

Método de selección de canales efectivos orientado al reconocimiento emocional mediante señales electroencefalográficas

Stephanie Erandi Guzman Ramírez¹, Alejandro Estrella Lopez¹, Antonio Ramos Robles¹, Oscar Almanza Conejo², Mario Alberto Ibarra Manzano²

¹ División de Ingenierías, Campus Irapuato-Salamanca, Universidad de Guanajuato. Carretera Salamanca - Valle de Santiago km 3.5 + 1.8 Comunidad de Palo Blanco, Salamanca, Gto.

² Laboratorio de Procesamiento Digital de Señales, Departamento de Ingeniería Electrónica, División de Ingenierías, Campus Irapuato-Salamanca, Universidad de Guanajuato. Carretera Salamanca - Valle de Santiago km 3.5 + 1.8 Comunidad de Palo Blanco, Salamanca, Gto.

se.guzmanramirez@ugto.mx, a.estrellalopez@ugto.mx, a.ramosrobles@ugto.mx, o.almanzaconejo@ugto.mx, ibarram@ugto.mx

Resumen

El presente estudio aborda el reconocimiento automático de emociones humanas a partir de señales electroencefalográficas (EEG), en el marco de la computación afectiva y las interfaces cerebro-computadora. Basado en las teorías emocionales de Ekman, se diseñó una metodología centrada en el uso del conjunto de datos SEED-VII. Este conjunto de datos incluye registros multimodales (EEG y seguimiento ocular) recolectados en un diseño experimental controlado, destinado a elicitarse seis emociones básicas y un estado neutro.

La metodología consistió en la estandarización de archivos ".mat" mediante un conjunto de instrucciones escritas en el lenguaje de programación de Python que corrigen variables mal nombradas, seguido de la extracción de 22 características por canal EEG a través de 62 electrodos. Posteriormente, se aplicó el algoritmo MRMR (Minimum Redundancy Maximum Relevance) en dos fases: primero para la selección de características discriminativas y luego para identificar los canales más relevantes, permitiendo la formación de matrices optimizadas para clasificación.

Los resultados muestran, en la redacción de un pseudocódigo, un alto porcentaje de efectividad en el modelo de 62 canales, destacando especialmente cuatro canales en la discriminación de emociones, representadas en una gráfica canales-afectividad. Además, el análisis MRMR permitió identificar los electrodos con mayor peso informativo, lo que optimiza el diseño de futuras interfaces EEG.

En discusión, se destaca la robustez del procesamiento automatizado y la viabilidad del modelo propuesto para su integración en sistemas inteligentes de interacción humano-máquina. La confiabilidad de los datos EEG, su carácter involuntario y la capacidad de modelado dinámico de las redes neuronales consolidan este enfoque como una herramienta eficaz en el análisis de estados afectivos, contribuyendo al avance de tecnologías empáticas basadas en señales neurofisiológicas.

Palabras clave: mRMR; Minimum Redundancy Maximum Relevance (Maxima Relevancia Minima Redundancia). EEG; Electroencephalogram (Electroencefalograma).

Introducción

Las emociones desempeñan un papel fundamental en la vida humana, ya que mediante nuestra interacción con el entorno influyen en la percepción, el pensamiento y la conducta. Según Schmidt-Atzert (1985), las emociones implican “un vivenciar subjetivo, una reacción fisiológica y un comportamiento”, lo que permite a cada ser humano ser consciente del estado en el que se encuentra. Además, el estudio emocional es un proceso que debe considerar la estrecha interacción entre las dimensiones afectiva y cognitiva; esta interacción sensibiliza la dimensión emocional (sentimiento), las dimensiones fisiológicas (cambios corporales internos), las dimensiones expresivas/motoras (manifestaciones conductuales externas) e implica una dimensión cognitiva (funcionamiento mental) [Palmero, 1997].

Ekman sostiene que existen seis emociones básicas: alegría, tristeza, ira, miedo, disgusto y sorpresa, estas se asocian con las expresiones faciales universalmente innatas. Ampliando el análisis de las emociones básicas de Ekman, nuestro modelo incluye el estudio de estas principales emociones y del estado neutro (Jiang *et al.*, 2025). Clasificadas en no fisiológicas, las expresiones faciales, los movimientos corporales y el habla son ejemplos de este tipo de señales. Adicionalmente, las señales fisiológicas, que se encuentran en el sistema nervioso autónomo (SNA) del cerebro humano, no se pueden activar de forma consciente o intencional (Lin & Li, 2023).

La Computación Afectiva es un campo de investigación emergente enfocado en desarrollar sistemas y dispositivos capaces de reconocer, interpretar, procesar y simular emociones humanas, con el objetivo de optimizar la interacción entre los usuarios y las computadoras (García-Hernández *et al.*, 2024). En conjunto, la Interfaz Cerebro-Computadora (BCI) es un sistema que facilita la comunicación e interacción con el mundo externo mediante actividades eléctricas cerebrales individuales con efectos artificiales (Houssein *et al.*, 2022).

Un electroencefalograma (EEG) aplicado en este proceso, permite monitorear o almacenar la actividad de la corteza cerebral, para el desarrollo de múltiples tecnologías tales como el reconocimiento emocional, manifestando la comunicación de los elementos electrofisiológicos, obtenidos del usuario [Kriesel, 2021]. La implementación de esta herramienta se realiza una vez que el participante es estimulado; se lee la señal de EEG directamente desde el cuero cabelludo mediante múltiples electrodos. Por lo que las señales de EEG, al no poder ser manipulables, no pueden simular un estado emocional, por lo que la información proporcionada es fiable, sin embargo, su dificultad está en la decodificación de esta información, ya que tiene que ser asociada a emociones muy específicas. En línea con estos avances, el reconocimiento automático de emociones mediante señales EEG se perfila como una herramienta con gran potencial para fortalecer las relaciones entre las personas y su entorno. Asimismo, esta tecnología desempeña un papel clave en el desarrollo de sistemas de inteligencia artificial orientados a la interacción humana, al permitir una comprensión más profunda y empática de los estados emocionales, demostrando un 80% de precisión en sus clasificaciones (Li *et al.*, 2022). Tecnologías tales como las redes neuronales artificiales constituyen una técnica de aprendizaje automático inspirada en la estructura del cerebro humano. Estas redes son capaces de modelar relaciones no lineales complejas entre entradas y salidas, lo que las hace adecuadas para tareas como la clasificación de emociones a partir de señales EEG, que adaptan dinámicamente cada situación que se le aplique un algoritmo, las redes neuronales artificiales pueden ser rápidas al momento de resolver un problema, pero sus capacidades son limitadas. [Jiang, 2024]. Sin embargo, el principio de incertidumbre persiste en la calidad y cantidad de datos introducidos a los algoritmos de aprendizaje computacional para resolver alguna problemática. Las condiciones no controladas durante la adquisición de señales, como factores ambientales, psicológicos, de salud e incluso familiares, pueden afectar la calidad de la información monitoreada o grabada, lo que introduce valores atípicos o ruido en el estudio. Para solucionar esta problemática, algunos autores sugieren el uso de un subconjunto reducido de datos que permite lograr resultados similares que, empleando, en el caso de señales EEG, todo el conjunto de la matriz de información de los canales electroencefalográficos. De este modo, mediante este subconjunto de información EEG, se reduce la incorporación de valores atípicos en el proceso de extracción de características.

Bajo este escenario, se propone un algoritmo para la selección de canales EEG efectivos asociados a múltiples estados emocionales. Estos estímulos han sido previamente seleccionados y capturados por un grupo de científicos expertos en el área de las emociones para crear una base de datos orientada al reconocimiento emocional. Para ello, se desarrolló una propuesta de algoritmo, centrada en las seis emociones básicas (felicidad, tristeza, miedo, disgusto, sorpresa e ira) y la neutra, permitiendo una interpretación de canales más relevantes de estas emociones a partir de una estimación global de la base de datos de dichas emociones.

Estudio del arte

En este apartado se presenta una revisión de los trabajos más relevantes que han abordado el reconocimiento emocional mediante señales EEG, con énfasis en las metodologías utilizadas para representar las emociones y en la identificación de características discriminativas. SEED-VII nos proporciona mediante EEG, movimiento ocular y un activo con 62 canales en un sistema internacional 10-20 y rastreador ocular Tobii Pro Fusion, demostró un nuevo conjunto de datos multimodales de las emociones básicas y el estado neutral, promoviendo MAET procesador de entradas unimodales y multimodales demostrando mejora en capacidad de discriminación de emociones del enfoque propuesto [Jiang, 2024]. Asimismo, Bolta *et al.* (2024). En 2024, se desarrolla un clasificador de emociones basado en aprendizaje automático que utiliza datos EEG. Modelo de aprendizaje profundo avanzado, con Redes neuronales convolucionales (CNN) y memoria a corto y largo plazo (LSTM); Aplico cinco bandas de frecuencia: delta (1-4 Hz), theta (4-8 Hz), alpha (8-14 Hz), beta (14-31 Hz) y gamma (31-50 Hz), generando un archivo de 310 características utilizando 62 canales, finalizando con un mapeo de etiquetas emocionales a valores numéricos a la par existe una combinación de todos los datos en un único DataFrame la cual proporciona una base de datos estructurada para la exploración de correlaciones en características EEG y estados emocionales [Bolta, 2024].

Por otro lado, García (2020) aplico la herramienta Emotiv EPOC, para el análisis de emociones de los usuarios mediante EEG provocadas por estímulos visuales, realizado con 14 electrodos divididos en sensores y 2 electrodos de referencia, con un sistema internacional 10-20 de EEG, Emotiv EPOC se posicionaba al ritmo de 128 muestras por segundo (2048 Hz de frecuencia base) el cual ayudo a detectar la conexión inalámbrica de la diadema y llave USB a cada electrodo, para ser almacenada y clasificada a un rango de 0 a 128, sin aplicación alguna de filtros, que posteriormente eran aplicados por un filtro de paso alto <0,2Hz ; en un paquete de análisis sólido como EEGLab.

El objetivo principal de la clasificación de emociones es desarrollar un modelo de reconocimiento emocional que identifique con precisión las emociones sin depender de la intensidad de los estímulos. Algunos enfoques en este ámbito utilizan los conjuntos de datos SEED y SEED-V, los cuales se basan principalmente en la recolección de datos de participantes chinos, aunque el conjunto SEED también incorpora sujetos de Alemania y Francia. En este sentido, en el estudio de Liu y autores (Liu *et al.*, 2022), se estableció una tarea de clasificación para estos tres grupos de participantes, llegando a la conclusión de que la característica de entropía diferencial obtuvo resultados superiores en el conjunto de datos chino en comparación con los otros dos. Asimismo, indicaron que los sujetos alemanes y franceses podrían compartir patrones emocionales similares, a diferencia de los chinos.

Metodología

La metodología seguida en este estudio se basa en el uso del conjunto de datos SEED-VII, el cual está diseñado para el reconocimiento de emociones a partir de señales EEG y de movimiento ocular. El siguiente gráfico muestra un resumen de la estructura experimental, que incluye estimulación emocional, participantes, diseño experimental y adquisición de señales.

Descripción General del Conjunto de Datos SEED-VII	
Estimulación Emocional	Participantes
- Se usaron videos (auditivo-visuales) para provocar 6 emociones básicas (alegría, tristeza, miedo, ira, asco, sorpresa) y un estado neutral. - Se recopilaron 80 videoclips (12 por emoción, excepto neutralidad con 8). - Los videos fueron seleccionados y puntuados por voluntarios (1-5 puntos) según su efectividad emocional. - Duración de los videos: 2 a 5 minutos; total = 14,098 segundos.	20 voluntarios (10 hombres, 10 mujeres), edad 19-26 años (media 22.5) - Todos diestros, con visión y audición normales. - Selección de participantes mediante el cuestionario de Personalidad de Eysenck (EPQ) - Se priorizó a los de alta extraversión, por su mejor capacidad de percibir emociones.
Diseño Experimental	Adquisición de Señales
4 sesiones experimentales por participante. - Cada sesión: 20 ensayos (ver video + autoevaluación emocional con una clasificación de nivel de intensidad emocional del 1-10). - Solo se inducían 5 de 7 emociones por sesión para evitar sobrecarga emocional.	- Se registraron señales EEG y de movimiento ocular simultáneamente. - EEG: Utilizo un sistema ESI NeuroScan a una frecuencia de muestreo de 1000 Hz, activo de 62 canales con un sistema internacional 10-20. - Eye-tracking: Con un rastreador ocular Tobii Pro Fusion, 250 Hz.

Figura 1. Descripción General del Conjunto de Datos SEED-VII.

Extraer características y etiquetas de matrices de celdas en matrices para MRMR

En la figura 2 se describen los pasos para convertir las características y etiquetas, originalmente almacenadas en estructuras de celdas, en matrices para la implementación de un algoritmo de selección de características empleando el método mRMR (Mínima Redundancia Máxima Relevancia, [Ding & Peng, 2005]). El procedimiento inicia creando matrices vacías para almacenar las características y etiquetas. Luego, dependiendo del conjunto de datos seleccionado (SEED-V o SEED-VII), se extraen y organizan los datos de forma estructurada, manteniendo la correspondencia entre cada muestra, canal y etiqueta. Esto permite una posterior evaluación uniforme de la importancia de los canales EEG.

Algoritmo 1: Descriptores de características y conjuntos de datos

```

1  $nCaract \leftarrow$  número de características computadas;
2  $matrizCaract \leftarrow$  ceros( $nArchivos * nEnsayos * nCanales, nCaract$ );
3  $vectorEtiquetas \leftarrow$  ceros( $nArchivos * nEnsayos * nCanales, 1$ );
4 /* Selección del dataset a usar */
5 para  $idDataset \leftarrow 1$  a número de casos hacer
6     seleccionar Conjunto de datos hacer
7         caso Seed V hacer
8             extraerSeedV()
9         fin caso
10        caso Seed VII hacer
11            extraerSeedVII()
12        fin caso
13    fin seleccionar
14 fin para

```

Figura 2. Descriptores de características.

La función describe el procesamiento de datos del conjunto SEED-V. Se cargan los archivos de datos y etiquetas para 16 sujetos, cada uno con 15 ensayos. A partir de las señales EEG crudas, se extraen matrices de características por canal utilizando una frecuencia de muestreo de 1000 Hz. Las etiquetas correspondientes se vinculan a cada conjunto de características, estructurando así los datos para su análisis posterior, como la selección de canales efectivos, como se muestra en la figura 3.

Algoritmo 2: Función para procesamiento de datos SEED V

```

1 Función extraerSeedV
2      $archDatos \leftarrow$  archivos de datos desde disco;
3      $nArchivos \leftarrow 16$ ;
4      $archEtiquetas \leftarrow$  archivos de etiquetas desde disco;
5      $fs \leftarrow 1000$ ;
6      $nEnsayos \leftarrow 15$ ;
7      $nCanales \leftarrow 62$ ;
8      $nombresCanales \leftarrow$  nombres de canales EEG;
9     /* Extraemos las características de todo el dataset */
10    para  $i \leftarrow 1$  a  $nArchivos$  hacer
11         $datos \leftarrow$  cargar( $archDatos$ );
12         $etiquetas \leftarrow$  cargar( $archEtiquetas$ );
13        para  $j \leftarrow 1$  a  $nEnsayos$  hacer
14             $eeg \leftarrow$  señal EEG;
15            /* ProcesarSeñal es una función definida por el
16               usuario */
17             $matrizCaract \leftarrow$  procesarSeñal( $eeg, nCanales, fs, args$ );
18            /* Argumentos adicionales pasados a la función */
19             $vectorEtiquetas \leftarrow$  obtenerEtiquetas( $\bullet$ );
20        fin para
21    fin para
22 fin Función

```

Figura 3. Función para procesamiento de datos SEED V.

De manera similar a SEED V, se cargan señales EEG y etiquetas de 20 sujetos, cada uno con múltiples sesiones y carpetas, lo que genera varios ensayos por sujeto. Las señales, muestreadas a 200 Hz, se segmentan en ventanas y se extraen matrices de características por canal, como se muestra en la figura 4. Las etiquetas correspondientes se asignan a cada segmento, estructurando los datos para su uso en análisis como selección de características.

Algoritmo 3: Función para procesamiento de datos SEED VII

```

1 Función extraerSeedVII
2   archDatos ← archivos de datos;
3   archEtiquetas ← archivos de etiquetas;
4   nSujetos ← 20;
5   fs ← 200;
6   longVentana ← un cuarto de fs;
7   [nSesiones, nCarpetas] ← tamaño de la matriz de etiquetas;
8   nEnsayos ← nSesiones × nCarpetas;
9   nCanales ← 62;
10  nombresCanales ← nombres de canales EEG;
11  /* Extraemos las características de todo el dataset */
12  para i ← 1 a nSujetos hacer
13    datos ← cargar(archDatos);
14    etiquetas ← cargar(archEtiquetas);
15    para j ← 1 a nEnsayos hacer
16      eeg ← señal EEG;
17      /* ProcesarSeñal es una función definida por el
18         usuario */
19      matrizCaract ← procesarSeñal(eeg, nCanales, fs, args);
20      /* Argumentos adicionales pasados a la función */
21      vectorEtiquetas ← obtenerEtiquetas(j);
22    fin para
23  fin para
24 fin Función

```

Figura 4. Función para procesamiento de datos SEED VII.

Para cada canal de EEG, se desarrolla un método de extracción de características descriptivas de cada señal. Para lograr esto, primero, la matriz de 62 canales de EEG, obtenida de cada estímulo, por cada prueba y por cada sujeto, se procesa mediante un algoritmo basado en la transformada wavelet. Este método disminuye la cantidad de ruido en una señal de entrada a partir de la ecuación

$$W(s, \tau) = \frac{1}{\sqrt{s}} \int_{-\infty}^{\infty} \overline{\psi\left(\frac{t-\tau}{s}\right)} x(t) dt,$$

donde $\overline{\psi\left(\frac{t-\tau}{s}\right)}$ representa la transformada del núcleo conjugado complejo; s y τ representan los factores de escala y desplazamiento temporal, respectivamente. Según lo propuesto en este documento, la reducción de ruido por wavelets se realiza con la función *wdenoise* de MATLAB. Posteriormente se realiza una normalización de valores entre 0 y 1 para realizar una estimación de patrones locales binarios (Local Binary Pattern, LBP) siguiendo el criterio

$$LBP(t) = \sum_{p=1}^{\tau} s(\delta) * 2^{p-1},$$

donde

$$s(\delta) = \begin{cases} 1, & \text{si } EEG(t + \tau) \geq EEG\left(\left\lceil \frac{t + \tau}{2} \right\rceil\right) \\ 0, & \text{cualquier otro caso} \end{cases}$$

Donde, en este contexto, la longitud de la ventana se refiere al tamaño para analizar el LBP, tomando el valor central $EEG\left(\left\lceil \frac{t + \tau}{2} \right\rceil\right)$ de la ventana actual $EEG(t + \tau)$ como umbral. Los valores máximo y mínimo del LBP dependerán de la longitud τ . Para los propósitos de este estudio, utilizamos la configuración clásica de 8 bits. Este proceso se repite para cada ventana en el tensor de entrada. Un ejemplo del procesamiento de LBP se muestra en la figura 5. La figura 6 muestra el pseudocódigo empleado en este proceso.

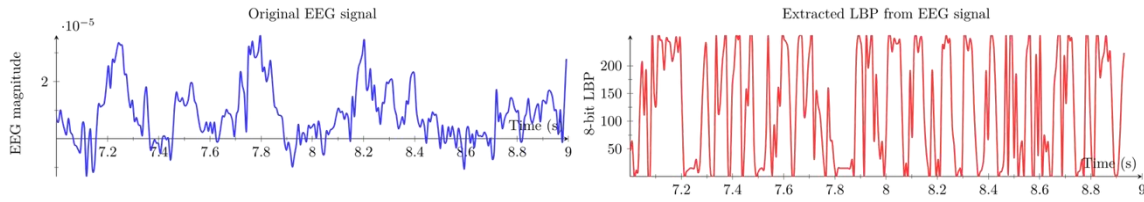


Figura 5. Comportamiento en el dominio del tiempo del paso de una señal EEG a LBP.

Algoritmo 4: Procesado de la señal mediante filtro wavelet

```

1 Función procesarSeñal(EGG, nCanales, fs):
2   nivWavelet  $\leftarrow$  floor(log2(fs));
3   /* Filtramos la señal EEG */
4   filtradaEGG  $\leftarrow$  wdenoise(EGG, nivWavelet);
5   /* Encontramos los Patrones Locales Binarios con una
   funcion definida por el usuario */
6   matrizLBP  $\leftarrow$  Vec2LBP(filtradaEGG, 3);
7   /* normalizamos matrizLBP */
8   para cn  $\leftarrow$  1 a nCanales hacer
9     cnLBP  $\leftarrow$  matrizLBP(cn, :);
10    matrizLBP(cn, :)  $\leftarrow$ 
      (cnLBP - min(cnLBP)) / (max(cnLBP) - min(cnLBP));
11  fin para
12  /* Extraemos las características de la señalEGG en otra
   funcion */
13  caracteristicas  $\leftarrow$  extraerCaracteristicasEGG(matrizEGG);
14  devolver caracteristicas

```

Figura 6. Procesado de la señal mediante filtro wavelet

Algoritmo 5: Cálculo de los patrones locales binarios

```

1 Función Vec2LBP(codigoLBP, nSegmentos):
2   sMatriz  $\leftarrow$  codigoLBP;
3   radioVentana  $\leftarrow$  nSegmentos;
4   longitudVentana  $\leftarrow$  radioVentana2;
5   pesosLBP  $\leftarrow$  [1, 2, 4, 8, 0, 16, 32, 64, 128];
6   [nCanales, ]  $\leftarrow$  size(sMatriz);
7   solaparVentana  $\leftarrow$  longitudVentana - 1;
8   nSegmentos = length(sMatriz) - solaparVentana;
9   segIdxs  $\leftarrow$  1 : nSegmentos;
10  segmentMatrix  $\leftarrow$  (segIdxs)-1 + (zeros(1, longitudVentana));
11  compVector  $\leftarrow$  (zeros(nSegmentos, 1))-1 + (0 :
      longitudVentana - 1)-1;
12  compVector  $\leftarrow$  offsetVector-1;
13  segmentComp  $\leftarrow$  segmentMatrix + compVector;
14  codigoLBP  $\leftarrow$ 
      zeros(nCanales, length(sMatriz) - solaparVentana);
15  para cn  $\leftarrow$  1 a nCanales hacer
16    sCanal  $\leftarrow$  sMatriz(cn, :);
17    canalVentana  $\leftarrow$  sCanal(segmentComp);
18    valoresCentrales  $\leftarrow$  canalVentana(:, 5);
19    matBinLBP  $\leftarrow$  canalVentana > valoresCentrales;
20    CodigoLBP(cn, :)  $\leftarrow$  matBinLBP * pesosLBP-1;
21  fin para
22  devolver [codigoLBP, nSegmentos]

```

Figura 7. Cálculo de los patrones binarios locales

Como complemento al pseudocódigo mostrado en la figura 6, la figura 7 modela el proceso de extracción de características estadísticas, tiempo-frecuencia y morfológicas para cada canal de una señal EEG. Se incluyen medidas temporales como la media, varianza, RMS, sesgo y curtosis, así como métricas derivadas de la señal como el contenido de energía, entropía diferencial y factores de forma. Las diferencias entre muestras sucesivas también se usan para extraer descriptores como la movilidad de Hjorth, HMO e IDM. El resultado es una matriz de tamaño [*nCanales* $\times$ 22], donde cada fila resume cuantitativamente un canal EEG, permitiendo su posterior análisis o clasificación. La sección apreciada en la figura 8 implementa la selección de características relevantes para clasificación emocional usando el algoritmo mRMR (Minimum Redundancy Maximum Relevance). Primero, se ensamblan las matrices globales de características y etiquetas desde

estructuras de celdas. Luego, se eliminan entradas con etiquetas faltantes y se aplica MRMR para identificar las características más informativas. Estas se filtran por categoría emocional y sus puntajes de importancia se normalizan. Posteriormente, se reorganizan para evaluar la relevancia de cada canal EEG. Finalmente, se identifican los canales más representativos mediante MRMR sobre la matriz reorganizada, lo que permite reducir la dimensionalidad al enfocarse en los canales más efectivos.

Algoritmo 6: Extraemos los 22 canales del encefalograma.

```

1 Función extraerCaracteristicasEGG(matrizEGG):
2   [nCanales, nMuestras]  $\leftarrow$  size(matrizEGG);
3   mtrCaracter  $\leftarrow$  zeros(nCanales, 22);
4   /* Designamos una variable con la distancia desde 1.0 al
      siguiente número de doble precisión mayor. */
5   valMin  $\leftarrow 2^{-52}$ ;
6   para cnIdx  $\leftarrow$  1 a nCanales hacer
7     sCanal = matrizEGG(cnIdx, :)-1;
8     sAbsoluta = abs(sCanal);
9     sDiferencia = diff(sCanal);
10    sPromedio = mean(sCanal);
11    sRMS = sqrt(mean(sCanal.2));
12    maxValor = 0,5 * (max(sCanal) - min(sCanal));
13    sDesEstnd = std(sCanal);
14    caracter  $\leftarrow$  zeros(1, 22);
15    caracter(1)  $\leftarrow$  mean(abs(sDiferencia));
16    caracter(2)  $\leftarrow$  var(sCanal);
17    caracter(3)  $\leftarrow$  sum(sqrt(sAbsoluta));
18    caracter(4)  $\leftarrow$  maxValor / (mean(sqrt(sAbsoluta)))2;
19    caracter(5)  $\leftarrow$  sRMS;
20    caracter(6)  $\leftarrow$  maxValor / sRMS;
21    caracter(7)  $\leftarrow$  sRMS / mean(sAbsoluta);
22    caracter(8)  $\leftarrow$  exp(mean(log(sAbsoluta + valMin)));
23    caracter(9)  $\leftarrow$  std(sDiferencia) / stdSignal;
24    caracter(10)  $\leftarrow$  sum(sAbsoluta);
25    caracter(11)  $\leftarrow$  sPromedio;
26    caracter(12)  $\leftarrow$  skewness(sCanal);
27    caracter(13)  $\leftarrow$  kurtosis(sCanal);
28    caracter(14)  $\leftarrow$  mean(abs(sCanal - sPromedio));
29    caracter(15)  $\leftarrow$  mean(abs(sDiferencia));
30    HMOtemp  $\leftarrow$  1 / (1 + abs(sDiferencia))
      caracter(16)  $\leftarrow$  mean(HMOtemp);
31    IDMtemp  $\leftarrow$  1 / (1 + abs(sDiferencia))
      caracter(17)  $\leftarrow$  mean(IDMtemp);
32    caracter(18)  $\leftarrow$  0,85 * sum(sCanal);
33    caracter(19)  $\leftarrow$  (-0,5) * log2(sum((sCanal / (nMuestras *
      sPromedio)).3 + valMin));
34    caracter(20)  $\leftarrow$  (sum(sqrt(sAbsoluta)))2;
35    caracter(21)  $\leftarrow$  sum(sDiferencia);
36    caracter(22)  $\leftarrow$  (0,5) * log(2 * pi * exp(1) * var(sCanal));
37    matCaracter(cnIdx, :) = caracter
38  fin para
39  devolver matCaracter

```

Figura 8. Extracción de características estadísticas, tiempo-frecuencia, y morfológicas.

Esta sección describe el proceso de extracción y selección de características relevantes mediante el método mRMR. Primero, se organiza la información de características y etiquetas mediante una estructura de celdas en matrices numéricas. Posteriormente, se eliminan valores nulos y se aplica MRMR para identificar las características más informativas. Finalmente, se evalúa la importancia de cada canal mediante los puntajes

MRMR, lo que permite identificar los canales EEG más relevantes para tareas de clasificación emocional, como se muestra en el pseudocódigo de la figura 9.

Algoritmo 7: Extraer características y etiquetas MRMR

```

1  $nCaract \leftarrow$  número de características computadas;
2  $matrizCaract \leftarrow$  ceros( $nArchivos * nEnsayos * nCanales, nCaract$ );
3  $vectorEtiquetas \leftarrow$  ceros( $nArchivos * nEnsayos * nCanales, 1$ );
4 /* Extrae el peso de cada etiqueta */
5 para  $idArchivo \leftarrow 1$  a  $nArchivos$  hacer
6    $matrizCaract[(idArchivo-1)(nEnsayos*nCanales) + 1 :$ 
    $idArchivo(nEnsayos*nCanales),:] \leftarrow$   $caract[idArchivo]$ ;
7   para  $idEnsayo \leftarrow 1$  a  $nEnsayos$  hacer
8      $vectorEtiquetas[.] \leftarrow$   $lbls[idArchivo][idEnsayo]$ ;
9   fin para
10 fin para
11 /* Selecciona los pesos mas importatantes */
12  $cNulos \leftarrow$  celdas vacias de  $vectorEtiquetas$ ;
13  $nNulos \leftarrow$  numero de  $cNulos$ ;
14  $mtxDatosLBP(cNulos, :) \leftarrow []$ ;
15  $EtiquetasUnicas \leftarrow$   $vectorEtiquetas$  únicas;
16  $[idxEmp, puntEmp] \leftarrow$ 
    $fscmr(mtxDatosLBP, vectorEtiquetas, 'Verbose', 1)$ ;
17  $caractMasImport \leftarrow$   $features(idxEmp(1 : 3))$ ;
18 para  $i \leftarrow 1$  a  $EtiquetasUnicas$  hacer
19    $caractMasImport \leftarrow$   $mtxDatosLBP(categorical(EtiquetasUnicas(i)) ==$ 
    $vectorEtiquetas, idxEmp(1 : 3))$ ;
20 fin para
21  $puntEmpNorm \leftarrow$ 
    $(puntEmp - \min(puntEmp)) / (\max(puntEmp) - \min(puntEmp))$ ;
22  $idx \leftarrow$   $find(puntEmp \geq 0.5)$ ;
23 si  $idx$  entonces
24    $idx \leftarrow$   $find(puntEmp \geq puntEmp(idxEmp(6)))$ ;
25 fin si
26 /* Selecciona los datos en las celdas de los pesos mas
   importantes */
27  $nCaract \leftarrow$   $length(idx)$ ;
28  $pesoCanales \leftarrow$ 
    $repmat(pesoCanales', [nArchivos * nEnsayos - nNulos, 1])$ ;
29  $mrmlrLBP \leftarrow$   $mtxDatosLBP(:, idx)$ ;
30  $matrizCaractCanal =$ 
    $zeros((nArchivos * nEnsayos - nNulos) * nCaract, nCanales)$ ;
31 para  $i \leftarrow 1$  a  $200 - nNulos$  hacer
32    $matrizCaractCanal[(i-1) * nCaract + 1 : i * nCaract, :] \leftarrow$ 
    $seleccionarCaract(:, (i-1) * nCanales + 1 : i * nCanales)$ ;
33 fin para
34  $[idxCanalSel, puntCanalSel] \leftarrow$ 
    $fscmr(matrizCaractCanal, pesoCanales, 'Verbose', 1)$ ;
35  $canales \leftarrow$   $nombresCanales$  res. $idxCanalesEfectivos \leftarrow$   $idxCanalSel$ 
   res. $puntajesNormalizados \leftarrow$   $puntCanalSel$ 

```

Figura 9. Extracción de características y etiquetas MRMR

Este procesamiento permite obtener una relevancia asociada a cada característica, que posteriormente se emplea como descriptor de características para cada canal, mediante el reacomodo de la matriz inicial de características a una matriz de dimensión [44640×13]. Para este caso, cada canal de EEG actuará como una

característica para que el algoritmo mRMR calcule la relevancia de cada canal de entrada. Es claro que este proceso puede aplicarse no solo a señales EEG, sino también a cualquier base de datos en la que el objetivo sea calcular la relevancia de un conjunto de datos de entrada provenientes de la misma fuente de adquisición.

En la figura 10 se muestra la sección que grafica los puntajes de importancia de cada canal EEG obtenidos con MRMR. Primero, normaliza los puntajes entre 0 y 1. Luego, genera una gráfica de barras que destaque los cuatro canales más relevantes. También se crea una leyenda con sus valores y se exporta la figura en formato EPS. Finalmente, se guardan los puntajes normalizados en una tabla de resultados para su posterior análisis.

Algoritmo 8: Graficar pesos de canales

```

1 [idxCanal, puntajes] ←
  fsmrmr(matrizCaractCanal, etiquetasCanales, Verbose = 1);
2 res.idxCanalesEfectivos ← idxCanal;
3 res.puntajesCanalesEfectivos ← puntajes;
4 canales ← nombresCanales;
5 vecCanales ← 1 a nCanales;
6 /* Normalizar puntajes entre 0 y 1 */
7 puntajes ←
  (puntajes - mín(puntajes))/(máx(puntajes) - mín(puntajes));
8 res.puntajesNormalizados ← puntajes;
9 /* Obtener los 4 mejores canales */
10 res.top4Canales ← canales[puntajes ≥ puntajes[idxCanal[4]]];
11 figura ← crearFigura(maximizada);
12 barra ← barras(vecCanales, puntajes, 'azul');
13 /* Configurar ejes y etiquetas */
14 marcasX ← vecCanales;
15 etiquetasX ← canales;
16 leyenda[1] ← 'Magnitud de canales efectivos';
17 para i ← 1 a 4 hacer
18   punto ← (idxCanal[i], puntajes[idxCanal[i]]);
19   marcador ← marcador(punto, 'círculo', grande);
20   leyenda[i + 1] ← unir(canales[idxCanal[i]], ':',
    , puntajes[idxCanal[i]]);
21 fin para
22 /* Etiquetas globales */
23 etiquetaX ← 'Etiqueta del canal';
24 etiquetaY ← 'Magnitud efectiva normalizada';
25 /* Exportar figura y guardar resultados */
26 exportarFigura(figura, 'ec.i1.eps');
27 out[idDataset] ← res;
28 tabla ← tabla(nombresCanales,
  vecCanales, res.puntajesNormalizados);
29 guardar(tabla, [dataset_canalesEfectivos.txt]);

```

Figura 10. Grafica de los pesos de los canales.

Resultados

Tras la extracción de características del conjunto de datos EEG (SEED V y SEED VII), se aplicó el método mRMR para evaluar la relevancia de las características respecto a las etiquetas emocionales. Como resultado, se identificaron los índices de las características más relevantes y se normalizaron sus puntajes entre 0 y 1 para facilitar la comparación. Posteriormente, se construyó una matriz extendida canal-característica, que permitió evaluar la importancia de cada canal en función de las características seleccionadas. A través de mRMR, se obtuvieron los puntajes de cada canal, lo que permitió generar un vector de pesos normalizados que representa la contribución relativa de cada canal en un espacio global multi-emocional, es decir, que modela la relevancia de cada canal para todas las emociones del conjunto de datos.

Descripción general de funcionamiento

Este script recorre de forma recursiva todos los subdirectorios de una carpeta especificada, identificando los archivos con extensión “.mat”. Posteriormente, cada archivo es procesado por la función `fix_mat_file`, la cual realiza las siguientes operaciones:

1. Carga de archivos: Con la función `scipy.io.loadmat()`, importa el archivo con el contenido encontrado en “.mat” como un diccionario de Python.
2. Filtrado de datos internos: Se eliminan automáticamente las claves reservadas de MATLAB que no corresponden a datos útiles para el análisis.
3. Validación y corrección de nombres de variables: Cada variable es inspeccionada; si el nombre comienza con un dígito, se le antepone el prefijo `TRIAL` para regularizarlo.
4. El almacenamiento del archivo corregido consiste en guardar una nueva versión, denominada con el sufijo `_fixed`, para preservar el archivo original sin modificaciones.

Aplicación Práctica

Este script fue utilizado como paso previo en el procesamiento del conjunto de datos SEED-VII, el cual contiene señales EEG preprocesadas y almacenadas en archivos “.mat” organizados por sujeto y ensayo; Lo que con la herramienta desarrollada, fue posible evitar errores de carga de MATLAB y otros elementos compatibles, asegurando la estandarización de los nombres de las variables, como características y análisis de señales, las cuales se especificaron que existen cuatro conjuntos de datos a procesar, donde cada celda con los veintidós descriptores de sus características son extraídas de cada canal de señal, mediante sesenta y dos electrodos de SEEDV, para posteriormente generar matrices de dos rondas (MRMR), la primera genera la selección de características, el cual convierte `channelLabels` a `categorical` y aplica `fscmr` para ordenar las 22 características según su poder discriminativo, donde también se encarga de extraer tres características top, creando una sub-matriz eligiendo aquellas con un score normalizado ≥ 0.5 , en caso de que no se supere el umbral, elige las seis primeras de la clasificación y la segunda ronda, los canales crean una matriz la cual contiene sus características, la matriz transpuesta, reorganiza sus datos de modo que cada fila represente un canal completo (todas las características seleccionadas) en un ensayo, para evaluar canales; Además, en una segunda selección de canales determina los electrodos más informativos, dándonoslos como resultados:

- `LbpFeatures`: matriz final de características usadas
- `idxEffectiveChannels` y `scoresEffectiveChannels`: clasificación y puntuaciones normalizadas de los canales más relevantes.

Como podemos visualizarla en la siguiente tabla, este bloque visualiza, destaca y guarda qué electrodos (canales) de EEG son más informativos según MRMR.



donde la trazabilidad y la reproducibilidad son prioritarias. Esta herramienta resulta especialmente valiosa en proyectos que involucran conjuntos de datos complejos como SEED-VII, donde cada sujeto puede tener múltiples sesiones y cientos de ensayos almacenados en archivos individuales. En estos casos, la corrección manual de nombres de variables sería impráctica y propensa a errores. La automatización proporcionada por este script permite escalar el procesamiento de manera segura, eficiente y replicable, eliminando cuellos de botella en la etapa inicial de carga de datos y facilitando la posterior aplicación de algoritmos de extracción de características, clasificación o análisis estadístico. Asimismo, su diseño modular y portátil permite una fácil integración en diferentes sistemas operativos y entornos de programación, ampliando su aplicabilidad a una variedad de escenarios en el análisis computacional de señales biomédicas. La inclusión de una validación robusta, el manejo de excepciones y la compatibilidad con estructuras de carpetas recursivas hacen de esta herramienta una contribución práctica y esencial en el preprocesamiento de archivos “.mat”.

En resumen, el algoritmo desarrollado no solo mitiga una limitación técnica puntual, sino que contribuye significativamente a la confiabilidad del procesamiento de datos en investigaciones neurocientíficas y de neuro tecnología, al ofrecer una solución sencilla, automatizable y reproducible para la normalización de archivos de datos.

Bibliografía/Referencias

- A Brief Introduction to neural networks. (2017). D. Kriesel.
https://www.dkriesel.com/en/science/neural_networks
- Ballester, B., & Pablo, J. (2024). Clasificación de emociones mediante señales EEG y técnicas de aprendizaje automático. Universitat Politècnica de València. <https://riunet.upv.es/handle/10251/210228>
- Ding, C., & Peng, H. (2005). Minimum redundancy feature selection from microarray gene expression data. *Journal of Bioinformatics and Computational Biology*, 3(2), 185–205.
<https://doi.org/10.1142/s0219720005001004>
- García-Hernández, R. A., Luna-García, H., Celaya-Padilla, J. M., García-Hernández, A., Reveles-Gómez, L. C., Flores-Chaires, L. A., Delgado-Contreras, J. R., Rondon, D., & Villalba-Condori, K. O. (2024). A systematic literature review of modalities, trends, and limitations in emotion recognition, affective computing, and sentiment analysis. *Applied Sciences (Basel, Switzerland)*, 14(16), 7165.
<https://doi.org/10.3390/app14167165>
- Houssein, E. H., Hammad, A., & Ali, A. A. (2022). Human emotion recognition from EEG-based brain-computer interface using machine learning: a comprehensive review. *Neural Computing & Applications*, 34(15), 12527–12557. <https://doi.org/10.1007/s00521-022-07292-4>
- Jiang, W.-B., Liu, X.-H., Zheng, W.-L., & Lu, B.-L. (2025). SEED-VII: A multimodal dataset of six basic emotions with continuous labels for emotion recognition. *IEEE Transactions on Affective Computing*, 16(2), 969–985. <https://doi.org/10.1109/taffc.2024.3485057>
- Li, J. W., Barma, S., Mak, P. U., Chen, F., Li, C., Li, M. T., Vai, M. I., & Pun, S. H. (2022). Single-channel selection for EEG-based emotion recognition using brain rhythm sequencing. *IEEE Journal of Biomedical and Health Informatics*, 26(6), 2493–2503. <https://doi.org/10.1109/JBHI.2022.3148109>
- Lin, W., & Li, C. (2023). Review of studies on emotion recognition and judgment based on physiological signals. *Applied Sciences (Basel, Switzerland)*, 13(4), 2573. <https://doi.org/10.3390/app13042573>
- Liu, W., Zheng, W.-L., Li, Z., Wu, S.-Y., Gan, L., & Lu, B.-L. (2022). Identifying similarities and differences in emotion recognition with EEG and eye movements among Chinese, German, and French People. *Journal of Neural Engineering*, 19(2), 026012. <https://doi.org/10.1088/1741-2552/ac5c8d>
- Palmero, F. (2000). Emoción. Breve reseña del papel de la cognición y el estado afectivo. *Revista Española de Motivación y Emoción*, 163–167.
- Pinzon, G., & Alexandro, G. (2020). Procesamiento de datos fisiológicos para detectar estados emocionales en el proceso de evaluación de la experiencia de usuario [Tecnológico Nacional de México]. <https://rinacional.tecnm.mx/jspui/handle/TecNM/4172>
- Schmidt-Atzert, L. (1985). Psicología de las emociones. Editorial Herder.