Evaluación del Impacto del Aprendizaje Auto-Supervisado en la Precisión de Interfaces Cerebro-Ordenador basadas en Imaginación Motora

S. Pérez Velasco[1,2], D. Marcos-Martínez[1,2], E. Santamaría-Vázquez[1,2], R. Ruiz-Gávez[1], R. Hornero[1,2]

[1] Grupo de Ingeniería Biomédica (GIB), Universidad de Valladolid, Valladolid, España, sergio.perezv@uva.es

[2] Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), España

Resumen

Las interfaces cerebro-ordenador (BCIs) buscan proporcionar vías de comunicación directas entre el cerebro humano y dispositivos externos. No obstante, la decodificación precisa de las intenciones del usuario es todavía un desafío, en parte por las limitaciones inherentes a la electroencefalografía (EEG), como son una resolución espacial limitada y relación señal-ruido reducida. Este estudio aborda estos desafíos mediante la aplicación de técnicas de aprendizaje auto-supervisado (SSL) en el preentrenamiento de una red basada en la arquitectura de transformer. Nuestra aproximación descompone la señal EEG en segmentos y utiliza un enmascarado y reconstrucción para obtener representaciones más robustas y efectivas.

Evaluamos el impacto de estas técnicas en la mejora de la clasificación de un sistema BCI basado en imaginación motora (MI) con una base de datos pública de 109 sujetos, utilizando un esquema de validación cruzada inter-sujeto k-fold (k=5). Comparamos tres escenarios diferentes: un modelo sin preentrenamiento SSL, un modelo con sondeo lineal de las características extraídas, y un modelo con fine tunning de toda la red. Nuestros resultados indican que el preentrenamiento con técnicas de SSL mejora significativamente la precisión de la clasificación de MI. Concretamente, la precisión se incrementa desde un 78.4%±12.2%, hasta un 79.9%±11.9% utilizando sondeo lineal, y alcanza un 82.4%±11.7% cuando se aplica fine tunning a toda la red. Este trabajo demuestra el potencial del SSL aplicado a redes basadas en transformer para avanzar en la interpretación de señales EEG.

1. Introducción

La Ingeniería Biomédica (IB) supone una sinergia entre medicina e ingeniería, orientada hacia el desarrollo de herramientas innovadoras para abordar enfermedades y trastornos humanos [1]. Este estudio se focaliza en una herramienta particularmente prometedora: la interfaz cerebro-ordenador (*brain-computer interface*, BCI) [2]. Estas interfaces ofrecen un camino de comunicación alternativo y directo entre el cerebro y el entorno [2]. Entre los métodos para captar la actividad cerebral, la electroencefalografía (EEG) destaca por su alta resolución temporal, facilidad de uso y coste relativamente bajo, convirtiéndola en una de las mejores elecciones de registro para los investigadores BCI [2]. No obstante, la EEG también presenta desafíos, como su baja resolución espacial y una relación señal-ruido (*signal to noise ratio*, SNR) reducida. Estas limitaciones subrayan la necesidad de perfeccionar tanto los sistemas de registro como las estrategias de procesamiento de señales para interpretar de forma precisa las intenciones del usuario BCI.

Paralelamente, el campo del aprendizaje profundo (*deep learning*, DL) ha experimentado avances significativos en la última década, algunos de los cuales se han aplicado con éxito en el análisis de señales EEG para una variedad de aplicaciones médicas y de investigación [3]. Gran parte de estos avances han sido posibles debido a la capacidad que tiene DL de transferir el aprendizaje de unas tareas a otras y a la disponibilidad de grandes conjuntos de datos etiquetados. No obstante, el aprendizaje auto-supervisado (*self-supervised learning*, SSL) ofrece una oportunidad adicional para optimizar las capacidades del DL, al permitir el preentrenamiento en datos no etiquetados, más abundantes y menos costosos de adquirir. Este método ya ha sido usado en estudios de señales EEG, según las referencias [4], [5]. En el primer estudio, se adapta el exitoso proceso de SSL anteriormente empleado al reconocimiento de habla [4]. En este enfoque, la red diseñada primero comprime la señal con una red de capas convolucionales (CNN) en un espacio latente, que luego se enmascara. Después, se intenta reconstruir este espacio latente tras introducirlo en una arquitectura basada en *transformer* [6]. Un segundo estudio se basa en el primero mencionado [5]. Al igual que en el estudio anterior, la red simplifica la señal en un espacio latente y luego lo enmascara. Sin embargo, la diferencia clave aquí es que la red intenta recuperar la señal original, en lugar del espacio latente enmascarado. Si bien estos trabajos han tenido éxito en probar la utilidad de SSL, la tarea que realizan limita el aprendizaje de la red puesto que se basa en el enmascarado del espacio latente y no el enmascarado de la señal original.

En el presente estudio, abordamos estos desafíos al aplicar técnicas de SSL en el preentrenamiento de una red de aprendizaje profundo basada en la arquitectura de *transformer*. Descomponemos la señal EEG en segmentos, similar al enfoque utilizado en las redes *transformer* de visión (ViT) [7], y empleamos una técnica de enmascarado para ocultar y posteriormente reconstruir fragmentos de la señal original. Esta tarea, con un alto ratio de enmascarado (75%), fuerza a la red a aprender representaciones más robustas y efectivas de la señal EEG [8]. Además, una mayor relación de enmascarado acelera la fase de preentrenamiento SSL, ya que solo se procesan los segmentos no enmascarados de la señal.

El objetivo principal de esta investigación es evaluar el impacto del preentrenamiento con técnicas de SSL, basadas en enmascarado y reconstrucción de señal EEG para la mejora de la clasificación en sistemas BCI basados

ISBN: 978-84-17853-76-1

en imaginación motora. Este incremento se evaluará en una base de datos pública de 109 sujetos [9]. El estudio prueba por primera vez el uso de redes basadas completamente en la arquitectura *transformer* para la extracción de características relevantes del EEG.

2. Métodos

2.1. Dataset

Se ha examinado la base de datos pública Physionet [9]. Esta base de datos incluye registros EEG a 160 Hz de 109 participantes, con un rango de 42 a 46 eventos de MI por individuo. Después de una señal auditiva, una flecha señalaba qué tipo de MI se debía efectuar (i.e., imaginar abrir o cerrar la mano izquierda o derecha) durante un periodo continuo de tres segundos. La actividad se llevó a cabo en una única sesión y los participantes no recibieron retroalimentación acerca de la tarea mental que estaban ejecutando. Esta base de datos se utilizará tanto para la tarea de preentrenamiento como para la clasificación.

Aplicamos el siguiente preprocesamiento sobre la señal EEG [10]: (1) se extraen los electrodos F7, F3, T7, C3, P7, P3, O1, PZ, CZ, F8, F4, T8, C4, P8, P4 y O2; (2) se aplica un filtro notch para eliminar la señal de línea eléctrica; (3) se realiza un filtrado espacial de referencia promedio común (CAR) a estos 16 electrodos; (4) se aplica un remuestreo a 128 Hz evitando aliasing con un filtrado paso-bajo a 63 Hz; (5) se extraen los ensayos con una ventana de tiempo de 1.5 segundos después de la señal auditiva; y (6) se aplica estandarización z-score por canal y ensayo.

2.2. Aprendizaje auto-supervisado (SSL)

En este trabajo adaptamos de forma rigurosa el SSL basado en la arquitectura de *autoencoders* con enmascaramiento (*masked auto-encoders*, MAE) originalmente aplicada en el análisis de imágenes [8]. Este proceso se articula en cuatro etapas fundamentales, esquematizadas en la *Figura 1*. En la primera etapa, diseñamos una capa neuronal específica que fragmenta la señal EEG en segmentos temporales de 250 ms, equivalente a 32 muestras a una frecuencia de 128 Hz. En la segunda etapa, introducimos una capa de *embedding*, que cumple una doble función: añadir información posicional a cada segmento y enmascarar un subconjunto de segmentos seleccionados aleatoriamente siguiendo una distribución uniforme. La necesidad de añadir información posicional surge debido a la naturaleza invariante al orden de las redes *transformer*, que no reconocen el orden secuencial de las entradas. En la tercera etapa, un *encoder* compuesto por una red *transformer* de seis capas se encarga de procesar los segmentos de señal que permanecen visibles. Finalmente, en la cuarta etapa, un *decoder* más ligero de una sola capa *transformer* se encarga de la tarea de reconstrucción de la señal EEG original, utilizando tanto los segmentos codificados como los tokens asociados a los segmentos enmascarados.

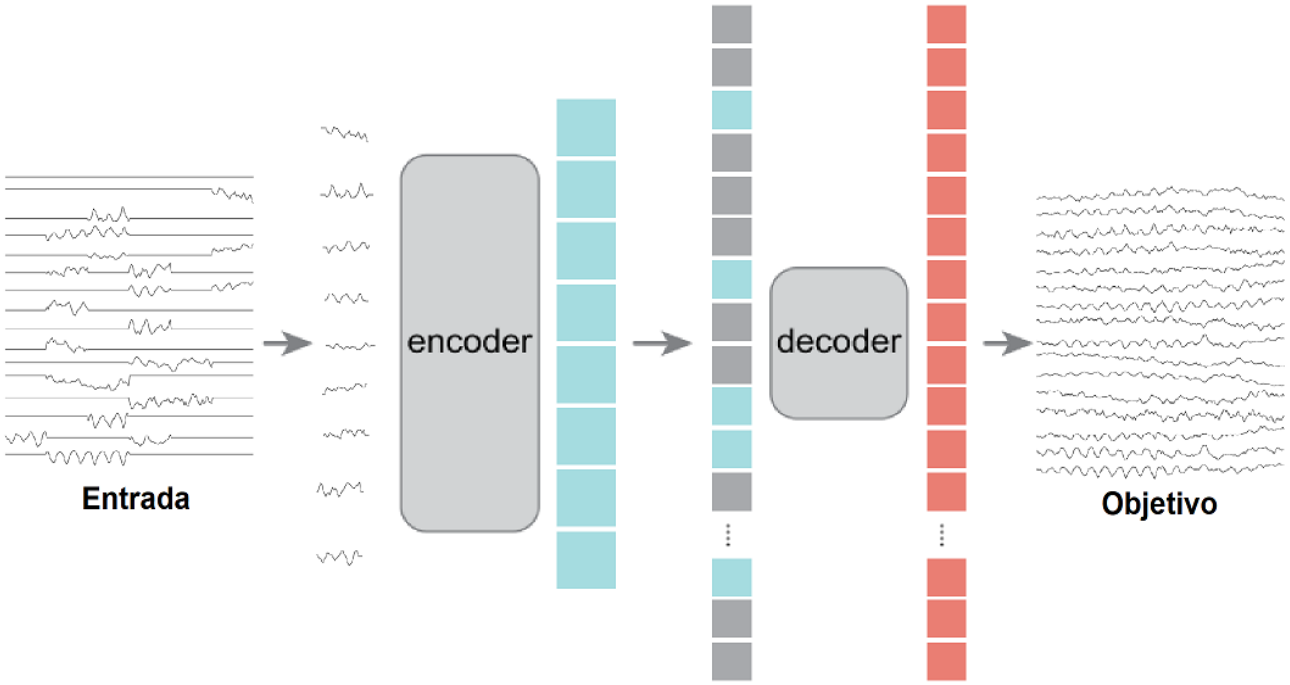

Figura 1. *Arquitectura de autoencoder con enmascaramiento (masked auto-encoder, MAE) optimizada para aplicaciones en electroencefalografía (EEG). En la fase de preentrenamiento, una proporción significativa de la señal EEG (e.g., 75%) se enmascara intencionadamente. El encoder se encarga de procesar únicamente los segmentos de señal que permanecen visibles. A continuación, el decoder opera tanto sobre los segmentos codificados como sobre los tokens de enmascaramiento, con el objetivo de reconstruir fielmente la señal EEG original. Tras finalizar el preentrenamiento, se prescinde del decoder, y el encoder se emplea en el procesamiento de señales EEG sin enmascarar para llevar a cabo la clasificación de eventos de imaginación motora (MI). Figura adaptada* [8].

ISBN: 978-84-17853-76-1

Los hiperparámetros utilizados en la fase de preentrenamiento se detallan a continuación: (1) optimizador *AdamW*, (2) *learning rate* de $1x10^{-3}$, (3) *weight decay* de $1x10^{-5}$, (4) 500 épocas de entrenamiento, (5) detención temprana (*early stopping)* con un margen de espera de 25 épocas, (6) tamaño de *embedding* de 256, (7) 6 cabezas del mecanismo de atención, y (8) función de pérdidas basada en el error medio cuadrado (*mean squared error*, MSE).

2.3. Clasificación de imaginación motora (MI)

Tras completar la fase de preentrenamiento, descartamos el *decoder* y utilizamos exclusivamente el *encoder* para procesar la señal EEG completa sin enmascaramiento, que abarca un intervalo temporal de 1.5 segundos, con el objetivo de clasificar MI. Para evaluar la eficacia de nuestro enfoque, se establecen tres escenarios comparativos: (1) un modelo que no ha sido preentrenado con técnicas de SSL; (2) la aplicación de un sondeo lineal utilizando las características extraídas por el *encoder* en el preentrenamiento; y (3) *fine tunning* de toda la arquitectura de la red a partir del *encoder* preentrenado.

Optamos por medir el rendimiento mediante un esquema de validación cruzada inter-sujetos *k*-fold (*k*=5). En otras palabras, en cada uno de los tres escenarios mencionados, el modelo se entrena para clasificar MI en el 80% de los participantes, y posteriormente se valida en el 20% de los sujetos que no fueron parte del conjunto de entrenamiento. Repetimos este proceso cinco veces, lo que nos permite calcular la precisión de clasificación en todos los sujetos incluidos en la muestra de entrenamiento. Al comienzo de cada iteración los parámetros del *encoder* se inicializarán aleatoriamente en el escenario (1), mientras que se usarán los obtenidos durante el preentrenamiento en los escenarios (2) y (3).

3. Resultados

3.1. Reconstrucción de la señal

Si bien la tarea de reconstrucción de la señal no es el objetivo final que se busca con la aplicación del SSL, es de gran interés observar qué grado de éxito alcanza en esta tarea el conjunto *encoder/decoder*. La *Figura 2* ilustra la evolución del valor del MSE entre la señal reconstruida y la señal original a lo largo del proceso de entrenamiento del SSL. El gráfico evidencia una evolución paralela en el MSE, tanto para el conjunto de entrenamiento como para el conjunto de validación. El entrenamiento se detiene antes de llegar a las 500 épocas debido a la activación del mecanismo de *early stopping*. La *Figura* 3, por otro lado, muestra el proceso de enmascaramiento y reconstrucción llevado a cabo por nuestro conjunto *encoder*/*decoder* tras el preentrenamiento basado en MAE.

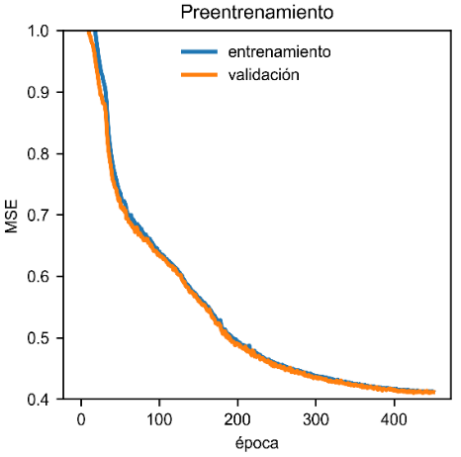

Figura 2. *Evolución del valor del error medio cuadrado (MSE) de la tarea de reconstrucción durante el aprendizaje auto-supervisada (SSL).*

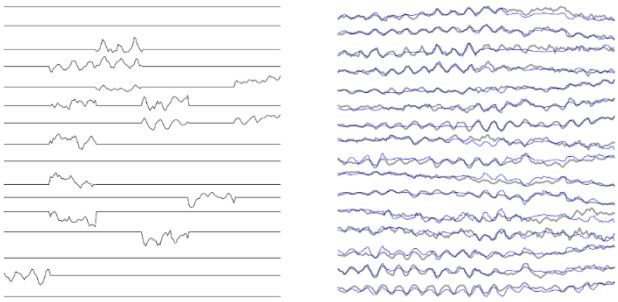

Figura 3. *Representación del enmascarado y reconstrucción de una señal tras el preentrenamiento. A la izquierda la señal enmascarada. A la derecha en gris transparente la señal original y en azul la señal reconstruida a partir de la señal no enmascarada.*

3.2. Clasificación en Physionet

En la *Tabla 1* se pueden observar los resultados de precisión para los 109 sujetos de Physionet [9] en cada una de las tres condiciones: (1) sin preentrenamiento; (2) preentrenamiento con sondeo lineal; y (3) preentrenamiento con *fine tunning* de la arquitectura completa. Se presenta la media de clasificación y la desviación estándar de las precisiones de clasificación de la imaginación motora inter-sujeto. Se observa como la condición de preentrenamiento SSL basado en MAE junto con *fine tunning* de toda la red logran los mejores resultados de precisión. Este comportamiento es análogo al observado en el trabajo original aplicado a imágenes [8].

	μ±σ
Sin preentrenamiento	78.4±12.2
SSL + sondeo lineal	79.9±11.9
SSL + *fine tunning*	**82.4±11.7**

Tabla 1. *Comparación de precisiones para los tres escenarios. μ±σ: precision media y desviación estándar obtenida para todos los sujetos siguiendo un k-fold con 5 iteraciones. Los mejores resultados están en* ***negrita****.*

4. Discusión

El presente estudio aborda el desafío persistente en el ámbito BCI referente a la decodificación precisa de las intenciones del usuario. La combinación de técnicas de SSL basadas en MAE y la arquitectura de *transformer* representa un avance significativo en el procesamiento de señales EEG. El uso de un enmascarado de alta proporción (75%) directamente sobre la señal EEG es novedoso,

ISBN: 978-84-17853-76-1

acelera el entrenamiento, y se demostró eficaz en la extracción representaciones relevantes de la señal. Además, la utilización de un esquema de validación cruzada inter-sujeto *k*-fold proporciona un marco riguroso para evaluar el rendimiento del modelo.

Los resultados sugieren que el uso de SSL podría tener aplicaciones más amplias en el campo de la IB, particularmente en la interpretación y el análisis de datos EEG. Sin embargo, hay algunas limitaciones que deben considerarse. Por ejemplo, aunque el dataset Physionet incluye gran cantidad de usuarios, la inclusión de más bases de datos podría ofrecer un preentrenamiento más potente y mejorar las representaciones obtenidas. Sobre todo, bases de datos con miles de usuarios y horas de registros EEG como la base de datos de Temple University Hospital (TUH) [11]. Aunque esta metodología ha mostrado avances al eliminar completamente el uso de capas convolucionales, los resultados obtenidos aún no superan a los trabajos que se consideran el estado del arte y que utilizan CNNs para la clasificación.

Comparado con trabajos anteriores que emplean SSL en el análisis de EEG [4, 5], este estudio supera ciertas limitaciones al centrarse no solo en el enmascarado del espacio latente creado por una etapa convolucional, sino también en el enmascarado de la señal original. Este enfoque permite a la red componerse solo de módulos *transformer*, que permiten acelerar el preentrenamiento al solo introducirse los segmentos de señal visibles.

Las futuras investigaciones podrían explorar el *fine tunning* del modelo a otros paradigmas BCI. Además, se podría emplear mayor cantidad de datos de EEG con gran variedad de paradigmas como datos de preentrenamiento, y no solo la misma base de datos que se usará para evaluar.

5. Conclusión

Este trabajo ha investigado la aplicabilidad y eficacia de técnicas de SSL para mejorar la precisión en la decodificación de las intenciones del usuario en BCIs basadas en EEG. Hemos demostrado que la arquitectura de *transformer*, que está preentrenada mediante técnicas SSL, permite segmentar y procesar eficazmente las señales EEG.

La validación de nuestro modelo en la clasificación de MI sobre una base de datos pública de 109 sujetos, mediante un enfoque de validación cruzada inter-sujeto *k*-fold, ha revelado una mejora significativa en precisión con respecto a la misma red sin este preentrenamiento. Los resultados indican que el preentrenamiento con técnicas de SSL no solo mejora la precisión de la clasificación, sino que también ofrece un nuevo camino hacia la interpretación más precisa de las señales EEG, lo cual es crucial para el avance de las BCIs.

Agradecimientos

Este estudio ha sido financiado por los proyectos TED2021-129915B-I00, PID2020-115468RB-I00 y RTC2019-007350-1 financiadas por el Ministerio de Ciencia e Innovación/Agencia Estatal de investigación/10.13039/501100011033/', FEDER Una forma de hacer Europa; y por 'Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBERBBN)' a través de 'Instituto de Salud Carlos III'. S. Pérez-Velasco y D. Marcos-Martínez son beneficiarios de una ayuda PIF de la Consejería de Educación de la Junta de Castilla y León.

Referencias

[1] J. Enderle *et al.*, *Introduction to biomedical engineering*. Academic press, 2012.

[2] J. R. Wolpaw and E. W. Wolpaw, *Brain-Computer Interfaces: Principles and Practice*. Oxford University Press, 2012.

[3] A. Craik *et al.*, "Deep learning for electroencephalogram (EEG) classification tasks: a review," *J Neural Eng*, vol. 16, no. 3, p. 031001, Jun. 2019.

[4] D. Kostas *et al.*, "BENDR: Using Transformers and a Contrastive Self-Supervised Learning Task to Learn From Massive Amounts of EEG Data," *Front Hum Neurosci*, vol. 15, no. June, pp. 1–15, Jun. 2021.

[5] H.-Y. S. Chien *et al.*, "MAEEG: Masked Auto-encoder for EEG Representation Learning," no. NeurIPS, Oct. 2022.

[6] A. Vaswani *et al.*, "Attention Is All You Need," *Adv Neural Inf Process Syst*, vol. 2017-Decem, no. Nips, pp. 5999–6009, Jun. 2017.

[7] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," *ICLR 2021 - 9th International Conference on Learning Representations*, Oct. 2020.

[8] K. He *et al.*, "Masked Autoencoders Are Scalable Vision Learners," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2022, pp. 15979–15988.

[9] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals.," *Circulation*, vol. 101, no. 23, Jun. 2000.

[10] S. Perez-Velasco *et al.*, "EEGSym: Overcoming Inter-Subject Variability in Motor Imagery Based BCIs With Deep Learning," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 1766–1775, 2022.

[11] A. Harati *et al.*, "The TUH EEG CORPUS: A big data resource for automated EEG interpretation," in *2014 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, IEEE, Dec. 2014, pp. 1–5.