivity	Specificity
[59] Left H.	-	0.846 [0.22]	0.803 [0.3]
[59] Right H.	-	0.715 [0.15]	0.608 [0.41]
[51] SVC + CI	0.803 [1.14]	0.802 [1.41]	0.793 [2.17]
[51] SVC + wIFCG	0.970 [1.89]	0.967 [1.89]	0.951 [1.98]
[139] EEG 4.8 Hz	0.783 [0.12]	0.471 [0.31]	0.896 [0.24]
[139] EEG 16 Hz	0.827 [0.04]	0.325 [0.01]	1.000 [0.09]
[139] EEG 40 Hz	0.733 [0.15]	0.441 [0.33]	0.836 [0.26]
[58] EEG 4.8 Hz	0.86 [0.03]	0.92 [0.02]	0.78 [0.06]
[58] EEG 16 Hz	0.90 [0.02]	0.93 [0.02]	0.86 [0.03]
This work 4.8 Hz	0.738 [0.13]	0.791 [0.11]	0.611 [0.14]
This work 16 Hz	0.830 [0.11]	0.932 [0.09]	0.686 [0.15]
This work 40 Hz	0.929 [0.08]	0.931 [0.05]	0.922 [0.07]

MVL exhibited poorer performance, likely due to its reliance on high-quality signal segments [21]. Its performance decreases in real-world noisy environments such as ours. A similar observation can be made for PLV. Notably, PLV and MVL perform better with the θ -stim (4.8 Hz), possibly due to the higher phase resolution of the stimulus (over 100 points per cycle compared to 12 points per cycle for the 40 Hz stimulus at $F_s = 500$). This higher resolution benefits PAC estimation, although their performance remains modest (< 0.7 accuracy). On the other hand, gcPAC assumes signal Gaussianity, which was not supported by our closer inspection of the EEG signals.

Other studies have classified dyslexic and control subjects using entrainment data with EEG [58, 59, 139], MEG [51], and MRI [191]. Table 7.1 compares the performance of these methods with ours.

Our algorithm achieves state-of-the-art performance while maintaining high interpretability. Unlike methods requiring assumptions for connectivity calculations, our approach excels in performance among EEG-based methods, which are more cost-effective than MRI or MEG. Additionally, our results align with current theories of language [90], and the SVC-learned coefficients provide insights into the neurological basis of dyslexia.

The classification coefficients (Fig. 7.5) reveal the contributions of electrode-specific entrainment. High entrainment in large areas of the right hemisphere (FC6, C4, CP2) and C3 contributes to classifying a subject as DLX, while greater entrainment in left-hemisphere regions (F3, FC1, P3) favors classification as CTL. The distribution of Gamma entrainment with the γ -stim is detailed in Fig. 7.6, showing higher band-average PAC for controls at F3, FC1, and P3, but lower values at FC6, C3, and CP2.

Key areas in the left hemisphere, crucial for normal reading [62, 90], exhibit higher gamma entrainment to phoneme stimuli. Conversely, right hemisphere regions (CP2, FC6) are more active in dyslexics, negatively correlating with reading performance. These findings support recent evidence suggesting that dyslexic language performance is compensated through right-hemisphere mechanisms in both adults [183] and ado-

lescents [95]. Our analysis confirms the utility and interpretability of the proposed methodology, which not only aids in differential diagnosis but also enhances understanding of the neurobiological basis of phonological processes.

Chapter 8

Cerebral Lateralization Assessment using Deep Learning

Deep Learning (DL) is one of the most powerful methods to learn patterns from data. Specifically, DL allows learning features from data and discover complex patterns by linear or non-linear relationships between features. However, the practical use of DL architectures in biomedical problems is challenging, due to the need of large amounts of data for generating the model by the learning process.

This chapter proposes the use of deep learning techniques for analyzing cerebral lateralization, focusing on the differentiation of functional and structural asymmetry between brain hemispheres, particularly in the context of neurological disorders like dyslexia. Brain lateralization refers to the specialized roles of the left and right hemispheres, which are morphologically similar yet functionally distinct. Disorders such as Alzheimer's, schizophrenia, and dyslexia have been linked to atypical lateralization patterns. Traditional approaches to studying lateralization often rely on statistical analyses of MRI or EEG data. However, subtle differences in brain activity can be challenging to detect with conventional methods. To address this, the chapter proposes a deep learning architecture that estimate phase-amplitude coupling from raw signals, incorporating the capabilities of **Recurrent Neural Network (RNN)** with **Long Short-Term Memory (LSTM)** units and **Convolutional Neural Network (CNN)** for processing EEG time-series data. This module is pre-trained on synthetic data by means of a EEG synthetic data generation method, overcoming the problem of limited data availability. Furthermore, the transfer learning paradigm is used to construct a Siamese architecture allowing comparing data pairs, making it ideal for analyzing hemispheric differences. Additionally, the proposed architecture integrates a Channel Attention Layer to prioritize the most relevant EEG channels, reducing computational complexity and enhancing interpretability.

Results from analyzing EEG data of dyslexic and control groups demonstrate the network's ability to identify significant lateralization patterns. Notably, the network's findings align with existing research on altered alpha-gamma wave coupling in dyslexics. Topographical maps generated by the attention mechanism provide insights into the spatial significance of brain regions in lateralization, emphasizing the role of frontal and auditory cortical areas. The study's framework achieves competitive classification performance, as evidenced by area under the curve (AUC) metrics, and offers a robust

approach for examining lateralization in dyslexia. Beyond classification, the method provides valuable insights into the functional asymmetries of the brain, with potential applications in broader neurological research.

8.1 Brain Lateralization

The brain is mainly divided into two clearly differentiated hemispheres. Although morphologically symmetrical, we know that each of the hemispheres processes different information in different ways and controls different kinds of motor behavior. This is what is called brain lateralization [153]. This phenomenon has been found across species from insects to birds and mammals [176]. It is important to note that brain lateralization is not a strict division of labor. The two hemispheres are interconnected, allowing for constant communication and collaboration between the two sides. This interplay ensures that most cognitive tasks involve both hemispheres work together, increasing the cognitive capacity.

Several studies have found that this lateralization is atypical in brains with disorders; from Alzheimer's disease [128], schizophrenia [173], depressive disorders [16], attention-deficit hyperactivity syndrome [127] or several speech disorders including dyslexia [152].

Usually, the study of laterization is trough **MRI** or **EEG** signals. Several statistical metrics are then calculated and compared between the control and the affected group. For example, in the work [16] compared the alpha waves (8-12Hz) in both hemispheres and found that there are relatively less activity over left than right frontal sites in depressed subjects. In this case, it is enough to analyze the wave in both hemispheres and realize the difference, but sometimes the differences can be so subtle that they escape us. It is in these scenarios when **AI** could help. Tools like **Deep Learning** have proven to be useful when analyzing complex data, like **EEGs**. In the context of kind of data, **RNNs** (section 4.3.2), especially those with **LSTM** (section 4.3.2.1) blocks, are the best suited for processing the EEG time series data. These models incorporate a *memory* module that handles long-term dependencies, being the best match to perform brain analysis [26, 58].

Another network architecture to analyze time-series are **CNNs** (section 4.3.1). Although initially developed to work with images, the application to temporal data has been successfully realized. 1D **CNNs** are widely used for their ability to learn hierarchical features, capturing spatial-temporal trends [1]. These features can be fed into another neural network or any IA algorithm to perform a supervised or unsupervised task [26]. One common example is EEGNet, which has been successfully used in numerous applications, such as the study of visual evoked potentials or to distinguish ADHD subtypes and controls[47, 119, 199].

In the analysis of the lateralization of the brain, a special architecture called Siamese network in combination with **CNNs** layers and **LSTMs** is used.

8.2 Siamese Networks

Siamese networks are an interesting and flexible architecture in machine learning, especially when it comes to problems related to similarity or recognition. Siamese neural networks have received their name from Siamese twins, which represent the functioning of two similar sub-networks in parallel.

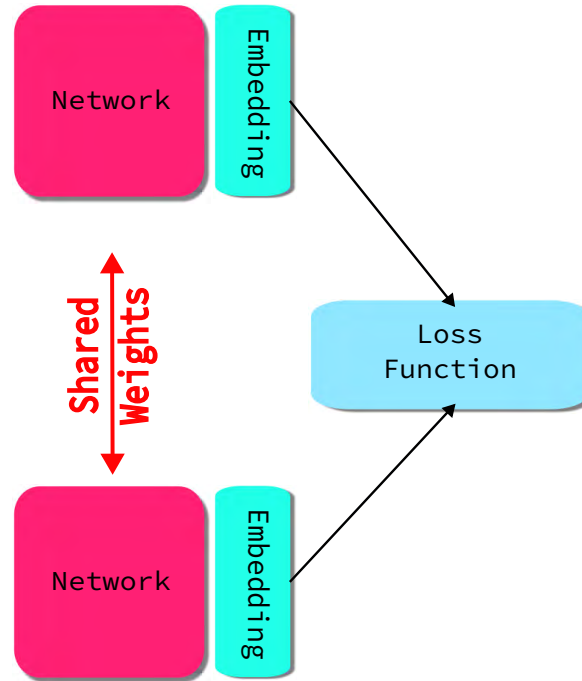


Figure 8.1: A Siamese network is actually the same model in both situations. But the input is considered in pairs to apply some special loss function between the output vectors, commonly called embeddings.

They learn and compare features from pairs of inputs. Therefore, they are very useful for tasks in which judging similarity or dissimilarity is important. The idea of a Siamese network is to make two inputs go through the same neural networks with shared weights enabling the extraction of meaningful features and computation of a similarity score, effectively learning a similarity metric directly from the data.

One of the most significant advantages of Siamese networks is their efficiency in handling scenarios with limited training data. While traditional neural networks need a huge amount of data points for each class, Siamese networks can generalize great even after being trained on very few examples. Which is what makes the way they function crucial in one-shot learning tasks, where a model needs to learn and generalize from very few (only 1 or minimal) training examples.

Siamese networks are applied in a wide range of domains. They are used frequently for face recognition, signature verification and image similarity in computer vision area. They have a role in different NLP tasks (from sentence similarity assessment to document comparison). This extends to biometrics where they are used for fingerprint matching or voice recognition.

Training these networks often involves specialized loss functions, such as contrastive loss or triplet loss. These functions are designed to minimize the distance between similar pairs while maximizing the distance between dissimilar pairs, effectively teaching the network to distinguish between matching and non-matching inputs.

Contrastive Loss:

$$L_{\text{contrastive}} = (1 - Y)\frac{1}{2}(D_w)^2 + Y\frac{1}{2}\max(0, m - D_w)^2 \quad (8.1)$$

Where: Y is the binary label (0 for dissimilar pairs, 1 for similar pairs) D_w is the Euclidean distance between the two sample embeddings m is the margin for dissimilar pairs.

Triplet Loss:

$$L_{\text{triplet}} = \max(0, D(a, p) - D(a, n) + \alpha) \quad (8.2)$$

Where: $D(a, p)$ is the distance between the anchor and positive samples $D(a, n)$ is the distance between the anchor and negative samples α is the margin enforced between positive and negative pairs.

8.3 Synthetic EEG signal generation

The method proposed here is entirely based on deep learning. This implies the need of a sufficient number of data samples to train the model. However, a common challenge in biomedicine is the limited data availability, especially in EEG and neurological research due to the inherent difficulty of acquisition procedures. Nevertheless, deep learning networks require enough data to be used at full potential in terms of generalization capability. This constitutes an important difficulty that can limit the scope and power of deep learning algorithms.

To overcome this problem, we propose that the generation of a synthetic dataset of EEG signals that can be coupled at will is possible, taking into account the properties of the EEG.

One of the main characteristics of EEG is its power law, which describes the relationship between the frequency and amplitude of brain activity as follows:

$$P(f) \sim \frac{1}{f^\beta} \quad (8.3)$$

where $P(f)$ is the power at frequency f , and β is the power law exponent. This means that higher frequencies have weaker amplitudes, while lower frequencies have higher amplitudes. Based on the β exponent, different types of noise can be defined, commonly named *colors*. For example, when $\beta = 0$ the noise is called *white noise*, when $\beta = 1$ it is called *pink noise*, and if $\beta = 2$ we called it *brown noise*. Typically, human brain activity exhibits a power law with an exponent $\beta \in [1, 2]$, corresponding to pink and brown noise [88].

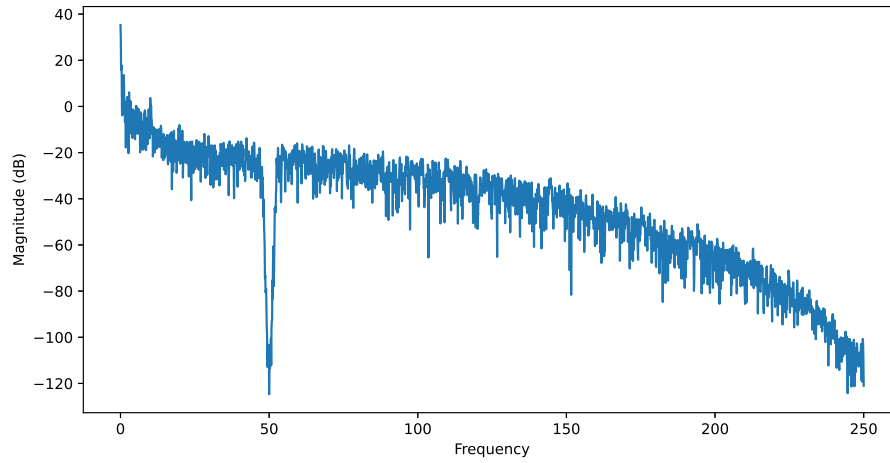


Figure 8.2: Example of EEG power law obtained from real data. The sudden depth in the 50-60 Hz range corresponds to a notch filter applied on the EEG to avoid noise introduced by the power line.

Several works were done using synthetic EEG data with similar methods to construct these signals [46, 100, 114, 147]. The approach used here is summarized in Algorithm 1, Fig. 8.4 and as follows:

1. **Generate colored noise:** The first step consists of generating a signal with characteristics similar to those of EEGs. This is accomplished by create a colored noise where $\beta = 2$ following the relation described in equation 8.3.

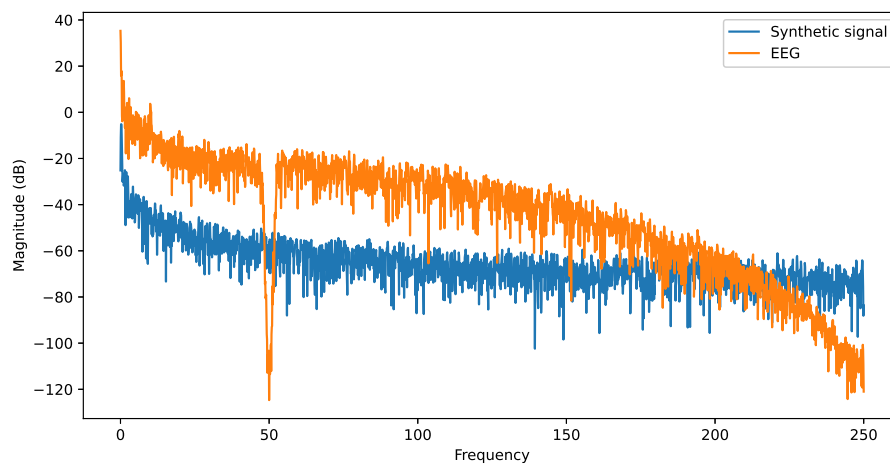


Figure 8.3: Example of EEG signal obtained from real data (orange) along with a synthetic signal (blue).

2. **Filtering using FIR filters:** The generated signals are filtered into specific frequency bands using FIR filters. We have to filter signals that will belong to

low frequencies (range 1-20Hz) and another that will belong to high frequencies (range 50-120Hz).

3. **Amplitude modulation:** This step is crucial to create a coupled pair of signals. This is achieved by applying a Hanning window centered on specific points in the phase of the low-frequency signal. With this, we can control in where phase point we want the coupling to appear. This is performed by obtaining the analytic signal of the phase signal via Hilbert Transform 3.4.1 and obtaining the time point at which the phase is the desired one. Then, the amplitude (high frequency) signal is multiplied at those points by a Hanning window. The duration of the Hanning window is calculated as

$$H = 0.5 * \frac{f_s}{f_p} \quad (8.4)$$

where H is the duration at several points, f_s is the sample rate and f_p is the frequency of the phase signal. The coupling force $\alpha \in (0, 1)$ is controlled by adjusting the height of the Hanning window. Typically, a scale factor between 0 and 1 is used, where 0 indicates no coupling, and 1 indicates maximum coupling. Additional Brownian noise is added to the generated signals to simulate actual EEG recording conditions.

Algorithm 1 summarizes the signal generation method used here, which constitutes an essential stage to generate the deep-learning based models described in Chapter 8. The proposed method allows generating signals in which different bands are coupled by PAC criteria. The specific PAC value has to be computed a posteriori with a specific PAC metric. On the other hand, Fig. 8.4 graphically shows the generation process.

Algorithm 1 Generate PAC Signal

Require: coupling_strength, coupling_phase, phase_frequency, amplitude_frequency, sample_rate

- 1: phase_signal $\leftarrow$ signal filtered to a low frequency (*phase_frequency*) and $\beta = 2$
- 2: amplitude_signal $\leftarrow$ signal filtered to a high frequency (*amplitude_frequency*) and $\beta = 2$
- 3: noise_signal $\leftarrow$ signal filtered to a high frequency (*amplitude_frequency*) and $\beta = 2$
- 4: phase_points $\leftarrow$ instantaneous phase points from *phase_signal* where phase = *coupling_phase*
- 5: HanningWindow $\leftarrow$ Hanning($0.5 * \text{sample_rate} / \text{phase_frequency}$)
- 6: HanningSignal $\leftarrow$ Signal storing all the Hanning Windows with same length as *phase_signal*
- 7: **for** i in phase_points **do**
- 8: HanningSignal[i] = HanningWindow * *coupling_strength*
- 9: **end for**
- 10: adding_signal $\leftarrow$ amplitude_signal * HanningSignal
- 11: amplitude_signal $\leftarrow$ amplitude_signal + adding_signal + noise_signal
- 12: **return** phase_signal, amplitude_signal

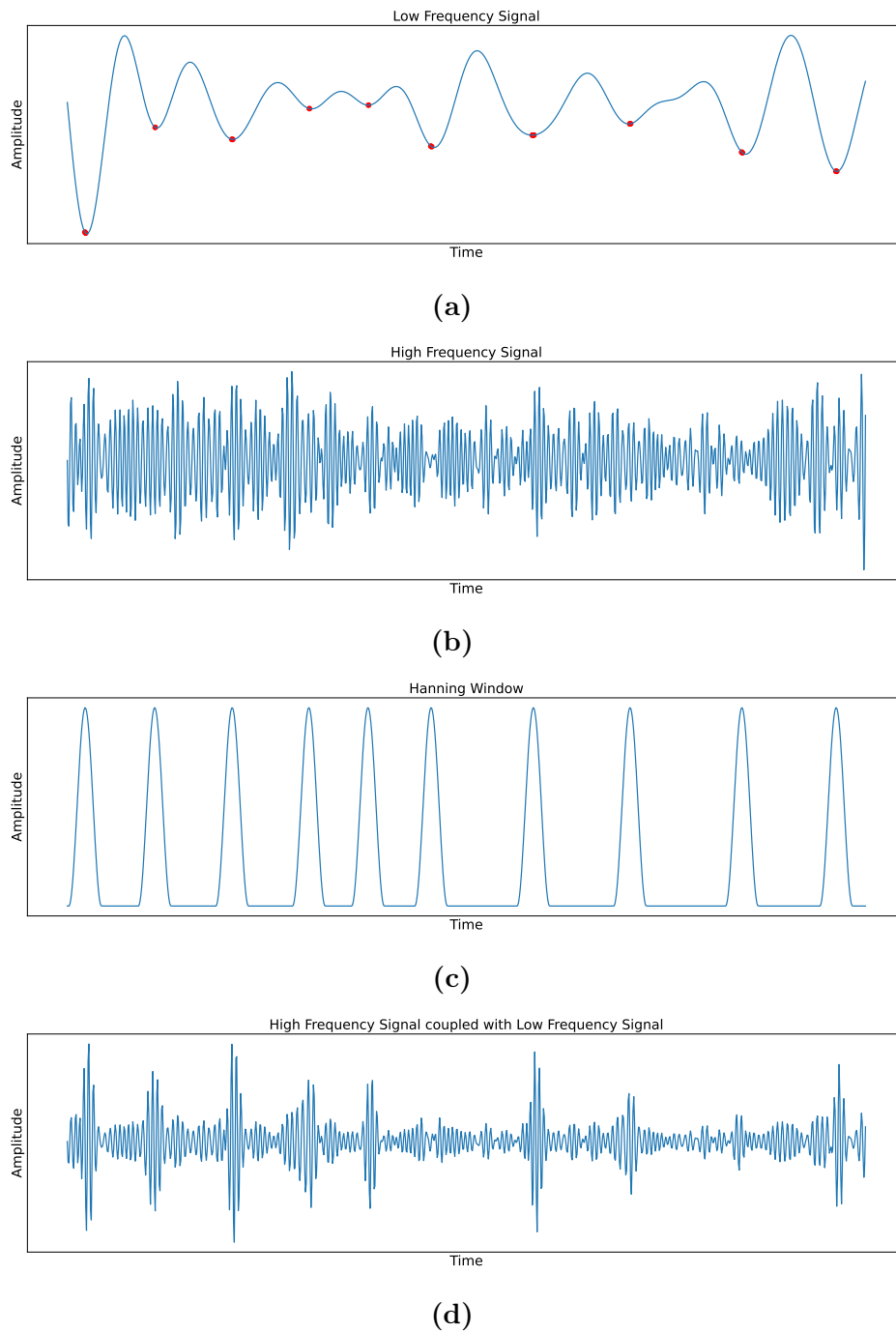


Figure 8.4: Figure a) depicts the low frequency generated signal along with the phase points (red). Figure b) is the high frequency signal that will be treated. Figure c) shows the Hanning windows that will be applied to the previous signal, showing the final result in Figure d). Notice how the amplitude is changed in the specific time points at which the low frequency signal reach the desired phase.

8.4 Siamese Architecture for Lateralization Analysis

The proposed deep learning architecture to analyze the lateralization of dyslexic children is a Siamese network in which the inputs are the raw EEG channels. Each side of the network is responsible for electrodes belonging to one hemisphere of the brain. Since we are relying on PAC computation, a block in the network is responsible to do so. This block needs a high frequency signal together with a low frequency signal to obtain the degree of coupling. The challenge lies in the multiple EEG channels (14 per hemisphere). Creating a network for each channel pair permutation would be prohibitively expensive in terms of both time and computational capacity, which, although technically feasible, is not optimal. Therefore, it is necessary to automatically select the most important channels. To address this, another block is chained just before the PAC Network. This block is responsible for selecting the most critical channels, reducing the total number of channels to one through the *Channel Attention Layer*. A depiction of the network is found in Fig. 8.5.

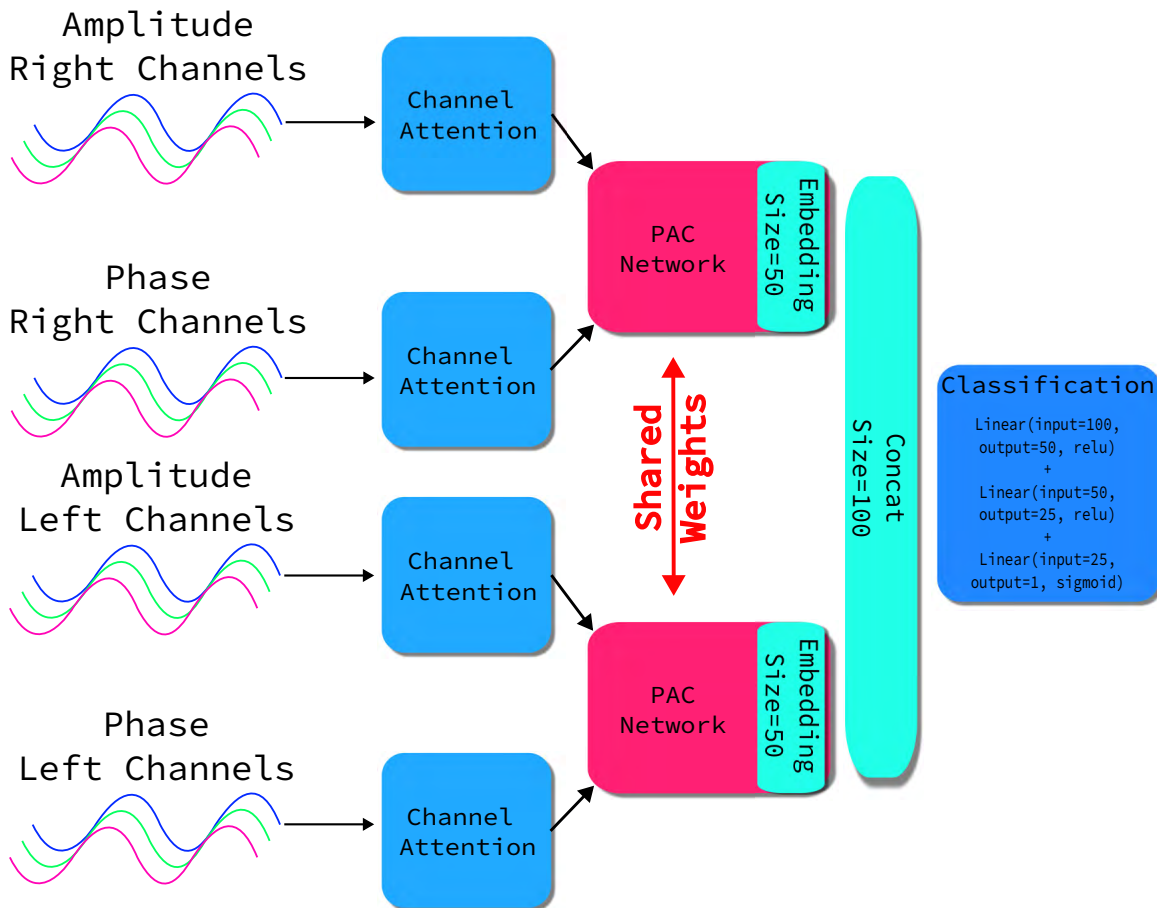


Figure 8.5: Siamese Neural Network. This network is able to process the two brain hemispheres at the same time while selecting the more important channels.

The global operation is to first train a PAC network that could compute the coupling grade between two signals. This network is shared then in a bigger, Siamese network that will perform the final classification using the transfer learning paradigm, in this

case dyslexic vs control group. Prior to that, a selection of channels is performed via a custom layer we called Channel Attention layer. This layer, automatically selects the most significant channels that will be fed to the Siamese network.

In the following sections, we analyze this network layer by layer to better understand the details and operation.

8.4.1 PAC estimation with Deep Learning

The first step is the creation and training of the *PAC Network*. This network is trained to estimate the grade of coupling between two signals, as in Section 3.4.1. This time, however, we are not using typical metrics — like MVL or PLV for example. Although useful in analyzing clear and low-noise signals, the analysis of real EEG data using these measurements faces important obstacles because of the innate complexity and noise that generally exists in real-world EEG signals.

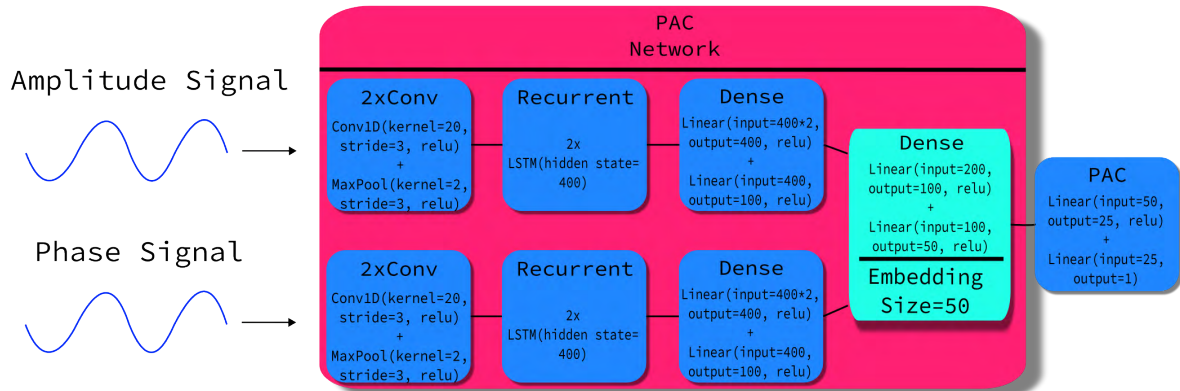


Figure 8.6: Neural Network for PAC calculation. The parameters of this network are frozen once trained, and only the layers until the embeddings are used in the siamese network (red part).

8.4.1.1 Architecture

This network features several layers already seen in section 4.3. As seen in Fig. 8.6, it has two inputs and work in parallel. One branch process on the amplitude signal (the signal with higher frequency) while the other process the phase signal (lower frequency). Both paths are similar and are composed of the same layers.

Convolutional Block

Firstly, a convolutional step is performed. The block features a *Conv1D* layer followed by a *Maxpool* layer. The work of 1D convolution is the same as in original CNN, but instead of working with images (2D matrices) we are working with time series. Fig. 8.7 depicts the process. To do so, filters or kernels are slid through the input signal, doing a dot product between this kernel and a patch of the signal, as follows:

$$y_t = \sum_{k=0}^{K-1} w_k x_{t-k} + b \quad (8.5)$$

where y_t is the output, x_t is the input at position t , w_k is the kernel weight at position k , K is the kernel size, and b is the bias term. The next kernel position can be controlled with a parameter called stride. The *maxpool* operation follows the same principle, but instead of doing a dot product, the maximum values of the extracted patch are selected. This entire process reduces the signal length at the same that can capture patterns retaining as much information as possible [119].

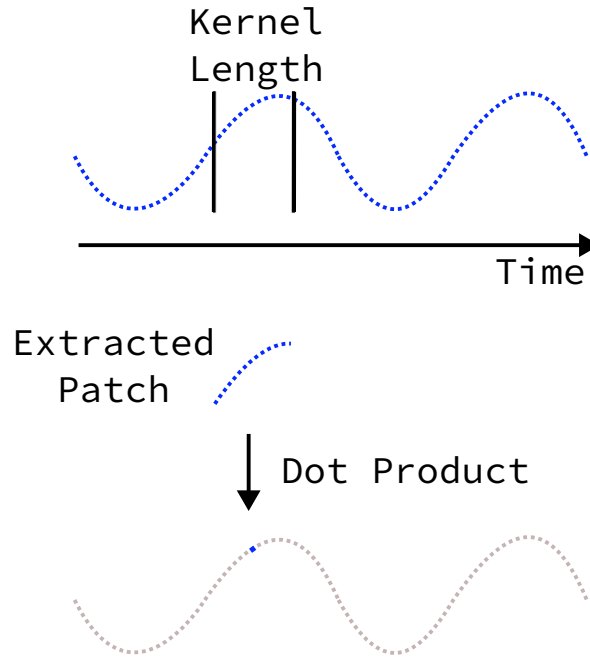


Figure 8.7: In 1D Convolution a segment of the signal is multiplied by a kernel resulting in a single value. This value contains information about the whole segment while reducing the signal length.

For this particular work, we select a *Conv1D* layer with a kernel of 20, stride of 3 followed by a *Maxpool* layer with kernel of 2 and stride of 3. This is repeated twice.

Recurrent Block

The next step on the architecture is a recurrent layer. This layer is a stack of two LSTM blocks with 400 cells each. They were explained in detail in Section 4.3.2.1. These types of networks can be used in different ways, depending on the task in mind. In time forecasting, when we are trying to predict the next token in the sequence, we would pass a time series and recover the last output of the network. This corresponds to $(h_n w)$ in figure 8.8b.

In our case, the task at hand needs to capture the contextual information of a time series, or summarize it. Thus, the part we want to deal with is the *memory*. This is represented by the h_n states of the network, but instead of a timewise way, we need

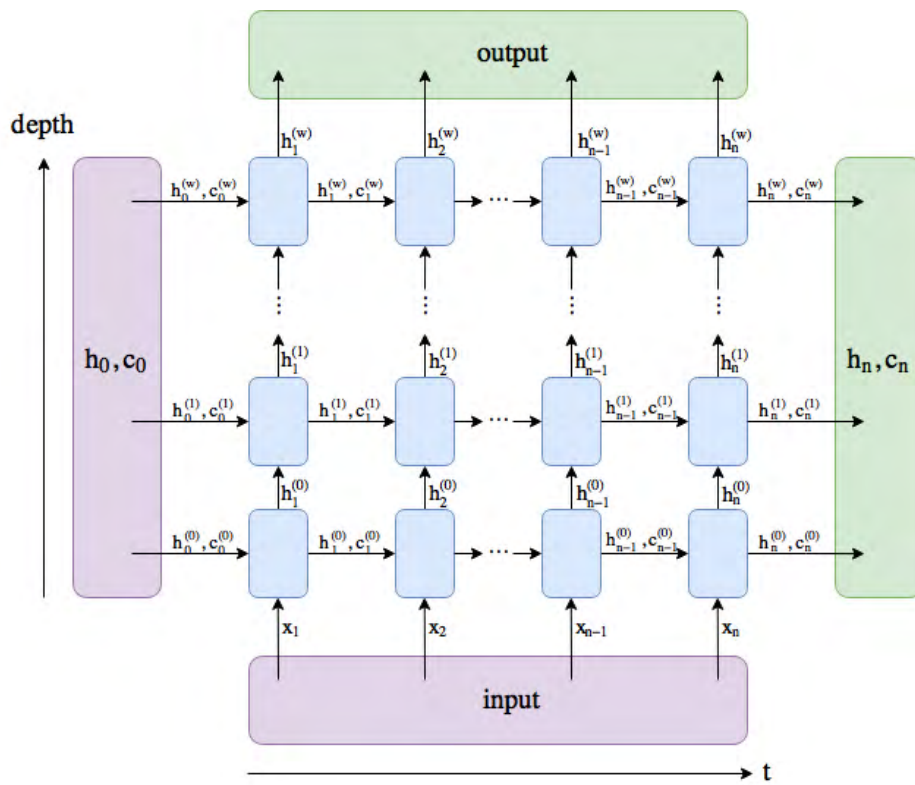
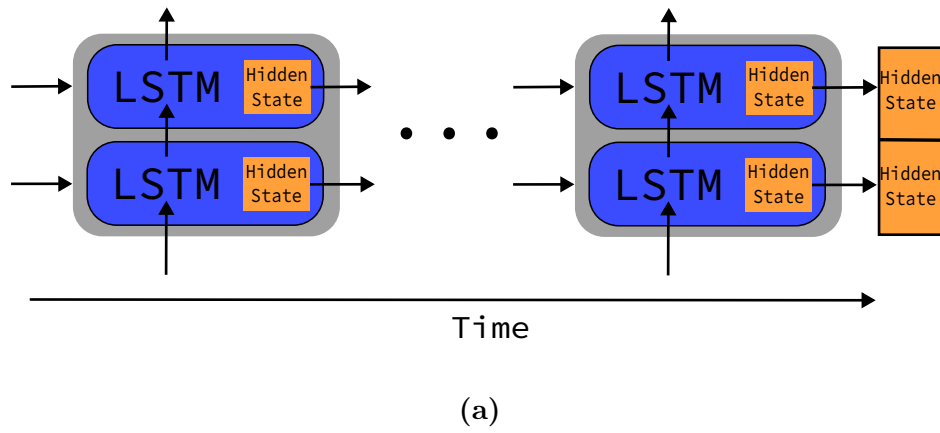


Figure 8.8: Depiction of a working LSTM layer. Each blue rectangle is an LSTM unit. We can see where each output and input are connecting inside the network.

it depthwise at the final step. This would correspond to the green matrix (h_n, c_n) in figure 8.8b. We use this hidden state to represent the EEG sequence, since its context is more crucial for us than forecasting the subsequent part of the signal.

Another characteristic of these networks is that the time series can be fed bidirectionally. This approach allows the network to process information from both past and future time steps, although we would deal with double the computational steps, it may lead to improved performance in many sequence-based tasks.

The final output for this block is a matrix of shape $(400, 2)$. This corresponds to the final output of the 400 cells defined for the layer, but double since we are using bidirectional feeding.

Dense layer and output

Finally, a common dense layer is used. This layer, called linear layer in Fig. 8.6 features fully connected neurons building a MLP (section 4.3). This architecture is common in the final stages of almost every deep network. It is usually in charge of the classification. All previous steps can be seen as an extraction of characteristics that lead to a classification or prediction of some real value.

If we look at Fig. 8.6, we have the branches that now is time to merge again. But prior to the merging, a reduction in the number of features is in place. This prevents a possible overfitting and summarizes the information, at the same time that helps in the next computational steps.

The merging process is accomplished by concatenating the output vectors from the preceding layers and then feeding this combined vector into another fully connected layer. The resulting output of this layer is referred to as embeddings. These embeddings serve as a dense representation of the input data, capturing the essential features and relationships learned by the network.

These embeddings are a condensed version of the input, which consists only meaningful information necessary to solve the task. This form is well suited for many downstream applications like classification or clustering/similarity comparisons. The embedding vectors will also generally be of lower dimensionality than the concatenated feature vectors, providing some form of distributed representation for keys that has reduced computational complexity in later operations.

Finally, another dense layer (**PAC** in Fig. 8.6), will learn if the two fed signals to the network are coupling, giving a real number that we clipped between $(0, 1)$, being 0 no coupling and 1 complete coupling.

The application of deep learning for PAC estimation on EEG data is a significant advancement, and provides a more powerful and efficient solution than traditional methods. This is especially important in the investigation of conditions such as dyslexia, for which revealing interplays among various brain waves are essential to understanding disorder etiology [58, 84].

Data and training process

Due to the nature of the problem, the data available is scarce in order to train a network from scratch. This is a common problem found in almost all medical investigations. To deal with this, sometimes the creation of synthetic data is possible. In

Section 8.3 a method for the creation of pairs of coupled signal is shown. This provides us with a virtually infinite dataset to train the network with.

The steps to train this network are: first, create pairs of coupled signal, each one with a different strength of coupling. The loss chosen to perform the training is MSE:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (8.6)$$

Where:

- n is the number of observations or data points
- y_i is the actual (observed) value for the i -th data point
- $\hat{y}$ is the predicted value for the i -th data point

As the data set is infinite, the training process is carried out until the loss is stabilized close to 0.

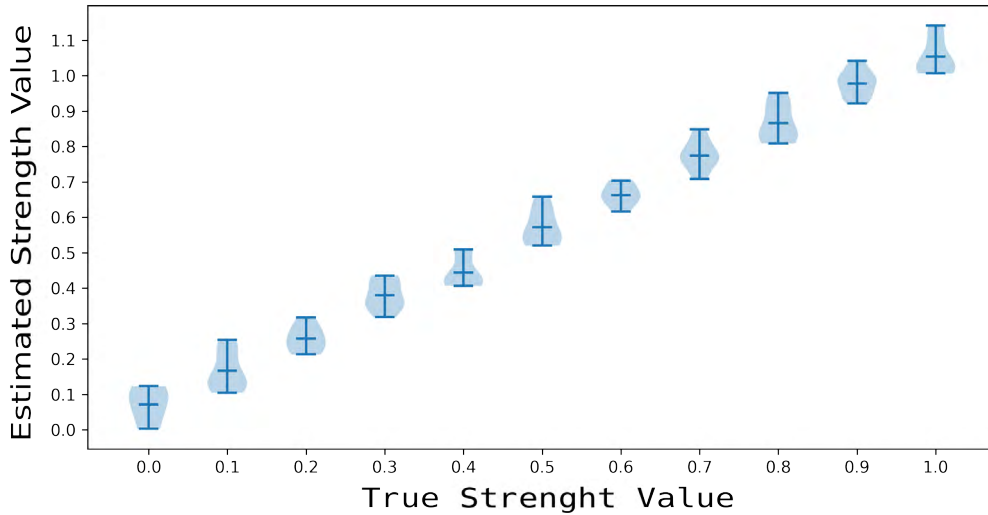


Figure 8.9: Distribution of the strength value α of the synthetic signals vs the output of the network. The network is able to correctly stimate the level of synchrony of the signals. This value is highly correlated with PAC metrics like MVL, PLV. The higher the strength, the higher these metrics.

Once the training phase is ended, in order to validate the network, a new set of synthetic signals are created of several strength coupling, ranging from 0 to 1 in 0.1 steps. The results are shown in Fig. 8.9. The discrepancies are within a tolerable range, therefore indicating that the network has successfully predict coupling forces in new signals. This amount of inaccuracy is acceptable for examining real-life EEG signals. This indicates that once trained, the PAC network can be taken as a reliable tool for analyzing phase-amplitude coupling of more diverse and experimental EEG signals.

To summarize, now the PAC network is trained and ready to be the part of the Siamese Network. Once added to the whole network, its weights will be frozen. This mean that the parameters of this network will not change in the training phase of the Siamese network.

8.4.2 Channel Attention Layer

This module is responsible for simultaneously handling numerous EEG channels. It filters out those channels that are less important to distinguishing between dyslexic and control groups. Its mission is to evaluate multiple inputs and rank them in importance, resulting in a single output channel that encapsulates the most important information, while reducing significantly the complexity of data.

The procedure is summarized in Fig. 8.10. It consists in two convolutional process (CNN + AvgPooling) with a kernel size of 20 for the CNN and a kernel size of 5 for the latter. After these operations, a dense layer with 40 hidden units is applied and a Softmax function is used as output. This produces a normalized weight vector for each EEG channel; the length of this weight vector correspond to the total number of channels. This vector can be seen as an index of importance for all original channels.

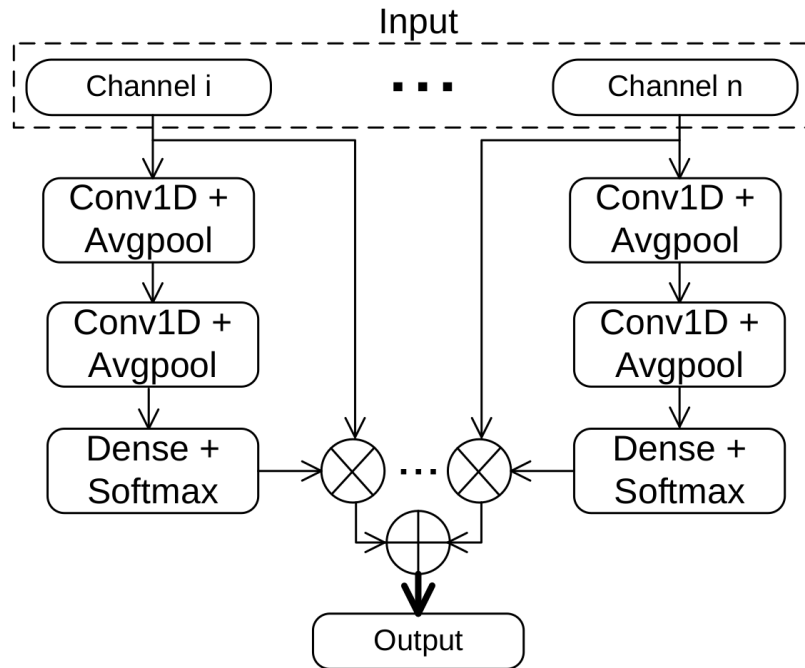


Figure 8.10: The channel attention layer facilitates channel selection by processing input EEG channels through multiple Conv1D and Average Pooling layers, ending in a Softmax layer. This final layer represents each channel's significance, enabling the consolidation of all channels into a single, information-rich channel by multiplying each by its importance and summing the results.

The process is explained in detail in the following equations. Let the input tensor be denoted as $\mathbf{X} \in \mathbb{R}^{C \times S}$ where C is the number of channels, and S is the length of the input sequence.

Layer 1:

1. Convolution Layer 1:

$$\mathbf{Y}_1 = \text{Conv1D}(\mathbf{X}, W_1, b_1), \quad (8.7)$$

where W_1 and b_1 are the weights and biases of the first convolution layer with kernel size $k_1 = 20$.

2. Activation and Average Pooling Layer 1:

$$\mathbf{Y}_2 = \text{AvgPool1D}(\text{ELU}(\mathbf{Y}_1), k_2), \quad (8.8)$$

where $k_2 = 5$ is the kernel size of the average pooling layer.

Layer 2:

3. Convolution Layer 2:

$$\mathbf{Y}_3 = \text{Conv1D}(\mathbf{Y}_2, W_2, b_2), \quad (8.9)$$

where W_2 and b_2 are the weights and biases of the second convolution layer with kernel size $k_1 = 20$.

4. Activation and Average Pooling Layer 2:

$$\mathbf{Y}_4 = \text{AvgPool1D}(\text{ELU}(\mathbf{Y}_3), k_3), \quad (8.10)$$

where $k_3 = 2$ is the kernel size of the average pooling layer.

Dense Layers and Attention Mechanism:

5. Flatten output and fully connected layer 1:

$$\mathbf{Y}_5 = \text{Flatten}(\mathbf{Y}_4), \quad (8.11)$$

6. Dense Layer 1:

$$\mathbf{Y}_6 = W_3 \mathbf{Y}_5 + b_3, \quad (8.12)$$

where W_3 and b_3 are the weights and biases of the first dense layer, transforming into a vector of size 40.

7. Activation and Dense Layer 2:

$$\mathbf{Y}_7 = \text{ELU}(\mathbf{Y}_6), \quad (8.13)$$

$$\mathbf{Y}_8 = W_4 \mathbf{Y}_7 + b_4, \quad (8.14)$$

where W_4 and b_4 are the weights and biases of the second dense layer, transforming into a vector of size M (number of channels).

8. Softmax Activation:

$$\mathbf{A} = \text{Softmax}(\mathbf{Y}_8), \quad (8.15)$$

where $\mathbf{A}$ represents the attention mechanism.

Attention Application:

9. Attention Application:

$$\mathbf{Z} = \sum_{m=1}^M \mathbf{A}_m \cdot \mathbf{X}_m \quad (8.16)$$

where $\mathbf{Z}$ represents the result of applying the attention $\mathbf{A}$ to the original input $\mathbf{X}$ through element-wise multiplication followed by summation over the channel dimension.

Output:

The final outputs of the network are:

$$(\mathbf{A}, \mathbf{Z}) \quad (8.17)$$

This approach offers numerous benefits for EEG analysis. It eliminates the need for computing PAC for every possible channel pair, substantially reducing computational demands. Furthermore, it autonomously identifies the most relevant channels for the classification task, effectively performing built-in feature selection. This aspect is crucial for model explainability, as the channel weights provide insight into which brain regions are most influential in differentiating dyslexic children from the control group.

8.4.3 Siamese network

The Siamese architecture was employed to identify brain lateralization using the already formulated PAC network and attention module. This was designed to detect and compare the disparities in brain activity patterns between the left and right hemispheres during dyslexia [174]. The Siamese network construction starts when the EEG channels are divided into two subsets, one corresponding to the right hemisphere and the other one corresponding to the left one. Then, they are filtered to two specific frequency bands: fast and slow. The fast band is always referred to as the gamma band, whereas the remaining group of bands, alpha, beta, delta, or theta, can stand in for the slow band with initial segregation, making it possible to examine how different brain wave patterns interact thoroughly.

For each hemisphere, two layers of attention are implemented, one dedicated to the fast band and one to the slow band. These layers of attention, as described above, identify and weigh the most relevant channels within each band and hemisphere [4]. The result of each layer of attention is a single, integrative signal that summarizes the information from the corresponding channels. The integrative signals from the attention layers are fed into the PAC network, whose weights have been pre-trained and frozen. This learning transfer approach makes it possible to take advantage of the knowledge acquired by the PAC network in the training phase with synthetic data [1].

The output of the Siamese network is an embedding vector of size 100, 50 of the values corresponding to one of the branches of the network, and the remaining 50 to the other branch. The network reaches its outcome by concatenating the final embeddings from the PAC networks obtained in each hemisphere, thus establishing the classification process [3]. These embeddings contain critical information related to variations in neural activities across two hemispheres, and thus can act as a higher-level feature extractor. At the same time, through this mechanism, the initial synthetic EEG

Table 8.1: Results obtained for different input stimuli and brainwaves. The * denotes no permutation test were done for that band combination.

Stim	Coupling	AUC	Sensitivity	Specificity	Balanced Accuracy	Precision	p-value
4.8	Alpha-Gamma	0.68 ± 0.15	0.63 ± 0.37	0.59 ± 0.48	0.61 ± 0.36	0.52 ± 0.41	*
	Beta-Gamma	0.74 ± 0.19	0.67 ± 0.17	0.60 ± 0.19	0.64 ± 0.14	0.71 ± 0.14	$< 10^{-5}$
	Delta-Gamma	0.61 ± 0.34	0.90 ± 0.13	0.37 ± 0.46	0.64 ± 0.35	0.53 ± 0.29	*
	Theta-Gamma	0.69 ± 0.28	0.80 ± 0.16	0.60 ± 0.49	0.70 ± 0.38	0.72 ± 0.34	*
16	Alpha-Gamma	0.82 ± 0.16	0.83 ± 0.15	0.59 ± 0.18	0.71 ± 0.12	0.69 ± 0.12	$< 10^{-5}$
	Beta-Gamma	0.76 ± 0.31	0.57 ± 0.29	0.97 ± 0.04	0.77 ± 0.20	0.77 ± 0.39	*
	Delta-Gamma	0.70 ± 0.34	0.60 ± 0.33	0.77 ± 0.39	0.69 ± 0.36	0.66 ± 0.42	*
	Theta-Gamma	0.69 ± 0.23	0.60 ± 0.33	0.74 ± 0.37	0.67 ± 0.33	0.55 ± 0.36	*
40	Alpha-Gamma	0.85 ± 0.14	0.77 ± 0.13	0.79 ± 0.19	0.78 ± 0.11	0.83 ± 0.17	$< 10^{-5}$
	Beta-Gamma	0.81 ± 0.14	0.80 ± 0.12	0.75 ± 0.38	0.78 ± 0.25	0.77 ± 0.26	*
	Delta-Gamma	0.72 ± 0.21	0.80 ± 0.16	0.60 ± 0.49	0.70 ± 0.28	0.72 ± 0.34	*
	Theta-Gamma	0.70 ± 0.22	0.67 ± 0.37	0.57 ± 0.47	0.62 ± 0.26	0.48 ± 0.36	*

data will become an informative representation of anything meaningful. After this, these fused embeddings go through a classifier layer where the model uses to identify individuals who have dyslexia and those who do not [125]. The classifier includes three fully connected layers with ReLU activation functions for the first two layers and then uses a sigmoid activation function during its output layer. This architecture ensures that control subjects are distinguished from those suffering from dyslexic conditions by exploiting differences in brain activity patterns captured by the Siamese network preceded by attentive layers [199]. Fig. 8.1 depicts this architecture.

8.5 Results

We assessed the effectiveness of our proposed methodology in three distinct scenarios, examining the impact of auditory stimuli at different frequencies (4, 16, and 40 Hz) on brain activity. Initially, synthetic signals were employed to train the PAC neural network before training the entire network with original data. Subsequently, separate models were trained for each frequency combination. Table 8.1 presents the optimal results for each stimulus-band pair combination.

The first column indicates the stimulus frequency, while the second column shows combinations of various brainwaves: alpha (9-12 Hz), beta (13-35 Hz), delta (1-4 Hz), theta (5-8Hz), and gamma (35-120 Hz). These brainwaves correspond to different cognitive states. In this study, all bands are paired with the gamma band, except for gamma-gamma. For the 4-Hz stimulus, the beta-gamma combination yielded the highest performance (AUC of 0.74). The 16-Hz stimulus showed best performance with the alpha-gamma combination (AUC of 0.82). The 40-Hz scenario achieved a maximum performance of $AUC = 0.85$, also with the alpha-gamma combination.

The results in Table 8.1 were derived through permutation testing, utilizing permuted data to establish the null distribution. Notably, high statistical significance (p -values below 10^{-5}) was observed for these scenarios. Fig. 8.11 illustrates the empirical distribution used to obtain statistical values. The superior performance of results with actual labels (indicated by a vertical red line) confirms that these outcomes were not due to chance.

As described in Section 8.4.2, the attention layer is crucial for model explainability,

enabling the detection of each channel's relative importance in PAC coupling computation. Fig. 8.12 displays the weights assigned by the attention layer to each channel in phase and amplitude components for the most discriminative cases, as per Table 8.1. Significant (p-value < 0.05) weight differences between controls and dyslexics are highlighted to reveal lateralization patterns identified by the network.

The results indicate that right hemisphere channels have greater importance for the 4.8 Hz stimulus, a pattern that persists in the 16 Hz stimulus. The 40 Hz stimulus, however, shows a slight bias towards the left hemisphere. In summary, our procedure effectively demonstrates consistent lateralization patterns with sufficient discriminative power to differentiate between controls and dyslexics.

Table 8.2: Comparison among works implementing different techniques for dyslexia classification. *Data not reported by the source.

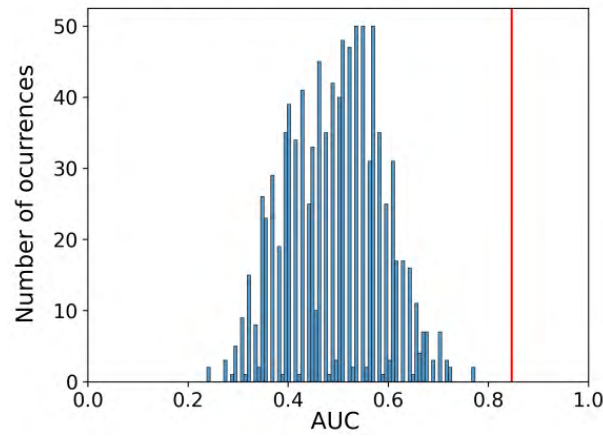
Method	Accuracy	Sensitivity	Specificity	AUC	Data
EEG-ERP / SVM (Memory task) [2]	0.79 ± *	0.89 ± *	0.58 ± *	*	Dyslexic 38 / Controls 19
EEG / SVM (Writing Task) [160]	0.72 ± *	0.76 ± *	0.67 ± *	*	Dyslexic 17 / Controls 15
EEG / SVM (Typing Task) [160]	0.78 ± *	0.88 ± *	0.67 ± *	*	Dyslexic 17 / Controls 15
EEG / SVM (Lexical decision) [60]	0.79 ± *	0.79 ± *	0.78 ± *	*	Dyslexic 20 / Controls 20
MRI / SVM [192]	0.80 ± *	0.82 ± *	0.78 ± *	*	Dyslexic 22 / Controls 15
fMRI/ CNN [213]	0.73 ± *	0.75 ± *	0.71 ± *	*	Dyslexic 19 / Controls 19
EEG / our method, (4.8 Hz, Beta-100 Hz)	0.71 ± 0.14	0.67 ± 0.17	0.60 ± 0.19	0.74 ± 0.19	Dyslexic 16 / Controls 32
EEG / our method, (16 Hz, Alpha-100 Hz)	0.69 ± 0.12	0.83 ± 0.15	0.59 ± 0.18	0.82 ± 0.16	Dyslexic 16 / Controls 32
EEG / our method, (40 Hz, Alpha-90 Hz)	0.77 ± 0.13	0.79 ± 0.19	0.85 ± 0.14	0.83 ± 0.17	Dyslexic 16 / Controls 32

8.6 Discussion

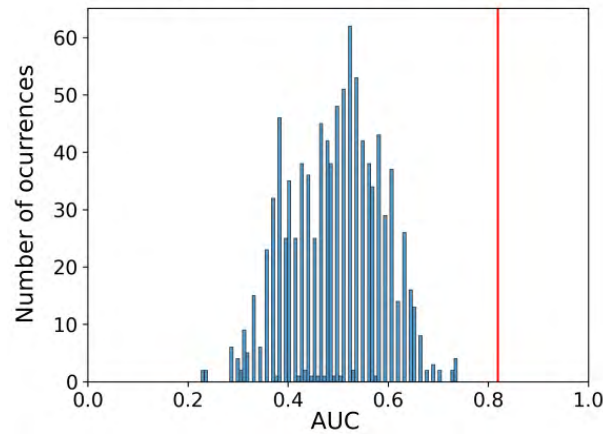
In this work, we present a novel approach for studying brain activity based on cerebral lateralization and signal coupling for the identification of differential patterns in functional activity by means of PAC, which allows understanding the interactions between brain rhythms [21]. In the context of EEG analysis, there are two main analyses that are usually performed: first, the study of the coupling between different frequency bands within the same electrode (local CFC). Second, the study of the coupling between different electrodes by examining how the phase of a low-frequency band signal in one electrode affects the amplitude of a high-frequency band signal on another electrode [23]. This second approach has been successfully used to study the interactions between different regions of the brain. Both methods offer helpful information regarding neural coordination over various periods and spatial locations, providing a greater understanding of how brain information processing works [157].

This work proposes a deep learning framework for PAC estimation, along with a method to overcome the usual drawbacks in EEG processing as well as a symmetric architecture, whose branches are fed with EEG channels (divided by hemisphere) to evaluate differences between them. Our work is based on several studies that support the theory of hemispheric differences in the coordination of brain function in dyslexic individuals [174, 206]. Finally, the classification decision is obtained and compared with the actual diagnosis (control/dyslexic) in order to evaluate the performance of the proposed architecture.

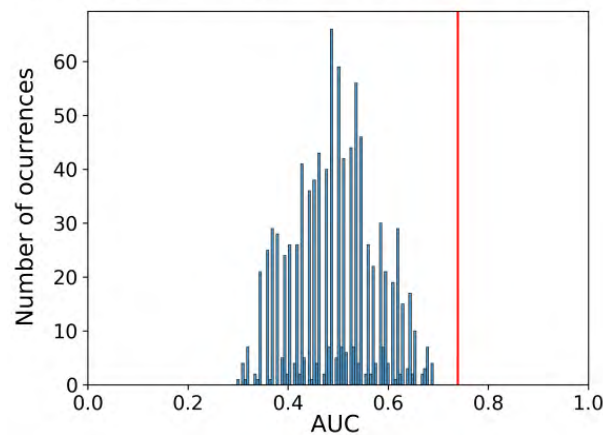
One of the most important contributions of this work is how the EEG channels that contain relevant information are identified. A crucial aspect in the training of

40 Hz, Alpha-Gamma PAC. AUC=0.82

(a)

16 Hz, Alpha-Gamma PAC. AUC=0.82

(b)

4.8 Hz, Beta-Gamma PAC. AUC=0.74

(c)

Figure 8.11: Best permutation results a) 4.8 Hz, b) 16 Hz and c) 40 Hz) stimuli. Null distributions obtained over 5000 permutations are shown in blue. The red line represents the AUC obtained when using the actual labels.

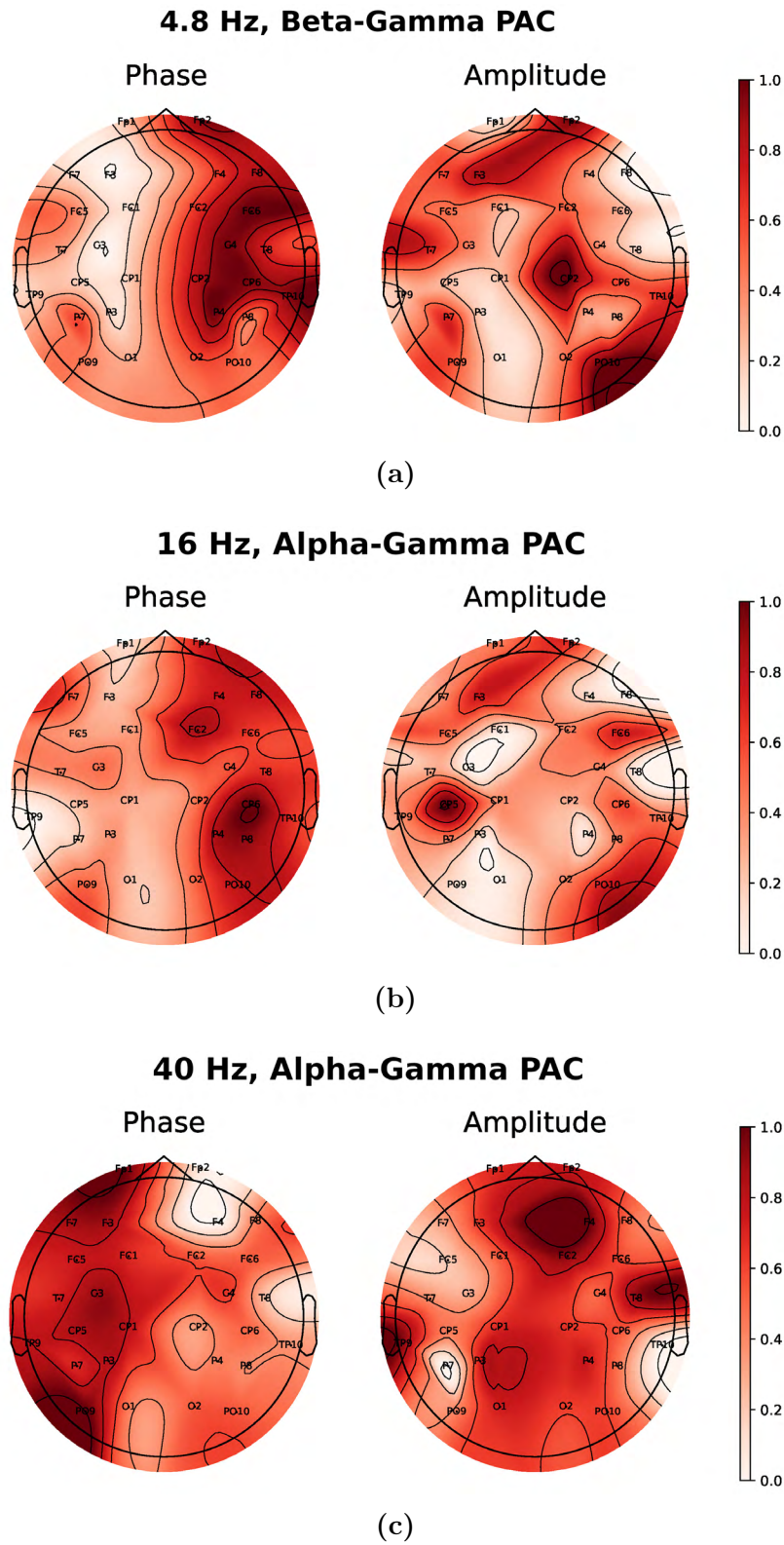


Figure 8.12: Topographical maps showing the significant differences ($p < 0.05$) between controls and dyslexics in channel importance computed by the attention layer for the a) 4.8 Hz, b) 16 Hz and c) 40 Hz stimuli used in this work.

an AI model is to isolate information from noise so that the model training is guided by real associations between labels and features. Most previous studies address this issue by applying the method on every pair of channels available, so that for every pairwise permutation only those with the highest performance are selected for the subsequent classification. This alternative entails a high computational cost, which is suboptimal from a computation perspective. The solution provided in this work (the use of an attentional block) overcomes the common alternatives by allowing the network to learn through a previous selection of the relevant channels, considerably reducing the computational burden. Thus, all channels are transformed into a single wave that carries all the relevant information, so that the contribution of each channel is directly proportional to the information provided individually.

It is worth mentioning that the results obtained are not derived from the brain as a whole but from the asymmetry between left and right hemispheres. Although results in EEG studies are based on the electrical activity recorded in a number of electrodes, conclusions in terms of spatial information are usually drawn at a regional level, where the region refers to an area previously identified (e.g., in an anatomical atlas). Despite this localization perspective of trying to identify where information is contained, the analysis of the different channels is simultaneously performed, so that all of them are considered as a common pattern. In the study of dyslexia, there are certain studies that have found differences in how the hemispheres work together in dyslexic subjects [174, 206]. These findings manifest the need to evaluate each hemisphere in a separate way, in order to study if they work differently compared to control subjects. The Siamese network proposed in this work was designed to account for this need, and the high classification performance achieved demonstrates the suitability of this architecture in this kind of analysis. In fact, it is extraordinarily relevant for the study of the brain basis of dyslexia, and paves the way for the development of new methodologies that exploit the role of the cerebral lateralization in this pathology.

It is important to highlight how we addressed in this work one of the main problems in the training of a deep neural network: the lack of a large training dataset. This condition is not always met in EEG studies, given the difficulty of the acquisition process and the high cost of the necessary equipment. Our method addresses this issue by generating synthetic data that simulate the behavior of the PAC. We pretrained a NN with this synthetic dataset prior to using our original dataset. After that, the classification results were validated through a permutation scheme, leading to very low p -values, which manifests the robustness of our classification framework and the validity of the results.

The results demonstrate the importance of the alpha/gamma band, where the optimum performance metrics are achieved. These observations corroborate the conclusions of other studies in which alpha waves are often altered in individuals with dyslexia [97, 211]. In fact, previous research found differences in the distribution of alpha power modulation among dyslexics and individuals with typical reading capabilities. Notably, alpha power is connected with not just spatial attention, but also the temporal organization of stimuli. One study, deeply explored auditory attention, discovering a synchronicity between alpha power lateralization in the brain and speech rate. This connection noticeably impacted stream-specific speech comprehension [211]. These findings point towards the substantial function of alpha oscillations in processing

and assimilating auditory information, highlighting its significance in language processing and reading concerning dyslexia. Thus, results obtained in this study are clearly connected with literature, showing that the framework proposed addresses important issues about the processing and study of brain EEG signals.

Other key points of this work are the spatial information provided by our method via the attentional block. This allows us to quantify the relevance of each individual electrode according to its weight, as a measure of how informative the brain region where it is located is, offering a crucial insight into what is happening in the brain of a dyslexic patient, which is highly valuable for clinicians and neuroscientists. Our results clearly demonstrate the anomalies in lateralization in dyslexic patients, as well as the asymmetry between hemispheres in different frequency bands [175]. Besides, the most relevant channels marked by our method clearly match the functional brain networks found in [67], where interconnected modules of brain regions were identified as indicative of segregated neural processing. In fact, the existence of small-worldness property as a topology of brain complex networks is compatible with our results, as well as its existence in multiple frequency bands [186]. Moreover, our method detected lateralization differences in the frontal lobe. This region plays a crucial role in speech production and comprehension [212], and thus is highly informative in the diagnosis of dyslexia. Finally, the most relevant channels identified by our method also overlap with the primary auditory cortex. An interaction between hemispheres in this region is associated with phonological processing, as reported in previous studies [69, 167].

Table 8.2 compares various studies using EEG with different stimuli as well as imaging data to compute differential patterns to classify control and dyslexic subjects. Our method yields classification results comparable to those obtained in other works, with a similar sample size. A key feature of our study is the exclusive use of passive auditory stimuli, which sets it apart from other approaches. This framework minimizes bias from mood or other factors affected by voluntary responses, unlike studies such as [2, 60, 160] that employ reading or writing tests. These methods used more complex experimental setups and require controlling for factors like the emotional and relaxation states. [171] acquired brain activity in resting state, which most closely resembles our experimental setup. However, our methodology can be categorized as a semi-resting state, as it incorporates auditory stimuli without requiring subject intervention. Other studies, such as [192, 213], focused on detecting structural and functional patterns using structural MRI and functional MRI. It is important to note that the primary aim of our work was not to achieve high classification rates but to provide a method for exploring differential lateralization patterns associated with language processing in dyslexia. Furthermore, the proposed method has potential applications in studying other neurological conditions or disorders beyond the specific case of language processing presented here.

Conclusions

This thesis presents different methods to study the brain connectivity by means of synchronization patterns. Methods proposed here combine signal processing techniques and artificial intelligence under the umbrella of explainability. Specifically, developed methods are based on cross-frequency coupling techniques focused to study the phenomenon of neural entrainment, aiming to compute biomarkers for differential diagnosis in language disorders and, particularly, [Developmental Dyslexia \(DD\)](#). As a summary, we provide the following conclusions of the contributions of this thesis:

- We have proposed different methods to study brain connectivity patterns based on cross-frequency coupling. Specifically, PAC methods are used. These methods are designed to be robust enough to avoid spurious values such as those caused by volume conduction.
- By employing amplitude-modulated white noise at speech-relevant frequencies, this study minimized biases introduced by language-specific tasks. This semi-resting-state paradigm provided robust data for studying the neural mechanisms of dyslexia across distinct EEG bands.
- In chapter [7](#), the method presented measures the entrainment between the EEG signals and auditory stimuli. In this study, five PAC statistics were used. Moreover, a feature aggregation method is proposed, and PAC values were aggregated within six frequency bands: Delta, Theta, Alpha, Beta, Gamma, and Epsilon (high Gamma). Aggregation involved counting the number of PAC values above a significance threshold. Statistical significance tests including significance corrections allowed to obtain results suggesting that the analysis of entrainment to rhythmic patterns in speech could lead to a fast, objective, and interpretable method for diagnosing dyslexia. This could pave the way for early interventions and reduce the impact of dyslexia on educational outcomes.
- The anomaly detection approach proposed in Chapter [6](#), which incorporates [Self-organizing map \(SOM\)](#) and Bayesian classification, identified distinct connectivity patterns in dyslexic brains. These patterns, based on phase connectivity using Pearson's circular correlation, have demonstrated their usefulness in differential diagnosis, yielding sensitivity and specificity values exceeding 90%, showcasing the potential of this method as a computer-aided diagnosis tool. Moreover, deviation from the normality is detected by difference in the quantization error, when the SOM is trained with normal patterns. The inherent variability present in dyslexic subjects, along with the unbalanced datasets and limited sample size

(as is common in biomedicine), makes this anomaly detection approach suitable for identifying experimental subjects as deviations from typical control behavior.

- The use of Deep Learning for processing EEG signals has also been explored in this thesis. In Chapter 8 we propose an architecture to assess the cerebral lateralization from EEG signals, according to the inter-channel synchronization. This is accomplished by designing a neural architecture combining convolutional layers and LSTM blocks to process EEG time series, which is trained to replicate PAC values from raw signals. In addition, a method was introduced to generate synthetic data that resembles the statistical properties of EEG for pre-training. The pre-trained network is then transferred to a symmetric (Siamese) architecture which will be responsible for detecting differences between hemispheres. This methodology overcomes the problem of limited sample size, which normally hinders the use of deep learning architectures with EEG data. Thus, the transfer learning paradigm and the construction of a Siamese network effectively quantify interhemispheric differences, providing a reliable measure of cerebral lateralization. The proposed approach demonstrated high classification performance, achieving up to 90% accuracy for dyslexic vs. control subjects. It revealed significant PAC strength anomalies between regions associated with language processing. Differential lateralization patterns were observed, with dyslexic brains showing atypical right-hemisphere activation compensating for deficits in left-hemisphere language processing areas. These findings align with theories of altered hemispheric dominance in DD.
- Also, for the sake of explainability, two main contributions have been made:
 - In the development of the deep learning architecture previously explained and presented in chapter 8, we present an attention mechanism within deep learning frameworks facilitated explainability and computational efficiency by focusing on critical EEG channels. This layer not only reduced the computational burden, but also enhanced model interpretability, enabling targeted insights into language-related brain activity. Thus, this layer enables to explore the channel selection performed by the network during the learning process and to observe the weights assigned to each channel, providing important clinical information about the brain areas in which differential patterns are observed. Furthermore, this method reveals how lateralization patterns arise depending on the frequency bands involved in the synchronization processes.
 - A projection method based on Lambert azimuthal projection is presented in chapter 7 in order to obtain a 2D representation of electrode positions, keeping the proportions and inter-electrode distances according to the 10-20 international system. Moreover, the projection of an MRI template is used to provide a representation in which anatomical and functional areas can be identified and linked to the electrodes. This provides a clear view of the areas involved in connectivity patterns.

Other Implications

- *Clinical Applications:* The methods proposed in this thesis pave the way for scalable, non-invasive, and objective diagnostic tools for dyslexia. These approaches can supplement traditional behavioral tests, offering clinicians a more comprehensive understanding of the disorder's neural underpinnings.
- *Educational and Therapeutic Interventions:* Insights into altered auditory processing and connectivity in dyslexia can inform tailored interventions. For instance, training programs could target specific neural rhythms to enhance phoneme segmentation and language comprehension.
- *Generalization and Future Work:* While the results are promising, they are currently specific to Spanish-speaking children. Future studies should expand these methodologies to diverse linguistic and demographic groups to validate generalizability. Moreover, high-density EEG could uncover additional fine-grained interactions between brain regions, offering further exploratory insights.

All the methods developed in this thesis demonstrates the potential of integrating signal processing methods with cutting-edge AI methodologies to tackle complex neurological conditions like dyslexia. The combination of interpretable machine learning models, efficient feature extraction, and a focus on brain connectivity offers a robust framework that with diagnostic applications but also providing knowledge about neurobiological insights.

References

- [1] S. U. Amin, M. Alsulaiman, G. Muhammad, M. A. Bencherif, and M. S. Hossain. Multilevel weighted feature fusion using convolutional neural networks for eeg motor imagery classification. *Ieee Access*, 7:18940–18950, 2019.
- [2] I. I. Andreadis, G. A. Giannakakis, C. Papageorgiou, and K. S. Nikita. Detecting complexity abnormalities in dyslexia measuring approximate entropy of electroencephalographic signals. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6292–6295. IEEE, 2009.
- [3] J. E. Arco, A. Ortiz, D. Castillo-Barnes, J. M. Górriz, and J. Ramírez. Ensembling shallow siamese architectures to assess functional asymmetry in alzheimer’s disease progression. *Applied Soft Computing*, 134:109991, 2023. ISSN 1568-4946. doi: <https://doi.org/10.1016/j.asoc.2023.109991>. URL <https://www.sciencedirect.com/science/article/pii/S1568494623000091>.
- [4] J. E. Arco, A. Ortiz, N. J. Gallego-Molina, J. M. Górriz, and J. Ramírez. Enhancing multimodal patterns in neuroimaging by siamese neural networks with self-attention mechanism. *International Journal of Neural Systems*, 33(04):2350019, 2023. doi: 10.1142/S0129065723500193.
- [5] P. Arpaia, E. D. Benedetto, and L. Duraccio. Design, implementation, and metrological characterization of a wearable, integrated ar-bci hands-free system for health 4.0 monitoring. *Measurement*, 177:109280, 2021. doi: 10.1016/j.measurement.2021.109280.
- [6] A. C. Atencio, T. F. B. Filho, A. Ferreira, and A. B. Benevides. Evaluation of erd/ers caused by unpleasant sounds to be applied in bcis. In *2013 ISSNIP Biosignals and Biorobotics Conference: Biosignals and Robotics for Better and Safer Living (BRC)*, pages 1–5, 2013. (see p. 26).
- [7] A. Bahramisharif, M. A. van Gerven, E. J. Aarnoutse, M. R. Mercier, T. H. Schwartz, J. J. Foxe, N. F. Ramsey, and O. Jensen. Propagating neocortical gamma bursts are coordinated by traveling alpha waves. *Journal of Neuroscience*, 33(48):18849–18854, 2013.
- [8] A. Bahramisharif, A. Mazaheri, N. Levar, P. R. Schuurman, M. Figuee, and D. Denys. Deep brain stimulation diminishes cross-frequency coupling in obsessive-compulsive disorder. *Biological psychiatry*, 80(7):e57–e58, 2016.

- [9] S. R. Benbadis. A spell in the epilepsy clinic and a history of “chronic pain” or “fibromyalgia” independently predict a diagnosis of psychogenic seizures. *Epilepsy & Behavior*, 6(2):264–265, 2005.
- [10] A. Benítez-Burraco and E. Murphy. Why brain oscillations are improving our understanding of language. *Frontiers in Behavioral Neuroscience*, page 190, 2019.
- [11] A. Betti, M. Tucci, E. Crisostomi, A. Piazzzi, S. Barmada, and D. Thomopoulos. Fault prediction and early-detection in large pv power plants based on self-organizing maps. *Sensors*, 21(5):1687, 2021.
- [12] S. Bhattacharyya, A. Khasnobish, A. Konar, D. N. Tibarewala, and A. K. Nagar. Performance analysis of left/right hand movement classification from eeg signal by intelligent algorithms. In *2011 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB)*, pages 1–8, 2011. doi: 10.1109/CCMB.2011.5952111.
- [13] B. Boashash. Chapter 16 - time-frequency methodologies in neurosciences. In *Time-Frequency Signal Analysis and Processing (Second Edition)*, pages 915–966. Academic Press, Oxford, second edition edition, 2016.
- [14] U. Braun, S. Muldoon, and D. Bassett. On human brain networks in health and disease. *eLS*, page 10, 02 2015.
- [15] R. P. Brenner. The interpretation of the eeg in stupor and coma. *The neurologist*, 11(5):271–284, 2005.
- [16] G. E. Bruder, J. W. Stewart, and P. J. McGrath. Right brain, left brain in depressive disorders: clinical and theoretical implications of behavioral, electrophysiological and neuroimaging findings. *Neuroscience & Biobehavioral Reviews*, 78:178–191, 2017.
- [17] A. Burgess. On the interpretation of synchronization in eeg hyperscanning studies: A cautionary note. *Frontiers in human neuroscience*, 7:881, 12 2013.
- [18] G. Buzsaki. *Rhythms of the Brain*. Oxford university press, 2006.
- [19] S. C. Health implications of binaural auditory beat (bab) its perils and applications. 2021. doi: 10.31234/osf.io/847gp.
- [20] W. Cai, D. Zhao, M. Zhang, Y. Xu, and Z. Li. Improved self-organizing map-based unsupervised learning algorithm for sitting posture recognition system. *Sensors*, 21(18):6246, 2021.
- [21] R. T. Canolty and R. T. Knight. The functional role of cross-frequency coupling. *Trends in cognitive sciences*, 14(11):506–515, 2010.
- [22] R. T. Canolty, E. Edwards, S. S. Dalal, M. Soltani, S. S. Nagarajan, H. E. Kirsch, M. S. Berger, N. M. Barbaro, and R. T. Knight. High gamma power is phase-locked to theta oscillations in human neocortex. *science*, 313(5793):1626–1628, 2006.

- [23] R. T. Canolty, C. F. Cadieu, K. Koepsell, R. T. Knight, and J. M. Carmena. Multivariate phase–amplitude cross-frequency coupling in neurophysiological signals. *IEEE Transactions on Biomedical Engineering*, 59(1):8–11, 2011.
- [24] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3):1–27, Apr. 2011. ISSN 2157-6904. doi: 10.1145/1961189.1961199.
- [25] M. Chaturvedi, J. G. Bogaarts, V. V. Kozak (Cozac), F. Hatz, U. Gschwandtner, A. Meyer, P. Fuhr, and V. Roth. Phase lag index and spectral power as qeeg features for identification of patients with mild cognitive impairment in parkinson’s disease. *Clinical Neurophysiology*, 130(10):1937–1944, 2019.
- [26] J. Chen, P. Zhang, Z. Mao, Y. Huang, D. Jiang, and Y. Zhang. Accurate eeg-based emotion recognition on combined features using deep convolutional neural networks. *IEEE Access*, 7:44317–44328, 2019.
- [27] Q. Chen, E. Middleton, and D. Mirman. Words fail: Lesion-symptom mapping of errors of omission in post-stroke aphasia. *Journal of Neuropsychology*, 13, 02 2018. doi: 10.1111/jnp.12148.
- [28] T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.
- [29] Z. Cheng, J. Wen, G. Huang, and J. Yan. Applications of artificial intelligence in nuclear medicine image generation. *Quantitative Imaging in Medicine and Surgery*, 11:2792–2822, 2021. doi: 10.21037/qims-20-1078.
- [30] W. Chiu, C. Su, C. Fan, C. Chen, and K. Yeh. Authentication with what you see and remember in the internet of things. *Symmetry*, 10:537, 2018. doi: 10.3390/sym10110537.
- [31] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014.
- [32] K. K. H. Chung, J. C. M. Lo, and C. McBride. Cognitive-linguistic profiles of chinese typical-functioning adolescent dyslexics and high-functioning dyslexics. *Annals of Dyslexia*, 68:229–250, 2018. doi: 10.1007/s11881-018-0165-y.
- [33] K. Clark, T. Helland, K. Specht, K. L. Narr, F. R. Manis, A. W. Toga, and K. Hugdahl. Neuroanatomical precursors of dyslexia identified from pre-reading through to age 11. *Brain*, 137:3136–3141, 2014. doi: 10.1093/brain/awu229.
- [34] M. X. Cohen. *Analyzing Neural Time Series Data: Theory and Practice*. 2014.
- [35] L. J. Colling, H. L. Noble, and U. Goswami. Neural entrainment and sensorimotor synchronization to the beat in children with developmental dyslexia: An eeg study. *Frontiers in Neuroscience*, 11:360, 2017.

- [36] L. J. Colling, H. L. Noble, and U. Goswami. Neural entrainment and sensorimotor synchronization to the beat in children with developmental dyslexia: An eeg study. *Frontiers in Neuroscience*, 11:360, 2017.
- [37] C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20:273–297, 1995.
- [38] R. J. Croft and R. J. Barry. Removal of ocular artifact from the eeg: a review. *Clinical Neurophysiology*, 111(1):5–9, 2000. doi: 10.1016/S1388-2457(99)00221-6. URL <https://www.sciencedirect.com/science/article/pii/S1388245799002216>.
- [39] R. J. Croft and R. J. Barry. Removal of ocular artifact from the eeg: a review. *Neurophysiologie Clinique/Clinical Neurophysiology*, 30(1):5–19, 2000.
- [40] M. Čukić, M. Stokić, S. Simic, and D. Pokrajac. The successful discrimination of depression from eeg could be attributed to proper feature extraction and not to a particular classification method. *Cognitive Neurodynamics*, 14:443–455, 2020.
- [41] T. Das. Consciousness created by brain waves. *J of Health Science*, 7, 2019. doi: 10.17265/2328-7136/2019.01.002.
- [42] Z. Davey, P. Gupta, D. Li, R. Nayak, and P. Govindarajan. Rapid response eeg: current state and future directions. *Current Neurology and Neuroscience Reports*, 22:839–846, 2022. doi: 10.1007/s11910-022-01243-1.
- [43] C. De Hemptinne, E. S. Ryapolova-Webb, E. L. Air, P. A. Garcia, K. J. Miller, J. G. Ojemann, J. L. Ostrem, N. B. Galifianakis, and P. A. Starr. Exaggerated phase–amplitude coupling in the primary motor cortex in parkinson disease. *Proceedings of the national academy of sciences*, 110(12):4780–4785, 2013.
- [44] E. De la Hoz, E. De La Hoz, A. Ortiz, J. Ortega, and B. Prieto. Pca filtering and probabilistic som for network intrusion detection. *Neurocomputing*, 164:71–81, 2015.
- [45] A. De Vos, S. Vanvooren, J. Vanderauwera, P. Ghesquière, and J. Wouters. Atypical neural synchronization to speech envelope modulations in dyslexia. *Brain and language*, 164:106–117, 2017.
- [46] A. De Vos, S. Vanvooren, J. Vanderauwera, P. Ghesquiere, and J. Wouters. A longitudinal study investigating neural processing of speech envelope modulation rates in children with (a family risk for) dyslexia. *Cortex*, 93:206–219, 05 2017.
- [47] X. Deng, B. Zhang, N. Yu, and K. Liu, K. and Sun. Advanced tsgl-eegnet for motor imagery eeg-based brain-computer interfaces. *IEEE Access*, 9:25118–25130, 2021. doi: 10.1109/access.2021.3056088.
- [48] G. Di Liberto, V. Peter, M. Kalashnikova, U. Goswami, D. Burnham, and E. Lalor. Atypical cortical entrainment to speech in the right hemisphere underpins phonemic deficits in dyslexia. *NeuroImage*, 175:70–79, 03 2018.

-
- [49] E. Dikici, M. Bigelow, L. Prevedello, R. White, and B. Erdal. Integrating ai into radiology workflow: levels of research, production, and feedback maturity. *Journal of Medical Imaging*, 7:1, 2020. doi: 10.1117/1.jmi.7.1.016502.
 - [50] S. I. Dimitriadis, P. G. Simos, J. M. Fletcher, and A. C. Papanicolaou. Aberrant resting-state functional brain networks in dyslexia: Symbolic mutual information analysis of neuromagnetic signals. *International Journal of Psychophysiology*, 126:20–29, 2018.
 - [51] S. I. Dimitriadis, P. G. Simos, J. M. Fletcher, and A. C. Papanicolaou. Aberrant resting-state functional brain networks in dyslexia: Symbolic mutual information analysis of neuromagnetic signals. *International Journal of Psychophysiology*, 126:20–29, 2018.
 - [52] N. Ding and J. Z. Simon. Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences*, 109(29):11854–11859, 2012.
 - [53] R. P. Duin. Classifiers in almost empty spaces. In *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, volume 2, pages 1–7. IEEE, IEEE, 2000.
 - [54] T. Ergüzel, G. Sayar, and N. Tarhan. Artificial intelligence approach to classify unipolar and bipolar depressive disorders. *Neural Computing and Applications*, 27:1607–1616, 2015. doi: 10.1007/s00521-015-1959-z.
 - [55] H. Ezawa, R. Yoneyama, S. Kandatsu, K. Yoshikawa, H. Tsujii, and K. Harigaya. Introduction of autopsy imaging redefines the concept of autopsy: 37 cases of clinical experience. *Pathology International*, 53:865–873, 2003. doi: 10.1046/j.1440-1827.2003.01573.x.
 - [56] M. Fatourehchi, A. Bashashati, R. K. Ward, and G. E. Birch. Emg and eeg artifacts in brain computer interface systems: A survey. *Clinical neurophysiology*, 118(3):480–494, 2007.
 - [57] S. Flanagan and U. Goswami. The role of phase synchronisation between low frequency amplitude modulations in child phonology and morphology speech tasks. *The Journal of the Acoustical Society of America*, 143:1366–1375, 03 2018. doi: 10.1121/1.5026239.
 - [58] M. A. Formoso, A. Ortiz, F. J. Martinez-Murcia, N. Gallego, and J. L. Luque. Detecting phase-synchrony connectivity anomalies in eeg signals. application to dyslexia diagnosis. *Sensors*, 21(21):7061, 2021.
 - [59] A. Frid and Z. Breznitz. An svm based algorithm for analysis and discrimination of dyslexic readers from regular readers using erps. In *2012 IEEE 27th Convention of Electrical and Electronics Engineers in Israel*, pages 1–4, 2012. doi: 10.1109/EEEI.2012.6377068.

- [60] A. Frid and L. M. Manevitz. Features and machine learning for correlating and classifying between brain areas and dyslexia. *arXiv preprint arXiv:1812.10622*, 2018.
- [61] A. D. Friederici. The brain basis of language processing: from structure to function. *Physiological reviews*, 91(4):1357–1392, 2011.
- [62] A. D. Friederici and S. M. Gierhan. The language network. *Current opinion in neurobiology*, 23(2):250–254, 2013.
- [63] P. Fries. A mechanism for cognitive dynamics. *Trends in Cognitive Sciences*, 9(10):474–480, 2005. doi: 10.1016/j.tics.2005.08.011.
- [64] K. J. Friston. Functional and effective connectivity: a review. *Brain Connectivity*, 1(1):13–36, 2011.
- [65] K. J. Friston, C. D. Frith, P. F. Liddle, and R. S. J. Frackowiak. Functional and effective connectivity in neuroimaging: A synthesis. *Human Brain Mapping*, 2(1-2):56–78, 1994.
- [66] N. Fujima. Current state of artificial intelligence in clinical applications for head and neck mr imaging. *Magnetic Resonance in Medical Sciences*, 22:401–414, 2023. doi: 10.2463/mrms.rev.2023-0047.
- [67] N. J. Gallego-Molina, A. Ortiz, F. J. Martínez-Murcia, I. Rodríguez-Rodríguez, and J. L. Luque. Assessing functional brain network dynamics in dyslexia from fnirs data. *International Journal of Neural Systems*, pages 2350017–2350017, 2023.
- [68] A. Ganapathiraju, J. E. Hamaker, and J. Picone. Applications of support vector machines to speech recognition. *IEEE Transactions on Signal Processing*, 52(8): 2348–2355, 2004.
- [69] M. Gavrilescu, S. Rossell, G. Stuart, T. Shea, H. Innes-Brown, K. Henshall, C. McKay, A. Sergejew, D. Copolov, and G. Egan. Reduced connectivity of the auditory cortex in patients with auditory hallucinations: A resting state functional magnetic resonance imaging study. *Psychological Medicine*, 40:1149 – 1158, 2010. ISSN 0033-2917.
- [70] geeksforgeeks.org. Bagging vs boosting in machine learning. URL <https://media.geeksforgeeks.org/wp-content/uploads/20210707140911/Boosting.png>. Accessed: 2024-12-01.
- [71] O. Ghitza and S. Greenberg. On the possible role of brain rhythms in speech perception: intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica*, 66(1-2):113–126, 2009.
- [72] S. Gibbon, A. Attaheri, Á. N. Choidealbha, S. Rocha, P. Brusini, N. Mead, P. Boutris, H. Olawole-Scott, H. Ahmed, S. Flanagan, et al. Machine learning accurately classifies neural responses to rhythmic speech vs. non-speech from 8-week-old infant eeg. *Brain and Language*, 220:104968, 2021.

-
- [73] J. Giehl, N. Noury, and M. Siegel. Dissociating harmonic and non-harmonic phase-amplitude coupling in the human brain. *NeuroImage*, 227:117648, 2021.
 - [74] A.-L. Giraud and D. Poeppel. Cortical oscillations and speech processing: emerging computational principles and operations. *Nature neuroscience*, 15(4):511–517, 2012.
 - [75] A.-L. Giraud and F. Ramus. Neurogenetics and auditory processing in developmental dyslexia. *Current opinion in neurobiology*, 23(1):37–42, 2013.
 - [76] D. V. Godoy. DL-visuals. URL <https://github.com/dvgodoy/dl-visuals>. Accessed: 2024-12-01.
 - [77] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *science*, 286(5439):531–537, 1999.
 - [78] U. Goswami. A temporal sampling framework for developmental dyslexia. *Trends in cognitive sciences*, 15(1):3–10, 2011.
 - [79] U. Goswami. Sensory theories of developmental dyslexia: three challenges for research. *Nature Reviews Neuroscience*, 16(1):43–54, 2015.
 - [80] U. Goswami. Speech rhythm and language acquisition: An amplitude modulation phase hierarchy perspective. *Annals of the New York Academy of Sciences*, 0(0), 2019. ISSN 1749-6632. doi: 10.1111/nyas.14137.
 - [81] U. Goswami. Toward realizing the promise of educational neuroscience: Improving experimental design in developmental cognitive neuroscience studies. *Annual Review of Developmental Psychology*, 2(1):133–155, 2020.
 - [82] U. Goswami, T. Fosker, M. Huss, N. Mead, and D. Szűcs. Rise time and formant transition duration in the discrimination of speech sounds: the ba-wa distinction in developmental dyslexia. *Developmental Science*, 14:34–43, 2010. doi: 10.1111/j.1467-7687.2010.00955.x.
 - [83] R. Goutagny, G. Ning, C. Cavanagh, J. Jackson, J.-G. Chabot, R. Quirion, S. Krantic, and S. Williams. Corrigendum: Alterations in hippocampal network oscillations and theta-gamma coupling arise before $\alpha\beta$ overproduction in a mouse model of alzheimer’s disease. *European journal of neuroscience*, 38(9-10), 2013.
 - [84] M. Habib. The neurological basis of developmental dyslexia and related disorders: A reappraisal of the temporal hypothesis, twenty years on. *Brain sciences*, 11(6): 708, 2021.
 - [85] J. A. Hämäläinen, A. Rupp, F. Soltész, D. Szűcs, and U. Goswami. Reduced phase locking to slow amplitude modulation in adults with dyslexia: an meg study. *Neuroimage*, 59(3):2952–2961, 2012.

- [86] A. Hassan, Q. Chen, Y. Liu, T. Jiang, G. Li, M. Jiang, N. Li, B. Lv, and S. Liu. Do plants affect brainwaves? effect of indoor plants in work environment on mental stress. *European Journal of Horticultural Science*, 85:279–283, 2020. doi: 10.17660/ejhs.2020/85.4.9.
- [87] M. Hata, H. Kazui, T. Tanaka, R. Ishii, L. Canuet, R. D. Pascual-Marqui, Y. Aoki, S. Ikeda, H. Kanemoto, K. Yoshiyama, M. Iwase, and M. Takeda. Functional connectivity assessed by resting state eeg correlates with cognitive decline of alzheimer’s disease – an eloreta study. *Clinical Neurophysiology*, 127(2):1269–1278, 2016.
- [88] B. J. He. Scale-free brain activity: past, present, and future. *Trends in cognitive sciences*, 18(9):480–487, 2014.
- [89] A. Herweg, J. Gutzeit, S. C. Kleih, and A. Kübler. Wheelchair control by elderly participants in a virtual environment with a brain-computer interface (bci) and tactile stimulation. *Biological Psychology*, 121:117–124, 2016. doi: 10.1016/j.biopsycho.2016.10.006.
- [90] G. Hickok and D. Poeppel. The cortical organization of speech processing. *Nature reviews neuroscience*, 8(5):393–402, 2007.
- [91] G. Hickok and D. Poeppel. The cortical organization of speech processing. *Nature reviews. Neuroscience*, 8:393–402, 06 2007. doi: 10.1038/nrn2113.
- [92] G. Hickok and D. Poeppel. Neural basis of speech perception. *Handbook of clinical neurology*, 129:149–160, 2015.
- [93] S. Hirano, A. Nakhnikian, Y. Hirano, N. Oribe, S. Kanba, T. Onitsuka, M. Levin, and K. M. Spencer. Phase-amplitude coupling of the electroencephalogram in the auditory cortex in schizophrenia. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 3(1):69–76, 2018.
- [94] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [95] F. Hoeft, B. D. McCandliss, J. M. Black, A. Gantman, N. Zakerani, C. Hulme, H. Lyytinen, S. Whitfield-Gabrieli, G. H. Glover, A. L. Reiss, et al. Neural systems predicting long-term outcome in dyslexia. *Proceedings of the National Academy of Sciences*, 108(1):361–366, 2011.
- [96] B. Horwitz and A. R. Braun. Brain network interactions in auditory, visual and linguistic processing. *Brain and language*, 89(2):377–384, 2004.
- [97] C. Huang and P. Lin. Effects of symbol component on the identifying of graphic symbols from eeg for young children with and without developmental delays. *Applied Sciences*, 9:1260, 2019. doi: 10.3390/app9061260.

-
- [98] H. Huang, J. Zhang, L. Zhu, J. Tang, G. Lin, W. Kong, X. Lei, and L. Zhu. Eeg-based sleep staging analysis with functional connectivity. *Sensors*, 21(6):1988, 2021.
 - [99] F. Huettig, T. Lachmann, A. Reis, and K. M. Petersson. Distinguishing cause from effect – many deficits associated with developmental dyslexia may be a consequence of reduced and suboptimal reading experience. *Language, Cognition and Neuroscience*, 33:333–350, 2017. doi: 10.1080/23273798.2017.1348528.
 - [100] M. J. Hülsemann, E. Naumann, and B. Rasch. Quantification of phase-amplitude coupling in neuronal oscillations: comparison of phase-locking value, mean vector length, modulation index, and generalized-linear-modeling-cross-frequency-coupling. *Frontiers in neuroscience*, 13:573, 2019.
 - [101] A. G. Huth, W. A. de Heer, T. L. Griffiths, F. E. Theunissen, and J. L. Gallant. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, 532(7600):453–458, Apr 2016. doi: 10.1038/nature17637.
 - [102] A. Hyafil, L. Fontolan, C. Kabdebon, B. Gutkin, and A.-L. Giraud. Speech encoding by coupled cortical theta and gamma oscillations. *elife*, 4:e06213, 2015.
 - [103] R. A. Ince, B. L. Giordano, C. Kayser, G. A. Rousselet, J. Gross, and P. G. Schyns. A statistical framework for neuroimaging data analysis based on mutual information estimated via a gaussian copula. *Human brain mapping*, 38(3):1541–1573, 2017.
 - [104] I. Isa, B. Zainuddin, Z. Hussain, and S. Sulaiman. Preliminary study on analyzing eeg alpha brainwave signal activities based on visual stimulation. *Procedia Computer Science*, 42:85–92, 2014. doi: 10.1016/j.procs.2014.11.037.
 - [105] Ixnay. A diagram for a one-unit recurrent neural network (rnn). from bottom to top : input state, hidden state, output state. u, v, w are the weights of the network. compressed diagram on the left and the unfold version of it on the right. URL https://commons.wikimedia.org/wiki/File:Recurrent_neural_network_unfold.svg. Accessed: 2024-12-01.
 - [106] S. R. Jammalamadaka and A. SenGupta. *Topics in Circular Statistics*. World Scientific, 1 edition, 2016. doi: 10.1142/4031.
 - [107] B. Jiang, N. Guo, Y. Ge, L. Zhang, M. Oudkerk, and X. Xie. Development and application of artificial intelligence in cardiac imaging. *British Journal of Radiology*, 93, 2020. doi: 10.1259/bjr.20190812.
 - [108] Jimenez-Bravo, Miguel and Marrero, Victoria and Benítez-Burraco, Antonio. An oscillopathic approach to developmental dyslexia: from genes to speech processing. *Behavioural Brain Research*, 329:84–95, 2017.
 - [109] M. Jiménez-Bravo, V. Marrero, and A. Benítez-Burraco. An oscillopathic approach to developmental dyslexia: from genes to speech processing. *Behavioural Brain Research*, 329:84–95, 2017. doi: 10.1101/108704.

- [110] G. John and P. Langley. Estimating continuous distributions in bayesian classifiers. *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence*, 1, 02 2013.
- [111] J. Jordan. Setting the learning rate of your neural network. URL https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcSKNuJAuvko5J5jSa4gyG9i94KO_Ig148xq4Q&s. Accessed: 2024-12-01.
- [112] W. Klimesch. Eeg alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain research reviews*, 29(2-3):169–195, 1999.
- [113] B. Koo and B. Shin. Using geometry based anomaly detection to check the integrity of ifc classifications in bim models. *Journal of KIBIM*, 7(1):18–27, 2017.
- [114] M. Kramer and U. Eden. Assessment of cross-frequency coupling with confidence using generalized linear models. *Journal of neuroscience methods*, 220(1):64–74, 2013. ISSN 0165-0270.
- [115] P. K. Kuhl. Early language acquisition: cracking the speech code. *Nature reviews neuroscience*, 5(11):831–843, 2004.
- [116] S. Kullback and R. A. Leibler. On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.
- [117] J.-P. Lachaux, E. Rodriguez, J. Martinerie, and F. J. Varela. Measuring phase synchrony in brain signals. *Human Brain Mapping*, 8(4):194–208, 1999.
- [118] Larhmam. Maximum-margin hyperplane and margin for an svm trained on two classes. samples on margins are called support vectors. URL https://commons.wikimedia.org/wiki/File:SVM_margin.png. Accessed: 2024-12-01.
- [119] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance. Eegnet: a compact convolutional neural network for eeg-based brain-computer interfaces. *Journal of neural engineering*, 15(5):056013, 2018.
- [120] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back. Face recognition: A convolutional neural-network approach. *IEEE transactions on neural networks*, 8(1):98–113, 1997.
- [121] Y. LeCun, Y. Bengio, et al. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10):1995, 1995.
- [122] K. Lehongre, F. Ramus, N. Villiermet, D. Schwartz, and A.-L. Giraud. Altered low-gamma sampling in auditory cortex accounts for the three main facets of dyslexia. *Neuron*, 72(6):1080–1090, 2011. ISSN 0896-6273. doi: <https://doi.org/10.1016/j.neuron.2011.11.002>. URL <https://www.sciencedirect.com/science/article/pii/S0896627311010014>.

- [123] Y. Lei, M. Yin, M. Yu, J. Yu, S. Zeng, W. Lv, J. Li, H. Ye, X. Cui, and C. Dietrich. Artificial intelligence in medical imaging of the breast. *Frontiers in Oncology*, 11, 2021. doi: 10.3389/fonc.2021.600557.
- [124] R. Lesser, H. Lüders, G. Klem, D. Dinner, H. Morris, and J. Hahn. Cortical after-discharge and functional response thresholds: results of extraoperative testing. *Epilepsia*, 25(5):615–621, 1984.
- [125] M. Li, K. Chang, B. Bearce, C. Y. Chang, A. J. Huang, and J. C. Campbell. Siamese neural networks for continuous disease severity evaluation and change detection in medical imaging. *npj Digit. Med.*, 3, 2020. doi: 10.1038/s41746-020-0255-1.
- [126] G. M. D. Liberto, J. O’Sullivan, and E. C. Lalor. Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Current Biology*, 25:2457–2465, 2015. doi: 10.1016/j.cub.2015.08.030.
- [127] F. J. Liégeois and A. T. Morgan. Neural bases of childhood speech disorders: lateralization and plasticity for speech functions during development. *Neuroscience & Biobehavioral Reviews*, 36(1):439–458, 2012.
- [128] H. Liu, L. Zhang, Q. Xi, X. Zhao, F. Wang, X. Wang, W. Men, and Q. Lin. Changes in brain lateralization in patients with mild cognitive impairment and alzheimer’s disease: a resting-state functional magnetic resonance study from alzheimer’s disease neuroimaging initiative. *Frontiers in neurology*, 9:3, 2018.
- [129] M. Lizarazu, M. Lallier, N. Molinaro, M. Bourguignon, P. M. Paz-Alonso, G. Lerma-Usabiaga, and M. Carreiras. Developmental evaluation of atypical auditory sampling in dyslexia: Functional and structural evidence. *Human Brain Mapping*, 36(12):4986–5002, 2015. doi: <https://doi.org/10.1002/hbm.22986>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.22986>.
- [130] S. M. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 4765–4774. Curran Associates, Inc., 2017. URL <http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf>.
- [131] H. Luo and D. Poeppel. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, 54:1001–1010, 2007.
- [132] J. Luo, Z. Zhang, Y. Fu, and F. Rao. Time series prediction of covid-19 transmission in america using lstm and xgboost algorithms. *Results in Physics*, 27: 104462, 2021.
- [133] W. Ma. Research on the application of artificial intelligence in pathology and experimental diagnosis. *Advanced Medical Research*, 2, 2023. doi: 10.62836/amr.v2i1.163.

- [134] W. Maass et al. Real-time computing without stable states. *Neural Computation*, 14(11):2531–2560, 2002. doi: 10.1162/089976602760407890.
- [135] K. Mandke, S. Flanagan, A. Macfarlane, F. Gabrielczyk, A. Wilson, J. Gross, and U. Goswami. Neural sampling of the speech signal at different timescales by children with dyslexia. *NeuroImage*, 253:119077, 2022.
- [136] F. J. Martínez Murcia. *Statistical Neuroimage Modelling, Processing and Synthesis Based on Texture and Component Analysis: Tackling the Small Sample Size Problem*. PhD Thesis, Universidad de Granada, 2017.
- [137] F. J. Martínez-Murcia, A. Ortiz, R. Morales-Ortega, P. López, J. L. Luque, D. Castillo-Barnes, F. Segovia, I. A. Illan, J. Ortega, J. Ramirez, et al. Periodogram connectivity of eeg signals for the detection of dyslexia. In *International Work-Conference on the Interplay Between Natural and Artificial Computation*, pages 350–359. Springer, 2019.
- [138] F. J. Martinez-Murcia, A. Ortiz, J.-M. Gorriz, J. Ramirez, and D. Castillo-Barnes. Studying the Manifold Structure of Alzheimer’s Disease: A Deep Learning Approach Using Convolutional Autoencoders. *IEEE Journal of Biomedical and Health Informatics*, 24(1):17–26, Jan. 2020. ISSN 2168-2194, 2168-2208. doi: 10.1109/JBHI.2019.2914970.
- [139] F. J. Martinez-Murcia, A. Ortiz, J. M. Gorriz, J. Ramirez, P. J. Lopez-Abarejo, M. Lopez-Zamora, and J. L. Luque. Eeg connectivity analysis using denoising autoencoders for the detection of dyslexia. *International Journal of Neural Systems*, 30(07):2050037, 2020. doi: 10.1142/S0129065720500379. URL <https://doi.org/10.1142/S0129065720500379>. PMID: 32466692.
- [140] C. y. D. Ministerio de Educación. *La atención al alumno con dislexia en el Sistema educativo en el contexto de las necesidades específicas de apoyo educativo*. Colección Eurydice. España-REDIE, 2012.
- [141] mlvidhya.com. Understanding bias variance tradeoff in machine learning. URL <https://mlvidhya.com/wp-content/uploads/2023/12/Bias-Variance-Tradeoff.jpg>. Accessed: 2024-12-01.
- [142] B. Mohammed, K. Bertling, A. Abbosh, and A. Rakie. brain waves. *Electronics Letters*, 52(16):1358–1358, 2016. doi: 10.1049/el.2016.2561.
- [143] Y. Mohan, S. S. Chee, D. K. P. Xin, and L. P. Foong. Artificial neural network for classification of depressive and normal in eeg. In *2016 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, pages 286–290. IEEE, 2016. (see p. 24).
- [144] N. Molinaro, M. Lizarazu, M. Lallier, M. Bourguignon, and M. Carreiras. Out-of-synchrony speech entrainment in developmental dyslexia. *Human Brain Mapping*, 37:2767–2783, 08 2016.

- [145] F. Mormann, J. Fell, N. Axmacher, B. Weber, K. Lehnertz, C. E. Elger, and G. Fernández. Phase/amplitude reset and theta-gamma interaction in the human medial temporal lobe during a continuous word recognition memory task. *Hippocampus*, 15(7):890–900, 2005.
- [146] S. Motamedi-Fakhr, M. Moshrefi-Torbati, M. Hill, C. M. Hill, and P. R. White. Signal processing techniques applied to human sleep eeg signals—a review. *Biomedical Signal Processing and Control*, 10:21–33, 2014.
- [147] T. T. Munia and S. Aviyente. Time-frequency based phase-amplitude coupling measure for neuronal oscillations. *Scientific reports*, 9(1):12441, 2019.
- [148] T. T. K. Munia and S. Aviyente. Time-frequency based phase-amplitude coupling measure for neuronal oscillations. *Scientific Reports*, 9:1–15, 08 2019.
- [149] J. Munilla, A. Ortiz, J. M. Górriz, and J. Ramírez. Construction and analysis of weighted brain networks from sice for the study of alzheimer’s disease. *Frontiers in Neuroinformatics*, 11:19, 2017.
- [150] E. Niedermeyer. *Niedermeyer’s electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2011.
- [151] Nvidia. Xgboost. URL <https://www.nvidia.com/content/dam/en-zz/Solutions/glossary/data-science/xgboost/img-2.png>. Accessed: 2024-12-01.
- [152] R. D. Oades. Frontal, temporal and lateralized brain function in children with attention-deficit hyperactivity disorder: a psychophysiological and neuropsychological viewpoint on development. *Behavioural brain research*, 94(1):83–95, 1998.
- [153] S. Ocklenburg and O. Güntürkün. *The lateralized brain: The neuroscience and evolution of hemispheric asymmetries*. Elsevier, 2024.
- [154] A. Ortiz, J. Munilla, J. M. Górriz, and J. Ramírez. Ensembles of deep learning architectures for the early diagnosis of the alzheimer’s disease. *International Journal of Neural Systems*, 26(07):1650025, 2016.
- [155] A. Ortiz, P. López, J. L. Luque, F. J. Martínez-Murcia, D. Aquino-Britez, and J. Ortega. An anomaly detection approach for dyslexia diagnosis using eeg signals. In *International Work-Conference on the Interplay Between Natural and Artificial Computation*, pages 369–378. Springer, 2019.
- [156] A. Ortiz, F. J. Martinez-Murcia, J. L. Luque, A. Giménez, R. Morales-Ortega, and J. Ortega. Dyslexia diagnosis by eeg temporal and spectral descriptors: An anomaly detection approach. *International Journal of Neural Systems*, 30(07):2050029, 2020.
- [157] T. E. Özkurt. Statistically reliable and fast direct estimation of phase-amplitude cross-frequency coupling. *IEEE Transactions on Biomedical Engineering*, 59(7):1943–1950, 2012.

- [158] J. Pereira, P. Ofner, A. Schwarz, A. Sburlea, and G. Müller-Putz. Eeg neural correlates of goal-directed movement intention. *Neuroimage*, 149:129–140, 2017. doi: 10.1016/j.neuroimage.2017.01.030.
- [159] H. Perera, M. F. Shiratuddin, and K. W. Wong. Review of eeg-based pattern classification frameworks for dyslexia. *Brain informatics*, 5:1–14, 2018.
- [160] H. Perera, M. F. Shiratuddin, K. W. Wong, and K. Fullarton. Eeg signal analysis of writing and typing between adults with dyslexia and normal controls. *International Journal of Interactive Multimedia and Artificial Intelligence*, 5(1):62, 2018.
- [161] R. Peterson and B. Pennington. Developmental dyslexia. *Lancet*, 379:1997–2007, 04 2012.
- [162] T. W. Picton, S. Bentin, P. Berg, E. Donchin, S. A. Hillyard, R. Johnson, G. A. Miller, W. Ritter, D. Ruchkin, M. Rugg, et al. Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology*, 37(2):127–152, 2000.
- [163] D. Poeppel. The analysis of speech in different temporal integration windows: cerebral lateralization as ‘asymmetric sampling in time’. *Speech Communication*, 41(1):245–255, 2003.
- [164] L. V. Popering, A. Tahmassebi, U. Meyer-Baese, M. Dyrba, J. Munilla, A. Ortiz, and A. Meyer-Baese. Identifying the diffusion source of dementia spreading in structural brain networks. In B. S. Gimi and A. Krol, editors, *Medical Imaging 2021: Biomedical Applications in Molecular, Structural, and Functional Imaging*, volume 11600, pages 58–63. International Society for Optics and Photonics, SPIE, 2021.
- [165] A. J. Power, N. Mead, L. Barnes, and U. Goswami. Neural entrainment to rhythmic speech in children with developmental dyslexia. *Frontiers in human neuroscience*, 7:777, 2013.
- [166] A. J. Power, L. J. Colling, N. Mead, L. Barnes, and U. Goswami. Neural encoding of the speech envelope by children with developmental dyslexia. *Brain and Language*, 160:1–10, 2016.
- [167] V. Rahimi, G. Mohammadkhani, J. Alaghband Rad, S. Z. Mousavi, and M. E. Khalili. Modulation of auditory temporal processing, speech in noise perception, auditory-verbal memory, and reading efficiency by anodal tdc in children with dyslexia. *Neuropsychologia*, 177:108427, 2022. ISSN 0028-3932.
- [168] C. Ramon and M. Holmes. Stochastic behavior of phase synchronization index and cross-frequency couplings in epileptogenic zones during interictal periods measured with scalp deeg. *Frontiers in Neurology*, 4, 2013. doi: 10.3389/fneur.2013.00057.

-
- [169] R. P. Rao and D. H. Ballard. Predictive coding in the visual cortex. *Nature Neuroscience*, 2(1):79–87, 1999. doi: 10.1038/4580.
 - [170] T. Rathcke and C. Lin. Towards a comprehensive account of rhythm processing issues in developmental dyslexia. *Brain Sciences*, 11:1303, 2021. doi: 10.3390/brainsci11101303.
 - [171] Z. Rezvani, M. Zare, G. Žarić, M. Bonte, J. Tijms, M. Van der Molen, and G. Fraga González. Machine learning classification of dyslexic children based on eeg local network features. *BioRxiv*, page 569996, 2019.
 - [172] F. Riaz, A. Hassan, S. Rehman, I. K. Niazi, and K. Dremstrup. Emd-based temporal and spectral features for the classification of eeg signals using supervised learning. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 24(1):28–35, Jan 2016.
 - [173] M. Ribolsi, G. Koch, V. Magni, G. Di Lorenzo, I. A. Rubino, A. Siracusano, and D. Centonze. Abnormal brain lateralization and connectivity in schizophrenia. *Reviews in the Neurosciences*, 20(1):61–70, 2009. doi: doi:10.1515/REVNEURO.2009.20.1.61. URL <https://doi.org/10.1515/REVNEURO.2009.20.1.61>.
 - [174] F. Richlan. Developmental dyslexia: dysfunction of a left hemisphere reading network. *Frontiers in human neuroscience*, 6:120, 2012.
 - [175] I. Rodríguez-Rodríguez, A. Ortiz, N. J. Gallego-Molina, M. A. Formoso, and W. L. Woo. Eeg interchannel causality to identify source/sink phase connectivity patterns in developmental dyslexia. *International Journal of Neural Systems*, 33(04):2350020, 2023.
 - [176] L. J. Rogers. Brain lateralization and cognitive capacity. *Animals*, 11(7):1996, 2021.
 - [177] K. Rothmaler and G. Ivanova. Circular correlation coefficients versus the phase-locking-value. *Biomedizinische Technik. Biomedical engineering*, page 2, 09 2013.
 - [178] P. Sauseng and W. Klimesch. What does phase information of oscillatory brain activity tell us about cognitive processes? *Neuroscience & Biobehavioral Reviews*, 32(5):1001–1013, 2008.
 - [179] Y. Scientific. Uncharted waters in dyslexia research, December 2014. URL <http://www.yalescientific.org/2014/12/uncharted-waters-in-dyslexia-research/>. Accessed: 2024-11-20.
 - [180] scikit learn.org. Cross-validation: evaluating estimator performance. URL https://scikit-learn.org/1.5/_images/grid_search_cross_validation.png. Accessed: 2024-12-01.
 - [181] T. Sejnowski and O. Paulsen. Network oscillations: Emerging computational principles. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 26:1673–6, 03 2006. doi: 10.1523/JNEUROSCI.3737-05d.2006.

- [182] L. S. Shapley. A value for n-person games. In H. W. Kuhn and A. W. Tucker, editors, *Contributions to the Theory of Games II*, pages 307–317. Princeton University Press, Princeton, NJ, 1953. URL <https://www.rand.org/pubs/papers/P295.html>.
- [183] S. E. Shaywitz, B. A. Shaywitz, R. K. Fulbright, P. Skudlarski, W. E. Mencl, R. T. Constable, K. R. Pugh, J. M. Holahan, K. E. Marchione, J. M. Fletcher, et al. Neural systems for compensation and persistence: young adult outcome of childhood reading disability. *Biological psychiatry*, 54(1):25–33, 2003.
- [184] S. Siuly, H. Wang, and Y. Zhang. Detection of motor imagery eeg signals employing naïve bayes based learning process. *Measurement*, 86:148–158, 03 2016. doi: 10.1016/j.measurement.2016.02.059.
- [185] J. O. Smith. *Introduction to Digital Filters with Audio Applications*. W3K Publishing, 2007. URL <https://www.dsprelated.com/freebooks/filters/>.
- [186] O. Sporns and C. J. Honey. Small worlds inside big brains. *Proceedings of the National Academy of Sciences*, 103(51):19219–19220, 2006.
- [187] Stackoverflow. What’s the difference between hidden and output in pytorch lstm? URL <https://stackoverflow.com/questions/48302810/whats-the-difference-between-hidden-and-output-in-pytorch-lstm>. Accessed: 2024-12-01.
- [188] B. E. Stein and T. R. Stanford. Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, 9:406–406, 2008. URL <https://api.semanticscholar.org/CorpusID:265846612>.
- [189] M. Steriade. Grouping of brain rhythms in corticothalamic systems. *Neuroscience*, 137(4):1087–1106, 2006.
- [190] J. Sung. Diffusion tensor imaging studies on arcuate fasciculus in stroke patients: A review. *Frontiers in Human Neuroscience*, 7(749), 2013. doi: 10.3389/fnhum.2013.00749.
- [191] P. Tamboer, H. Vorst, S. Ghebreab, and H. Scholte. Machine learning and dyslexia: Classification of individual structural neuro-imaging scans of students with and without dyslexia. *NeuroImage: Clinical*, 11:508–514, 2016.
- [192] P. Tamboer, H. Vorst, S. Ghebreab, and H. Scholte. Machine learning and dyslexia: Classification of individual structural neuro-imaging scans of students with and without dyslexia. *NeuroImage: Clinical*, 11:508–514, 2016.
- [193] A. B. Tort, M. A. Kramer, C. Thorn, D. J. Gibson, Y. Kubota, A. M. Graybiel, and N. J. Kopell. Dynamic cross-frequency couplings of local field potential oscillations in rat striatum and hippocampus during performance of a t-maze task. *Proceedings of the National Academy of Sciences*, 105(51):20517–20522, 2008.

-
- [194] A. B. Tort, R. Komorowski, H. Eichenbaum, and N. Kopell. Measuring phase-amplitude coupling between neuronal oscillations of different frequencies. *Journal of neurophysiology*, 104(2):1195–1210, 2010.
 - [195] A. B. L. Tort, R. Komorowski, H. Eichenbaum, and N. Kopell. Measuring phase-amplitude coupling between neuronal oscillations of different frequencies. *Journal of Neurophysiology*, 104(2):1195–1210, 2010.
 - [196] J. Twilt, K. Leeuwen, H. Huisman, J. Fütterer, and M. Rooij. Artificial intelligence based algorithms for prostate cancer classification and detection on magnetic resonance imaging: a narrative review. *Diagnostics*, 11:959, 2021. doi: 10.3390/diagnostics11060959.
 - [197] B. Türkbey and M. Haider. Deep learning-based artificial intelligence applications in prostate mri: brief summary. *British Journal of Radiology*, 95, 2021. doi: 10.1259/bjr.20210563.
 - [198] S. A. Unde and R. Shriram. Coherence analysis of eeg signal using power spectral density. In *2014 Fourth International Conference on Communication Systems and Network Technologies*, pages 871–874, 2014.
 - [199] A. Vahid, A. Bluschke, V. Roessner, S. Stober, and C. Beste. Deep learning based on event-related eeg differentiates children with adhd from healthy controls. *JCM*, 8:1055, 2019. doi: 10.3390/jcm8071055.
 - [200] S. Vanhatalo, J. M. Palva, M. Holmes, J. Miller, J. Voipio, and K. Kaila. Infralow oscillations modulate excitability and interictal epileptic activity in the human cortex during sleep. *Proceedings of the National Academy of Sciences*, 101(14): 5053–5057, 2004.
 - [201] R. VanRullen. Perceptual cycles. *Trends in Cognitive Sciences*, 20(10):723–735, 2016. doi: 10.1016/j.tics.2016.08.007.
 - [202] R. VanRullen and C. Koch. Is perception discrete or continuous? *Trends in Cognitive Sciences*, 7(5):207–213, 2003. doi: 10.1016/S1364-6613(03)00131-0.
 - [203] V. N. Vapnik. *Statistical Learning Theory*. John Wiley and Sons, Inc., New York, 1998.
 - [204] J. Vesanto and E. Alhoniemi. Clustering of the self-organizing map. *IEEE Transactions on neural networks*, 11(3):586–600, 2000.
 - [205] P. Virtala, S. Talola, E. Partanen, and T. Kujala. Poor neural and perceptual phoneme discrimination during acoustic variation in dyslexia. *Scientific Reports*, 10(1):1–11, 2020.
 - [206] K. E. Waldie, C. E. Haigh, G. Badzakova-Trajkov, J. Buckley, and I. J. Kirk. Reading the wrong way with the right hemisphere. *Brain Sciences*, 3(3):1060–1075, 2013.

- [207] A. Widmann, E. Schröger, and B. Maess. Digital filter design for electrophysiological data – a practical approach. *Journal of Neuroscience Methods*, 250:34–46, 2015. doi: 10.1016/j.jneumeth.2014.08.002. URL <https://www.sciencedirect.com/science/article/pii/S0165027014002842>.
- [208] P. H. Wolff, G. F. Michel, M. Ovrut, and C. T. V. Drake. Rate and timing precision of motor coordination in developmental dyslexia. *Developmental Psychology*, 26:349–359, 1990. doi: 10.1037/0012-1649.26.3.349.
- [209] J. R. Wolpaw and E. W. Wolpaw. Brain-computer interfaces: something new under the sun. *Brain-computer interfaces: principles and practice*, 14:3–12, 2012.
- [210] www.cs.us.es. Self organizing maps (som) in netlogo. URL <https://www.cs.us.es/~fsancho/Blog/posts/img/self-organizing-feature-map-2.png>. Accessed: 2024-12-01.
- [211] M. Wöstmann, B. Herrmann, B. Maess, and J. Obleser. Spatiotemporal dynamics of auditory attention synchronize with speech. *Proc. Natl. Acad. Sci. U.S.A.*, 113: 3873–3878, 2016. doi: 10.1073/pnas.1523357113.
- [212] E. Zaccarella, G. Papitto, and A. D. Friederici. Language and action in broca’s area: Computational differentiation and cortical segregation. *Brain and Cognition*, 147:105651, 2021. ISSN 0278-2626.
- [213] S. Zahia, B. Garcia-Zapirain, I. Saralegui, and B. Fernandez-Ruanova. Dyslexia detection using 3d convolutional neural networks and functional magnetic resonance imaging. *Computer Methods and Programs in Biomedicine*, 197:105726, 2020.
- [214] G. Zhang. What is the kernel trick? why is it important? URL https://miro.medium.com/v2/resize:fit:1400/1*mCwnu5kXot6buL7jeIafqQ.png. Accessed: 2024-12-01.
- [215] F. Zubler and A. Tzovara. Deep learning for eeg-based prognostication after cardiac arrest: from current research to future clinical applications. *Frontiers in Neurology*, 14, 2023. doi: 10.3389/fneur.2023.1183810.