ISSN 1870-4069

Evaluación de métodos de aprendizaje supervisado para la clasificación de palabras utilizando señales de electroencefalografía

Denise Alonso-Vázquez[1], Tonatiuh Hernández-del-Toro[2], Omar Mendoza-Montoya[1], Ricardo Caraza[3], Hector R. Martinez[3], Carlos A. Reyes-García[2], Javier M. Antelis[1]

Tecnológico de Monterrey, Escuela de Ingeniería y Ciencias, México

Instituto Nacional de Astrofísica Óptica y Electrónica, Departamento de Ciencias Computacionales, México

Tecnológico de Monterrey, Escuela de Medicina y Ciencias de la Salud, México

denise.alonso.v@tec.mx

Resumen. En este trabajo evaluamos diferentes métodos para la clasificación de palabras pronunciadas a partir de señales de electroencefalografía (EEG). Se utilizó la red neuronal convolucional EEGNet y se compararon los resultados obtenidos con el uso de características en el dominio de la frecuencia y clasificadores tradicionales: LDA, SVM y RF. Se utilizó una base de datos de cinco palabras en español: "si", "no", "agua", "comida" y "dormir", adquirida mediante 32 canales de electroencefalografía distribuidos uniformemente sobre el cuero cabelludo en participantes sanos. La clasificación se realizó intra-sujeto y en todos los casos se obtuvieron porcentajes de exactitud superiores al azar (20 %). Utilizando la EEGNet se obtuvo un mejor desempeño respecto a los otros métodos, obteniendo una exactitud promedio de 75.05±7.30 % entre todos los participantes. El participante con mayor exactitud obtuvo 84.00±4.54 % y el de menor desempeño 60.00±8.48 %. También se encontró que la palabra que mejor decodifica este método es "agua" y la peor es "dormir". Este trabajo es un estudio preliminar para la decodificación del intento del habla en pacientes con esclerosis lateral amiotrófica bulbar, utilizando un método de adquisición de señales no-invasivo como el EEG.

Palabras clave: Electroencefalografía, decodificación del habla, EEGNet.

Denise Alonso-Vázquez, Tonatiuh Hernández-del-Toro, Hector R. Martinez, et al.

Evaluation of Supervised Learning Methods for Words Classification using Electroencephalography Signals

Abstract. In this work, we evaluate different methods for the classification of pronounced words from electroencephalography (EEG) signals. The EEGNet convolutional neural network was used, and the results obtained were compared with models based on the traditional classifiers LDA, SVM, and RF; that used characteristics in the frequency domain. A database of five Spanish words was used: "si", "no", "water", "food" and "sleep". EEG recordings were acquired from 32 uniformly distributed electroencephalography channels on the scalp of healthy participants. The classification was carried out intra-subject, and for all methods, higher than chance accuracy percentages were obtained (20%). The best classification performance was obtained by EEGNet, in comparison to the other methods, obtaining an average accuracy of 75.05±7.30% among all the participants. The participant with the highest accuracy obtained 84.00±4.54 %, and the one with the lowest performance 60.00±8.48 %. It was also found that the word that was best decoded by this method was "water' and the worst was "sleep". This work is a preliminary study for the decoding of attempted speech in patients with bulbar amyotrophic lateral sclerosis using EEG as non-invasive signal acquisition method.

Keywords: Electroencephalography, speech decoding, EEGNet.

1. Introducción

Existen diferentes enfermedades en las que las neuronas motoras mueren progresivamente, y como consecuencia, se pierde la capacidad de hablar. Un ejemplo es el caso de la esclerosis lateral amiotrófica (ELA), enfermedad en la que aproximadamente el 85 % de los pacientes experimentan síntomas de disfunción bulbar, como disminución de la comunicación verbal y función de deglución, afectando significativamente su calidad de vida [14].

A nivel global más de 5,000 personas son diagnósticadas por año, con una prevalencia de 1 en 20,000 personas [1]. En México al menos hasta el 2018 se registraron más de 6,000 casos diagnosticados [7]. Actualmente, existen diversos dispositivos de asistencia para pacientes con limitaciones del habla.

En pacientes con ELA y como consecuencia tetraplegia y anartria (trastorno en la expresión del lenguaje que consiste únicamente en la imposibilidad de articular los sonidos) el dispositivo más utilizado es el dispositivo de comunicación por seguimiento ocular, en el que el usuario necesita señalar y mantener la mirada en los comandos que se muestran en el monitor, lo cual es detectado con una cámara infrarroja, sin embargo, son dispositivos de alto costo [6].

Otra alternativa que se ha estudiado ampliamente es el uso de las interfaces cerebro-computadora (BCIs por las siglas en inglés de brain-computer interface), las cuales detectan y cuantifican las características de las señales cerebrales que indican la intención del usuario, traducen estas mediciones en tiempo real en comandos para el dispositivo y proporcionan retroalimentación simultánea al usuario [22].

Existen diferentes mecanismos para medir la actividad cerebral, utilizando técnicas invasivas y no-invasivas, la electroencefalografía (EEG) es considerada como el método más común en la medición de señales cerebrales ya que tiene una alta resolución temporal, es fácil de usar, segura y asequible [15], por lo tanto, la mayoría de las BCIs utilizan señales adquiridas con EEG.

Algunas BCIs trabajan con potenciales evocados, como el potencial P300 o el potencial evocado visual de estado estacionario (SSVEP por las siglas en inglés de Steady-state visually evoked potential), y otras con tareas cognitivas, como el movimiento imaginado [19].

En el P300, por lo general, el participante mira la pantalla donde parpadean los caracteres y selecciona uno de ellos prestándole atención. En deletreadores que funcionan con SSVEP se utilizan estímulos visuales o recuadros que parpadean a distintas frecuencias. Cuando el usuario se concentra en el elemento que desea seleccionar, se genera un potencial en la corteza visual con la misma frecuencia de parpadeo que la imagen.

En el movimiento imaginado, el participante imagina que está moviendo una de sus extremidades sin realizar ningún movimiento, y con la decodificación de estas señales se controla una BCI. Estos paradigmas han sido ampliamente utilizados, sin embargo, no decodifican directamente la respuesta relacionada con el habla.

Se han realizado diferentes estudios decodificando el habla, ya sea pronunciada, susurrada, silenciosa (solamente se gesticula sin emitir sonidos) o imaginada (pronunciación interna de la palabra, sin gesticular ningún movimiento y sin emitir ningún sonido). En [16], se decodifica el intento del habla en un participante con anartria como consecuencia de un accidente cerebrovascular.

Se implantó una matriz de 128 electrodos en la corteza sensoriomotora del habla y se utilizó un modelo de lenguaje natural que calculó la probabilidad de la siguiente palabra dadas las palabras anteriores en una secuencia. En [2] se presentó un enfoque para sintetizar el habla audible a partir del habla imaginada y el habla susurrada, utilizando un diccionario de 100 palabras holandesas y electrodos estereotácticos profundos implantados en un participante.

A pesar de que estos trabajos decodifican el habla en sus diferentes paradigmas, utilizan métodos invasivos, lo que aporta un riesgo a la salud del paciente, al igual que incrementa los costos respecto al EEG. Algunos trabajos a partir de señales de EEG clasifican clases gramaticales, en [10] con habla imaginada (sustantivos y verbos) y en [3] utilizando habla pronunciada (adverbios de decisión y sustantivos).

También se han decodificado vocales en español [20] y palabras cortas y largas en inglés [17]. En [9] evalúan diferentes métodos de clasificación tradicionales y tres redes neuronales convolucionales distintas, con el objetivo de encontrar una óptima combinación de hiperparámetros para la clasificación de palabras en español.

La mayoría de los métodos que utilizan señales de EEG como método de adquisión, implementan sus protocolos con participantes sanos, es decir, sin ningún padecimiento neurológico diagnósticado y ningún transtorno del habla. Por lo tanto, en este trabajo se muestra un estudio preliminar de decodificación de palabras pronunciadas en una base de datos propia, que posteriormente será implementado en pacientes con ALS bulbar mientras realizan la tarea de intento del habla.

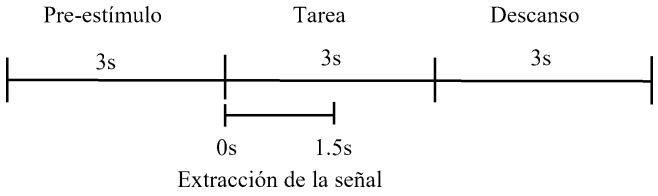

Fig. 1. Distribución de los estímulos en un ensayo.

Se evaluaron algoritmos de clasificación tradicionales ampliamente utilizados en BCIs (LDA, SVM y RF), utilizando una extracción de características en el dominio de la frecuencia, al igual que una red neuronal convolucional (EEGNet) que sigue la metodología comúnmente utilizada en las BCIs.

Se realizó una clasificación intra-sujeto de cinco clases correspondientes a las siguientes palabras: “si”, “no”, “agua”, “comida” y “dormir”. Los resultados obtenidos reflejan que la EEGNet tiene un mejor desempeño que los demás métodos de clasificación evaluados en este estudio, obteniendo una exactitud promedio entre todos los participantes de $75.05 \pm 7.30\,\%$ (nivel de azar del $20\,\%$).

2. Métodos

En esta sección se presenta la descripción de la base de datos, el preprocesamiento de las señales, los modelos de clasificación utilizados, el procedimiento de evaluación y las métricas de desempeño.

2.1. Descripción de la base de datos

La base de datos contiene el registro de señales de electroencefalografía (EEG) de diez participantes sanos, 6 hombres y 4 mujeres con edad promedio de 24.8 años (std=8.4 años), diestros y hablantes nativos del Español durante la tarea de habla pronunciada. Se utilizó un grupo de 5 palabras en Español que consideramos útiles para una persona con limitaciones en la comunicación: “si”, “no”, “agua”, “comida” y “dormir”. La duración de cada ensayo (ver Figura 1) fue de 9s divididos en estímulos de 3s. Los primeros 3s corresponden a una cruz de fijación donde la instrucción fue poner atención y evitar movimientos.

Posteriormente se muestra de forma aleatoria una de las cinco palabras, donde la tarea dada al participante fue pronunciar (una sola vez) la palabra de manera natural, es decir, la forma en la que normalmente habla, finalmente el bloque de descanso fue representado por una palmera con 3s de duración.

Se grabaron 4 bloques de 50 ensayos cada uno por participante, por lo tanto se obtuvieron 40 ensayos por cada palabra. Las señales fueron registradas mediante 32 electrodos activos (Ag/AgCl) distribuidos uniformemente sobre el cuero cabelludo de acuerdo al sistema 10-20.

El equipo utilizado fue el amplificador de bioseñales de alto rendimiento g.HIAMP 256 de g.tec. Los datos fueron adquiridos a una frecuencia de muestreo de 1200Hz,

se aplicó un filtro pasabanda Butterworth de 0.5Hz a 500Hz y un filtro Notch en 60Hz, colocando la referencia en el lóbulo de la oreja derecha y el electrodo de tierra en la posición AFz. Previo a la sesión todos los participantes declararon no tener ningún trastorno del habla o desorden neurológico diagnosticado, además de firmar el consentimiento informado y otorgar el permiso para el uso de sus datos.

2.2. Preprocesamiento

La señal fue submuestreada a 256 Hz, el segmento de tiempo utilizado fue de 0 s a 1.5 s, donde el cero corresponde al instante en el que la palabra aparece en la pantalla. De acuerdo a [11], el mayor pico de voltaje correspondiente a la señal de EEG contaminada por actividad muscular en la gesticulación de movimientos, se encuentra alrededor de los 30Hz, por lo que, se aplicó un filtro Butterworth pasabanda de 1 Hz a 20 Hz.

2.3. Modelos de clasificación

Para estudiar la discriminación entre las cinco palabras, por medio de la clasificación multiclase utilizando señales electroencefalográficas, se evaluaron dos alternativas, la primera basada en extracción de características y clasificadores convencionales, y la segunda basada en un modelo de aprendizaje profundo.

Extracción de características y clasificadores

– **Extracción de características utilizando potencia espectral:** La densidad de potencia espectral (PSD por las siglas en inglés de power spectral density) ha sido ampliamente utilizada en señales de EEG para proporcionar información de la distribución de la potencia en las distintas bandas de frecuencia que conforman la señal.

Utilizamos la transformación de frecuencia multitaper para estimar la potencia espectral usando ventanas de Hanning. Se calculó el promedio de la potencia espectral para cada canal, en cinco bandas de frecuencia con una resolución de 0.5Hz: delta (1-4Hz), theta (4-7 Hz), alfa (8-13 Hz), beta-baja (12-15) y beta-media (15 Hz - 20 Hz).

Por lo tanto, el vector de características resultante es $\mathbf{x} \in \mathbb{R}^{(N_{\text{Bandas}} \cdot N_{\text{Canales}}) \times 1}$, donde $N_{\text{Bandas}} = 5$ es el número de bandas de frecuencia y $N_{\text{Canales}} = 32$ es el número de canales de EEG, por consiguiente, la dimensión del vector es $N_{\text{Bandas}} \cdot N_{\text{Canales}} = 160$, con una etiqueta asociada $\mathbf{y} \in (\text{si, no, agua, comida, dormir})$.

Implementamos la selección de características con el objetivo de obtener una representación de baja dimensión del conjunto de datos original, pero con un alto poder de discriminación. Utilizamos un método de selección de características basado en el valor F de ANOVA, para seleccionar las 50 características mejor clasificadas. Por lo tanto, la dimensión final del vector es (1,50).

Para la clasificación, utilizamos algoritmos de aprendizaje máquina ampliamente utilizados en señales de EEG [18, 12], los cuales son análisis discriminante lineal, máquinas de soporte vectorial y árboles aleatorios, también conocidos como: LDA, SVM y RF respectivamente por sus siglas en inglés.

– **Análisis discriminante lineal (LDA):** es un algoritmo que puede ser utilizado para aprendizaje supervisado y no supervisado. Consiste en encontrar la proyección del hiperplano, definida por el vector discriminante $w*$, que maximice la distancia entre las medias de las clases al mismo tiempo que minimiza su varianza.

La función del discriminante toma la forma $f(x) = w^T x$, donde w es un vector aprendido de pesos y x representa el (n+1)-dimensional vector de características de la instancia a clasificar. El problema de optimización a resolver, es el siguiente:

$$w^* = \frac{w^T S_B w}{w^T S_W w}, \tag{1}$$

donde S_B es la varianza entre clases y S_W es la covarianza dentro de la clase[5]. La solución se puede obtener resolviendo un sistema de valores propios generalizado. El hiperplano obtenido puede ser utilizado para clasificación, reducción de dimensionalidad e interpretación de la importancia de las características dadas[23].

– **Máquina de soporte vectorial (SVM):** es un método utilizado para clasificación, regresión y estimación de densidad. Consiste en encontrar un hiperplano que discrimine entre las clases de manera que maximice los márgenes de separación entre él y los datos más cercanos en cada clase (llamados vectores de soporte). El problema de optimización a resolver, es el siguiente:

$$w^*, b^*, \zeta^* = \text{argmin}_{w,b,\zeta} \left(\frac{1}{2}\|w\|^2 + C \sum_{i=1}^{m} \zeta_i \right), \tag{2}$$

donde w es el vector de pesos, b el término de sesgo, ζ_i las variables de holgura, C un parámetro de regularización que determina el compromiso entre el ancho del margen y el error de entrenamiento y w^*, b^*, ζ^* los valores óptimos para el modelo [21]. También es posible que el límite de decisión sea no lineal a través de una función kernel, ya sea polinomial, radial o sigmoideo.

Para este trabajo utilizamos un kernel lineal con un parámetro de regularización de 1. El SVM es un clasificador binario, sin embargo, se puede ampliar fusionando varios de su tipo en un clasificador multiclase implementando el enfoque de "uno contra uno" [12].

– **Bosques aleatorios (RF):** son una modificación del bagging (empaquetado) que crea un grupo de árboles descorrelacionados y luego los promedia, con el objetivo de disminuir la varianza del modelo. En algunos problemas, el rendimiento de RF es similar a boosting (ayuda a disminuir el sesgo del modelo), y son más sencillos de entrenar y ajustar.

Un promedio $\bar{x}$ de variables aleatorias B cada una con varianza σ^2 tiene varianza σ^2/B. Si las variables son identicamente distribuidas, pero no necesariamente independientes con correlación positiva por pares ρ, la varianza del promedio es:

$$\text{var}(\bar{x}) = \rho\sigma^2 + \frac{1-\rho}{B}\sigma^2. \tag{3}$$

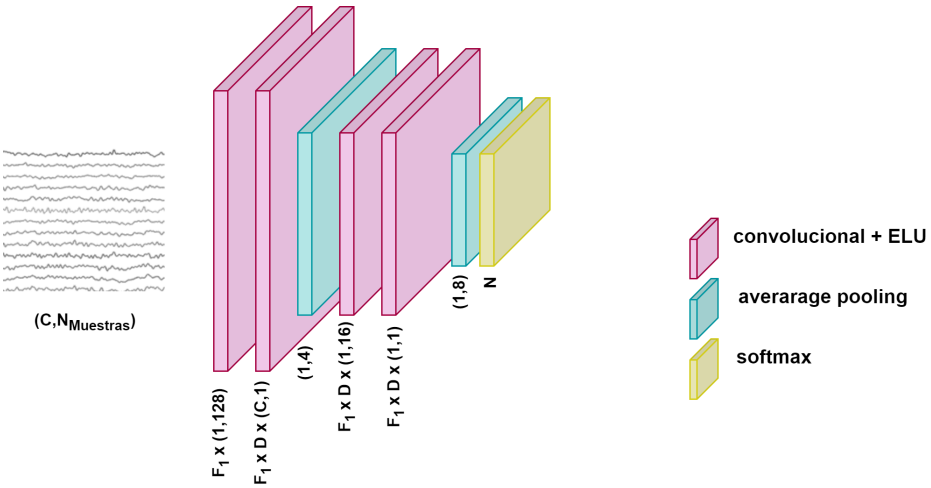

Fig. 2. Arquitectura de la red Neuronal EEGNet [8].

A medida que B aumenta, el segundo término desaparece, sin embargo, el primero permanece y el tamaño de la correlación de pares de árboles empaquetados, limita los beneficios del promedio. El objetivo de RF es mejorar la reducción de la varianza del bagging al reducir la correlación entre los árboles, sin aumentar demasiado la varianza. Esto se logra en el proceso de crecimiento de árboles a través de la selección aleatoria de las variables de entrada [4]. En este trabajo utilizamos un conjunto de 100 árboles.

Modelo de aprendizaje profundo

– **EEGNet:** La EEGNet es una red neuronal convolucional compacta para BCIs basadas en EEG, la cual se puede utilizar en diferentes paradigmas de BCI y entrenarse con una cantidad de datos muy limitados [8]. La arquitectura de la EEGNet (ver Figura 2) está compuesta por tres bloques: el primero se compone de una secuencia convolucional de dos pasos.

Comienza con una capa de convolución temporal para aprender filtros de frecuencia. Utilizamos $F1 = 8$, dónde $F1$ es el número de filtros temporales con un tamaño de kernel de (1,128), es decir, la mitad de la frecuencia de muestreo.

El segundo paso es una convolución profunda para aprender filtros espaciales para cada filtro temporal, con el objetivo de obtener una extracción eficiente de filtros espaciales específicos de frecuencia. El tamaño es (C, 1) con C definido como el número de canales, por lo tanto, C=32. El parámetro de profundidad D es es el número de filtros espaciales para aprender dentro de cada convolución temporal, D=2.

Este primer bloque está inspirado en el algoritmo Filter-Bank Common Spatial Pattern (FBCSP) [13]. Posteriormente se aplica Batch normalization entre la dimensión de los feature maps antes de aplicar el exponential linear unit (ELU) nonlinearity y la técnica Dropout para regularizar el modelo, con una probabilidad establecida en 0.85.

El bloque termina con una capa average pooling de tamaño (1, 4) que reduce la frecuencia de muestreo de la señal a 64Hz. El segundo bloque está compuesto por una

Tabla 1. Media aritmética ± desviación estándar de la exactitud total obtenida por participante en la clasificación de las cinco palabras (nivel de azar=20 %), para todos los métodos evaluados.

Participante	PSD+LDA	PSD+SVM	PSD+RF	EEGNet
1	38.00±9.09 %	43.00±12.3 %	39.00±5.75 %	81.00±2.23 %
2	37.00±7.37 %	34.00±5.18 %	29.00±5.18 %	79.00±7.41 %
3	59.50±8.73 %	62.50±3.53 %	44.00±8.22 %	75.50±9.25 %
4	29.00±5.75 %	32.00±7.58 %	30.00±6.85 %	68.00±7.79 %
5	41.50±2.23 %	36.50±7.42 %	38.00±5.12 %	73.50±9.11 %
6	32.00±4.80 %	38.00±5.97 %	31.00±2.85 %	79.00±4.54 %
7	53.00±7.34 %	50.50±7.37 %	48.00±4.11 %	70.00±8.10 %
8	50.00±10.5 %	54.00±12.3 %	49.50±4.81 %	84.00±4.54 %
9	43.50±9.94 %	45.00±9.84 %	37.50±9.01 %	60.00±8.48 %
10	31.00±3.35 %	32.00±8.73 %	31.00±2.24 %	80.50±6.71 %
Total	**41.45±10.1 %**	**42.75±10.3 %**	**37.70±7.54 %**	**75.05±7.30 %**

convolución separable formada por la combinación entre una convolución profunda (tamaño $(1, 16)$) y una convolución puntual, con $F2 = F1 * D$ dónde F2 es el número de filtros puntuales para aprender.

Al igual que en el bloque anterior se aplica Batch normalization entre la dimensión de los feature maps, el exponential linear unit (ELU) nonlinearity y la técnica Dropout para regularizar el modelo, con una probabilidad establecida en 0.85. El average pooling se establece en un tamaño de (1,8) para reducir las dimensiones.

En el bloque de clasificación se utiliza una softmax classification con N unidades, dónde N es el número de clases, por lo tanto N=5. La matriz que ingresa a la red neuronal tiene dimensión $\mathbf{x} \in \mathbb{R}^{N_{\text{Canales}} \times N_{\text{Muestras}}}$ con una etiqueta asociada $\mathbf{y} \in$ (si, no, agua, comida, dormir), donde $N_{\text{Canales}} = 32$ es el número de canales de EEG, y $N_{\text{Muestras}} = 385$ es el número de muestras contenidas en 1.5s de la señal.

Utilizamos los GPUs de Google Colab para el entrenamiento de la red desarrollada en Tensorflow utilizando Keras API. Se realizaron curvas de aprendizaje y basados en lo obtenido, se utilizaron 300 épocas. Para el ajuste del modelo se utilizó el optimizador Adam y la función de pérdida categorical cross-entropy.

2.4. Procedimiento de evaluación y métricas de desempeño

Los modelos se entrenaron intra-sujeto, es decir, el modelo se ajustó a cada uno de los participantes de forma independiente. Se realizó una clasificación multiclase de $N_{\text{Clases}} = 5$, donde cada clase corresponde a la pronunciación de cada una de las palabras. Para evaluar los modelos se utilizó validación cruzada de 5 iteraciones, es decir, el conjunto de características de cada participante se dividió en 5 grupos.

Las características de cuatro grupos se tomaron para ajustar los modelos de clasificación mientras que las características del grupo restante se utilizaron como un conjunto de datos de prueba para calcular las métricas de rendimiento.

Este procedimiento se repitió cinco veces, tomando un grupo diferente como conjunto de datos de entrenamiento y prueba en cada iteración, para garantizar que

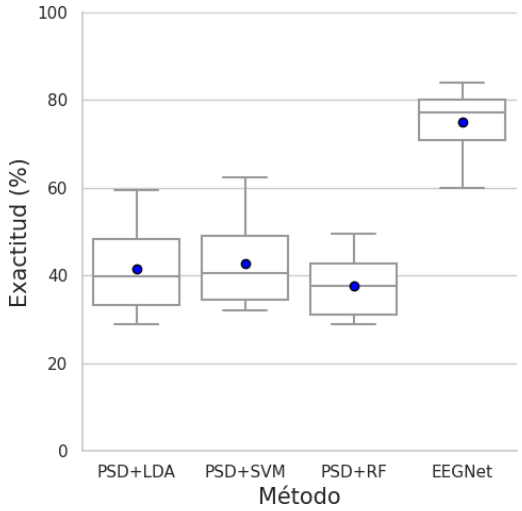

Fig. 3. Distribución de la exactitud total obtenida por cada método en la clasificación de las cinco palabras (nivel de azar=20 %). El punto corresponde a la media aritmética obtenida entre todos los participantes en cada uno de los métodos.

estos dos conjuntos siempre se excluyen mutuamente. Para evaluar el desempeño de los modelos, utilizamos las siguientes métricas:

– Exactitud de clasificación total, que representa el porcentaje total de valores correctamente clasificados.

– Precisión de clasificación por clase, indica qué tan confiable es nuestro modelo para predecir una clase específica.

– Sensibilidad por clase, se refiere al porcentaje de los elementos de la clase que fueron detectados correctamente entre todos los elementos de esa clase, es decir, el desempeño del modelo al detectar esa clase.

– Matriz de confusión para visualizar el número de predicciones de cada clase, respecto a las instancias en la clase real.

3. Resultados

La Tabla 1 contiene los resultados de la exactitud total obtenida por participante en la clasificación de las cinco palabras. Tomando en cuenta que el nivel de azar es del 20 % todos los participantes alcanzaron una exactitud promedio por encima del azar.

Utilizando la red neuronal convolucional EEGNet, se obtuvieron los porcentajes de exactitud más altos, donde el participante 8 logró el mejor desempeño con 84.00 ± 4.54 % de exactitud, mientras el peor caso con este método se obtuvo en el participante 9 con 60.00 ± 8.48 %.

El promedio de los resultados obtenidos utilizando la EEGNet fue de 75.05 ± 7.30 %, aproximadamente el doble del total obtenido en el método que combina

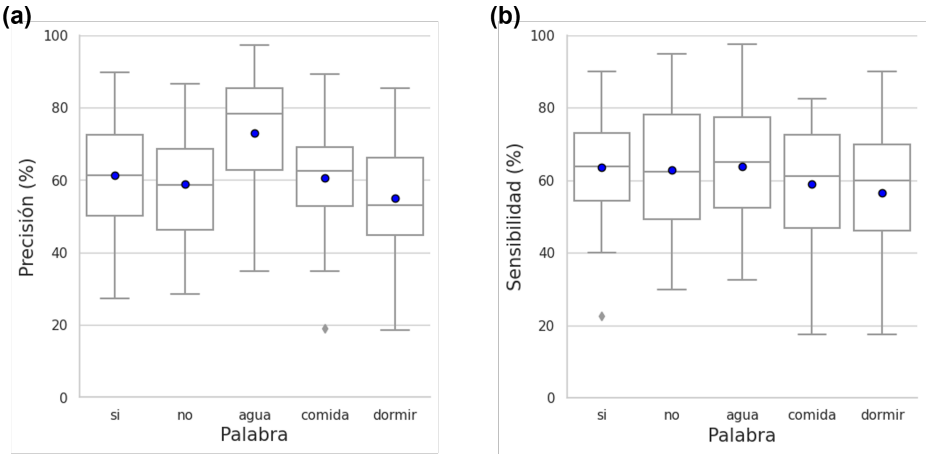

Fig. 4. Distribución de la precisión y sensibilidad por palabra, utilizando la EEGNet (nivel de azar=20 %). El punto corresponde a la media aritmética obtenida entre todos los participantes en cada una de las palabras. a) Porcentaje de precisión, b) porcentaje de sensibilidad.

PSD y RF donde se obtuvo el menor porcentaje de exactitud (37.70±7.54 %). Con PSD+RF el peor caso ocurre en el participante 2 (29.00±5.18 %) y su mejor caso al igual que con la EEGNet, en el participante 8 (49.50±4.81 %).

En los métodos que combinan PSD+LDA y PSD+SVM se obtuvieron resultados similares entre sí; 41.45±10.1 % y 42.75±10.3 % respectivamente. En estos dos casos los participantes que obtienen los porcentajes de exactitud más altos (participante 3 con 59.50±8.73 % y 62.50±3.53 %) y los mas bajos (participante 4 con 29.00±5.75 % 32.00±7.58 %) coinciden.

La Figura 3 muestra la distribución de los porcentajes de exactitud total alcanzados para cada uno de los métodos. Utilizando la EEGNet la mediana de los valores de exactitud total se encuentra en 77.25 % con una asimetría negativa, es decir, el 50 % de los valores están mayormente concentrados por encima de la mediana, mientras que, por debajo de ella el 50 % restante está más disperso.

Las combinaciones que se basan en la extracción de características de potencia junto con un algoritmo de clasificación, tienen la mediana de sus valores alrededor de 40 %, es decir, el doble del nivel de azar (20 %), además de una asimetría positiva.

Tomando en cuenta únicamente estos tres métodos, el promedio más alto se logra con la combinación PSD+SVM y el más bajo en PSD+RF. Con el objetivo de evaluar si la EEGNet tiene significativamente mejor desempeño que los métodos restantes, se realizó la prueba estadística Wilcoxon de una cola para cada uno de los métodos respecto a EEGNet.

Con un nivel de significancia $\alpha = 0.01$ y un p–valor=0.0009766 para cada una de las pruebas, se comprueba que estadísticamente los resultados obtenidos con la EEGNet son superiores respecto a cada uno de los 3 métodos restantes. Se calculó la distribución de los porcentajes de precisión y sensibilidad por clase obtenidos entre todos los participantes utilizando la EEGNet (ver Figura 4).

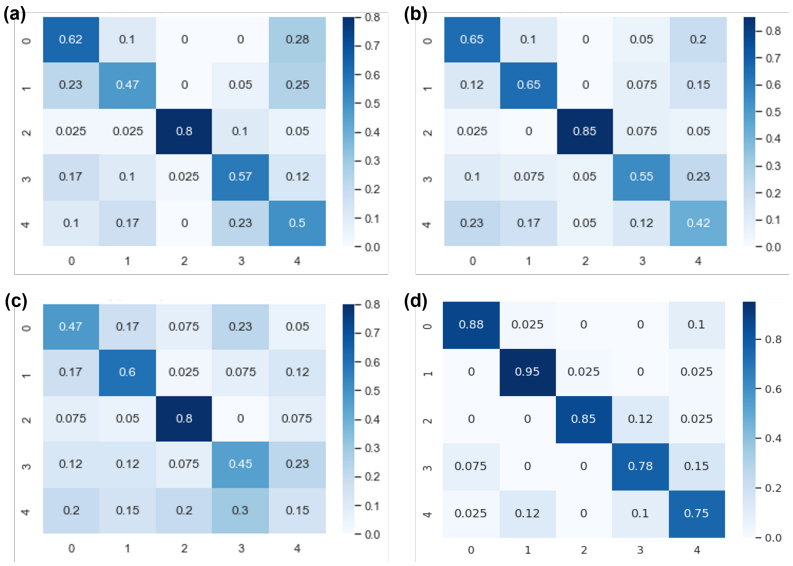

Fig. 5. Matrices de confusión obtenidas en los participantes con mejor porcentaje de exactitud. Método: a) PSD+LDA, b) PSD+SVM, c) PSD+RF y d)EEGNet. Cada valor numérico corresponde a cada una de las palabras de la siguiente manera: 0=“si”, 1=“no”, 2=“agua”, 3=“comida” y 4=“dormir”.

Con una precisión promedio por encima del 70 % se observa que la palabra que el modelo clasifica con mayor confiabilidad es “agua”. Mientras que, en la palabra “dormir” se obtiene la menor precisión promedio. La sensibilidad nos indica la relación de los elementos de la clase que fueron detectados correctamente entre todos los elementos de esa clase. Los resultados muestran que la sensibilidad del modelo es similar en todas las palabras (alrededor del 60 %).

La Figura 5 contiene las matrices de confusión de cada uno de los participantes con mejor desempeño en cada uno de los métodos. Se observa que para las combinaciones PSD+LDA, PSD+SVM y PSD+RF la palabra que mejor se reconoce es “agua”, resaltando esa clase entre las restantes de la diagonal con una exactitud igual o por encima de 80 %, mientras que el resto de los valores de la diagonal oscilan entre el 15 % y 65 %.

En contraste con lo anterior, utilizando la EEGNet la distribución de la exactitud en la diagonal es uniforme, es decir, solamente toma valores entre el 75 % y 95 %. La palabra que mejor se reconoce con este método es “no”, seguida de “si” y “agua”. En general, utilizando los cuatro métodos, la palabra que más se confunde con las demás es “dormir”, seguida de “comida”.

4. Conclusiones

En este trabajo se evaluó el desempeño de la red neuronal convolucional EEGNet y de tres métodos de clasificación tradicionales (LDA, SVM con kernel lineal y RF) en la

decodificación de palabras pronunciadas a través de señales adquiridas mediante EEG. Se utilizó un conjunto de 5 palabras en español "si", "no", "agua", "comida" y "dormir" pronunciadas por participantes sanos.

Para los clasificadores tradicionales se realizó una extracción de características basada en la densidad de potencia espectral de cinco bandas de frecuencia. La clasificación multiclase (nivel de azar del 20 %) se llevo a cabo por participante. El método con el mejor desempeño fue la EEGNet, obteniendo un promedio de 75.05 ± 7.30 % calculado con todos los participantes, donde el valor máximo fue 84.00 ± 4.54 % y el mínimo 60.00 ± 8.48 %. Utilizando este método, se encontró que la palabra que mejor clasifica el modelo es "agua" y la que menos reconoce es "dormir".

Los resultados indican que una red neuronal convolucional como la EEGNet tiene mejor desempeño que clasificadores tradicionales en la decodificación de palabras pronunciadas. Este trabajo presenta resultados prometedores en el uso de redes neuronales convolucionales como la EEGNet, para la decodificación de palabras a partir de señales de EEG. Este estudio preliminar nos otorga fundamentos para el desarrollo de un estudio posterior donde se realice la decodificación del intento del habla en pacientes con ALS utilizando un método no-invasivo como el EEG.

Referencias

1. ALS news today, how common is ALS? (2019) alsnewstoday.com/how-common-is-als/
2. Angrick, M., Ottenhoff, M. C., Diener, L., Ivucic, D., Ivucic, G., Goulis, S., Saal, J., Colon, A. J., Wagner, L., Krusienski, D. J., Kubben, P. L., Schultz, T., Herff, C.: Real-time synthesis of imagined speech processes from minimally invasive recordings of neural activity. Communications Biology, vol. 4, no. 1 (2021) doi: 10.1038/s42003-021-02578-0
3. Barrientos Rojas, S. J., Ramirez-Valencia, R., Alonso-Vazquez, D., Caraza, R., Martinez, H. R., Mendoza-Montoya, O., Antelis, J. M.: Recognition of grammatical classes of overt speech using electrophysiological signals and machine learning. In: IEEE 4th International Conference on BioInspired Processing (2022) doi: 10.1109/bip56202.2022.10032476
4. Bickel, P., Diggle, P., Fienberg, S., Gather, U.: Springer Series in Statistics (2005)
5. Bishop, C. M., Nasrabadi, N. M.: Pattern recognition and machine learning. vol. 4, no. 4, pp. 738 (2006)
6. Caligari, M., Godi, M., Guglielmetti, S., Franchignoni, F., Nardone, A.: Eye tracking communication devices in amyotrophic lateral sclerosis: Impact on disability and quality of life. Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration, vol. 14, no. 7–8, pp. 546–552 (2013) doi: 10.3109/21678421.2013.803576
7. Consejo nacional para el desarrollo y la inclusión de las personas con discapacidad. La Esclerosis Lateral Amiotrófica ELA (2018) www.gob.mx/conadis/articulos/la-esclerosis-lateral-amiotrofica-ela?idiom=es
8. Cooney, C., Korik, A., Folli, R., Coyle, D.: Evaluation of hyperparameter optimization in machine and deep learning methods for decoding imagined speech EEG. Sensors, vol. 20, no. 16, pp. 4629 (2020) doi: 10.3390/s20164629
9. Cooney, C., Korik, A., Folli, R., Coyle, D.: Evaluation of hyperparameter optimization in machine and deep learning methods for decoding imagined speech EEG. Sensors, vol. 20, no. 16, pp. 4629 (2020) doi: 10.3390/s20164629
10. Datta, S., Boulgouris, N. V.: Recognition of grammatical class of imagined words from EEG signals using convolutional neural network. Neurocomputing, vol. 465, pp. 301–309 (2021) doi: 10.1016/j.neucom.2021.08.035

11. Goncharova, I. I., McFarland, D. J., Vaughan, T. M., Wolpaw, J. R.: EMG contamination of EEG: spectral and topographical characteristics. Clinical Neurophysiology, vol. 114, no. 9, pp. 1580–1593 (2003) doi: 10.1016/s1388-2457(03)00093-2
12. Guler, I., Ubeyli, E. D.: Multiclass support vector machines for EEG-signals classification. IEEE Transactions on Information Technology in Biomedicine, vol. 11, no. 2, pp. 117–126 (2007) doi: 10.1109/titb.2006.879600
13. Keng-Ang, K., Yang Chin, Z., Zhang, H., Guan, C.: Filter bank common spatial pattern (FBCSP) in brain-computer interface. In: IEEE International Joint Conference on Neural Networks (2008) doi: 10.1109/ijcnn.2008.4634130
14. Lee, J., Madhavan, A., Krajewski, E., Lingenfelter, S.: Assessment of dysarthria and dysphagia in patients with amyotrophic lateral sclerosis: Review of the current evidence. Muscle and Nerve, vol. 64, no. 5, pp. 520–531 (2021) doi: 10.1002/mus.27361
15. Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., Arnaldi, B.: A review of classification algorithms for EEG-based brain–computer interfaces. Journal of Neural Engineering, vol. 4, no. 2, pp. R1–R13 (2007) doi: 10.1088/1741-2560/4/2/r01
16. Moses, D. A., Metzger, S. L., Liu, J. R., Anumanchipalli, G. K., Makin, J. G., Sun, P. F., Chartier, J., Dougherty, M. E., Liu, P. M., Abrams, G. M., Tu-Chan, A., Ganguly, K., Chang, E. F.: Neuroprosthesis for decoding speech in a paralyzed person with anarthria. New England Journal of Medicine, vol. 385, no. 3, pp. 217–227 (2021) doi: 10.1056/nejmoa2027540
17. Nguyen, C. H., Karavas, G. K., Artemiadis, P.: Inferring imagined speech using EEG signals: a new approach using Riemannian manifold features. Journal of Neural Engineering, vol. 15, no. 1, pp. 016002 (2017) doi: 10.1088/1741-2552/aa8235
18. Panachakel, J. T., Ramakrishnan, A. G.: Decoding covert speech from EEG-A comprehensive review. Frontiers in Neuroscience, vol. 15 (2021) doi: 10.3389/fnins.2021.642251
19. Rezeika, A., Benda, M., Stawicki, P., Gembler, F., Saboor, A., Volosyak, I.: Brain–uellers: A review. Brain Sciences, vol. 8, no. 4, pp. 57 (2018) doi: 10.3390/brainsci8040057
20. Sarmiento, L. C., Villamizar, S., López, O., Collazos, A. C., Sarmiento, J., Rodríguez, J. B.: Recognition of EEG Signals from Imagined Vowels Using Deep Learning Methods. Sensors, vol. 21, no. 19, pp. 6503 (2021) doi: 10.3390/s21196503
21. Schölkopf, B., Smola, A. J.: Learning with kernels: Support vector machines, regularization, optimization, and beyond. MIT press (2002) doi: 10.7551/mitpress/4175.001.0001
22. Wolpaw, J. R.: Brain–computer interfaces. Handbook of clinical neurology, pp. 67–74 (2013)
23. Xanthopoulos, P., Pardalos, P. M., Trafalis, T. B.: Linear discriminant analysis. Robust Data Mining, pp. 27–33 (2013) doi: 10.1007/978-1-4419-9878-1