

# Detección de señales EEG epilépticas utilizando redes convolucionales basada en la transformada *synchrosqueezing* acolchada

Sergio Villazana<sup>1</sup>, Guillermo Montilla<sup>2</sup>, Antonio Eblen<sup>3</sup>, Carlos Maldonado<sup>1</sup>

<sup>1</sup> Universidad de Carabobo,  
Centro de Procesamiento de Imágenes,  
Venezuela

<sup>2</sup> Yttrium-Technology Corp.,  
Panamá

<sup>3</sup> Universidad Diego Portales, Facultad de Medicina, Laboratorio de Neurociencia Translacional,  
Chile

montillaleon@yttrium-technology.com

**Resumen.** Este trabajo propone un algoritmo basado en redes neuronales convolucionales para clasificar señales electroencefalográficas (EEG) en las clases normal, preictal e ictal, como apoyo para el especialista médico para facilitar el diagnóstico de la condición de epilepsia. Las señales EEG se pre-procesan mediante la aplicación de la transformada *synchrosqueezing* basada en la transformada corta de Fourier acolchada (SS-QSTFT de sus siglas en inglés), que genera como salida una representación tiempo-frecuencia que se utiliza como entrada a la red neuronal convolucional. El entrenamiento de los clasificadores se realizan con los registros de la base de datos EEG de la Universidad de Bonn, compuesta de cinco conjuntos identificados como A, B, C, D y E. Las clases normal, preictal e ictal se formaron con los conjuntos A-B, C-D y E, respectivamente. La exactitud, sensibilidad y especificidad del mejor modelo clasificador CNN obtenido fueron de 99,61; 99,10 y 98,99, respectivamente. Adicionalmente, se desarrolló otro clasificador basado en las máquinas de vectores de soporte (SVM de sus siglas en inglés) utilizando como extractor de rasgos el modelo CNN entrenado, al cual se le eliminó la capa de salida. Los rasgos de entrada a la SVM se tomaron de la salida de la capa densamente conectada de la CNN. La SVM se entrenó con los mismos datos (representación tiempo-frecuencia de las señales) con los que se entrenó la CNN, y su

desempeño en exactitud, sensibilidad y especificidad fue del 100 %, tanto para los datos de entrenamiento como para los datos de prueba.

**Palabras clave.** Señales EEG epilépticas, redes neuronales convolucionales, SST-QSTFT.

## Epileptic Signal Detection Using Quilted Synchrosqueezing Transform Based Convolutional Neural Networks

**Abstract.** This work proposes a convolutional neural networks-based algorithm to classify electroencephalographic signals (EEG) in normal, preictal and ictal classes to supporting to the physicians to diagnose the epilepsy condition. EEG signals are preprocessed through the application of the synchrosqueezing transform based on the quilted short time Fourier transform (SS-QSTFT) to generate a time-frequency representation, which is the input to the convolutional neural network (CNN). CNN based classifiers are trained using the EEG database of the University of Bonn, which have five sets identified as A, B, C, D and E. Normal, preictal and ictal classes were composed with the combination of the sets A-B, C-D and E, respectively. Accuracy, sensitivity and specificity of the best CNN-based classifier were 99.61, 99.10 and 98.99, respectively. Furthermore, another support vector

machines (SVM)-based classifier was developed using the previous CNN model as feature extractor, which last output layer was removed. Input features to the SVM were taken from the fully-connected layer of the CNN. SVM were trained using the same data (time-frequency representation) utilized to train the previous CNN, and their performance in accuracy, sensitivity and specificity were 100% for training and testing sets.

**Keywords.** Epileptic EEG signals, convolutional neural networks, SST-QSTFT.

## 1. Introducción

Uno de los desórdenes neurológicos más importante que afecta a la actividad cerebral es la epilepsia. La epilepsia es una condición que padecen 50 millones de personas alrededor del mundo [13, 25], que produce crisis convulsivas que afecta la calidad de vida del paciente [18]. Para la detección de la condición de epilepsia el especialista tiene que analizar e interpretar un conjunto de señales electroencefalográficas (EEG) muy extensas, lo que hace que la detección sea un proceso muy tedioso, y además dependiente del criterio del especialista que realiza dicho análisis. Es imperativo automatizar la detección de señales epilépticas mediante el análisis de las señales EEG para coadyuvar al especialista en su tarea de analizar estas señales. Actualmente, existen técnicas para la clasificación de las señales EEG en normales, preictales e iciales, basados en algoritmos de inteligencia artificial, destacando en los últimos años las redes neuronales profundas (entre ellas las redes convolucionales y las redes recurrentes).

Una característica a destacar de las redes neuronales profundas es la capacidad de extraer información subyacente (rasgos) de los datos entrada por medio de sus numerosas capas intermedias, lo que elimina la necesidad de extraer rasgos a mano, y para ello utilizan una vasta cantidad de técnicas lineales y no lineales, basadas en el tiempo o en la frecuencia o combinación de ellas. Sin embargo, existen trabajos donde señales temporales son convertidas en imágenes por medio de transformaciones tiempo-frecuencia, entre ellas la transformada corta de Fourier, y las transformadas Wavelets [23], no para

extraer rasgos para entrenar a la red neuronal profunda, sino para convertir la señal temporal en una imagen con la cual entrenar una red convolucional. Un aspecto relevante de los análisis tiempo-frecuencia es que permiten descubrir las componentes frecuenciales y los instantes en que estos ocurren en una señal temporal.

Una de las técnicas de análisis tiempo frecuencia es la transformada Synchrosqueezing (SST) que descompone a la señal temporal y permite su reconstrucción en sus amplitudes y frecuencias instantáneas [7]. El propósito de la transformada SST es proveer una representación concentrada de las señales multicomponentes (también llamadas modos) en el plano tiempo-frecuencia, y dar un método de descomposición que permita separar y demodular los diferentes modos [14]. La concepción original de la transformada SST está basada en la transformada Wavelet continua [6], pero otros investigadores desarrollaron la técnica basada en la transformada corta de Fourier (STFT) [21] conocida como transformada Synchrosqueezing-STFT (SST-STFT). Una de las debilidades de las técnicas de las transformadas SST basadas en CWT o STFT es debido a la restricción en la resolución tiempo-frecuencia [5].

Últimamente se han desarrollado métodos que incluyen generalizaciones de la transformada SST donde utilizan distintas ventanas para adaptar la señal en instantes diferentes [4]. Esta técnica se conoce como la transformada SST basada en la transformada STFT acolchada (Quilted), mejor conocida por sus siglas SST-QSTFT [5, 4], y fundamentalmente se basa en ajustar a la ventana dependiendo de la región de interés en el plano tiempo-frecuencia. El efecto que se logra con esta ventana ajustable es una mayor separabilidad de las componentes frecuenciales en sus instantes de existencia.

La propuesta de la presente investigación es desarrollar un clasificador de señales EEG temporales en normales, preictales e iciales, las cuales son convertidas en imágenes utilizando la transformada SST-QSTFT [5, 4], para luego entrenar a una red neuronal convolucional.

La estructura del presente trabajo es la siguiente: La sección 2 aborda los trabajos previos

relacionados con investigación; la sección 3 presenta la fundamentación teórica de la SST-QSTFT, las CCNs y las máquinas de vectores de soporte (SVM de sus siglas en inglés); la sección 4 presenta la metodología del desarrollo de los clasificadores; la sección 5 presenta la discusión de los resultados, y se hace una comparación de los resultados de esta investigación con resultados de otros investigadores; finalmente se presentan conclusiones de este trabajo.

## 2. Trabajos relacionados

Uno de los primeros trabajos de clasificación de señales EEG utilizando redes neuronales convolucionales (CNN de sus siglas en inglés) fue realizado por Acharya y colaboradores [1], donde los autores presentan el desarrollo de una red convolucional donde la entrada es una señal temporal de EEG para clasificarla en una de las siguientes clases: normal, preictal e ictal. Las señales utilizadas para el entrenamiento fueron tomadas de la base de datos de la Universidad de Bonn (<http://epilepsy.uni-freiburg.de/database>) [3]. La base de datos dispone de cinco conjuntos de datos denominados A, B, C, D y E, de los cuales los autores tomaron los conjuntos B (normal), D (preictal) y E (ictal). La arquitectura de la red neuronal convolucional (de una dimensión) fue una secuencia de cinco combinaciones de capas de convolución seguida de una de submuestreo (maxpooling), más tres capas completamente conectadas, para un total de 13 capas [1].

La partición del conjunto de datos fue de la siguiente manera: 90% (270 señales) para entrenamiento y 10% (30 señales) para prueba. El conjunto de entrenamiento fue dividido a su vez en 70% (189 señales) para entrenar el modelo y el 30% (81 señales) para validar el modelo al final de cada época en la fase de entrenamiento. El entrenamiento consistió en un proceso de validación cruzada de 10 conjuntos. La exactitud, sensibilidad y especificidad globales de predicción para las clases normal, preictal e ictal, en el entrenamiento, fueron 88,67%, 95,00% y 95,00%, respectivamente [1].

Ullah y colaboradores [24] proponen una red convolucional piramidal con una señal temporal

como entrada, al igual que la red propuesta en [1], la señal de entrada a la red es unidimensional. El esquema es propuesto para mejorar el desempeño de clasificación para el caso de tres clases, normal, interictal e ictal, además de reducir en un 60% el número de parámetros con respecto a una arquitectura de red convolucional tradicional [24]. La base de datos de señales EEG utilizada fue la de la Universidad de Bonn [3].

El sistema de clasificación propuesto está compuesto de tres módulos: Un módulo para dividir la señal de prueba (de 4097 muestras) en cuatro subseñales de 1024 muestras cada una (sin solapamiento), luego cada submuestra se dividió en tres (3) segmentos de 512 muestras solapados un 50%. Un segundo módulo consistente de tres clasificadores con estructura piramidal que clasifica cada uno por separado un segmento obtenido en el primer módulo. El tercer y último módulo toma la salida de cada uno de los clasificadores y decide la clase por voto mayoritario.

El modelo propuesto por los autores está compuesto de siete bloques. Los primeros tres es una secuencia de una capa de convolución, una capa de normalización y una capa de función de activación ReLu. El cuarto bloque comprende una capa densamente conectada y una capa ReLu. El quinto, sexto y séptimo bloque es una capa de "dropout", una capa densamente conectada y la capa softmax de salida, respectivamente. Los autores emplearon una estrategia para aumentar el conjunto de entrenamiento al dividir la señal de 4097 muestras utilizando una ventana de 512 muestras (2,95 segundos) con un paso de 64 muestras (448 muestras de solapamiento o 87,5%), resultando en 57 subseñales o segmentos a partir de una señal.

Considerando que tomaron como conjunto de entrenamiento el 90% de cada clase de la base de datos (90 señales de cada conjunto A, B, C, D y E), el número de segmentos para entrenamiento fue de 25650, en vez de 450 si fueran tomado la señal completa para entrenar. Los autores proponen ocho modelos distintos, cuatro de ellos con una arquitectura de red convolucional estándar, y cuatro modelos con su propuesta de arquitectura piramidal [24]. Luego entrenan cada uno de los

ocho modelos con validación cruzada de 10 conjuntos, y el mejor modelo resultó el modelo piramidal con menor número de parámetros.

Los resultados reportados de exactitud muestran un desempeño de 99,1% de exactitud, la sensibilidad promedio para las clases normal (AB), interictal (CD) e ictal (E) fueron de 97,50; 90,42; 95,83, respectivamente. La especificidad promedio para las clases normal (AB), interictal (CD) e ictal (E) fue de 93,52; 97,21; 99,78, respectivamente. Los valores de sensibilidad y especificidad promedios del entrenamiento se calcularon de los valores reportados por los autores.

Türk y Özerdem [23] utilizan la base de datos de la Universidad de Bonn [3] y plantean cuatro clasificadores: a) Binario (A-B); b) Tres clases (A-B-E); c) Cuatro clases (A-C-D-E); y d) Cinco clases (A-B-C-D-E). La señal de entrenamiento corresponde a la señal completa (4097 muestras) del conjunto seleccionado para cada uno de los clasificadores, sin preprocesamiento. El esquema propuesto por los autores consiste en un bloque que realiza la transformada Wavelet continua de la señal, que tiene como salida una imagen (escalograma) de la señal de  $662 \times 536$  píxeles. Luego, esta imagen pasa por un bloque que la redimensiona a  $32 \times 32$  píxeles, mediante un método de interpolación cúbica [23].

Finalmente la imagen redimensionada pasa a un clasificador basado en la red convolucional. La función Wavelet madre que utilizaron fue la Morlet continua. La arquitectura de la red convolucional fue dos secuencias de capas de convolución y maxpooling (4 capas), una capa densa, y una capa softmax de salida con un número de nodos de salida dependiendo del número de clases [23].

Los datos se dividieron en un 90% (180, 270, 360, 450 señales para cada clasificador) para entrenamiento, y un 10% (20, 30, 40, 50 señales para cada clasificador) para pruebas. El conjunto de entrenamiento fue dividido en un 80% (144, 216, 288, 360 señales) para entrenar la red y un 20% (36, 54, 72, 90 señales) para medir el error de validación en cada época durante el entrenamiento.

El entrenamiento de cada clasificador lo hicieron con validación cruzada de 10 conjuntos. La

exactitud, sensibilidad y especificidad promedio globales fue de 98,01; 96,92; 98,59, para los clasificadores A versus C versus E, A versus D versus E, B versus C versus E, y B versus D versus E, respectivamente.

Zhao y colaboradores [26] proponen una red convolucional de una dimensión compuesta de tres bloques convolucionales y tres bloques con capas densamente conectadas, para la detección de señales epilépticas utilizando la base de datos de la Universidad de Bonn [3]. Los tres bloques convolucionales (para extracción de rasgos) consistieron de las siguientes capas: Convolución, normalización, ReLu, dropout y maxpooling [26]. Los primeros dos bloques densamente conectados fueron constituidos por una capa densamente conectada, un capa ReLu, y una capa dropout [26]. El último bloque consistió de una capa densamente conectada y una capa de salida softmax, con tantos nodos de salida como clases corresponda (2, 3 o 5). Los autores plantean tres modelos: a) Un clasificador binario de las clases normal y epiléptica; b) Un clasificador de tres clases, normal, interictal e ictal; Un clasificador de cinco clases A, B, C, D y E.

Los autores proponen 8 arquitecturas distintas de redes convolucionales, de las cuales seleccionan al modelo M7 [26] (por validación cruzada) que le arrojó el mejor desempeño en exactitud, sensibilidad y especificidad, para el caso de cinco clases. Cada señal del conjunto de datos fue dividida en 23 segmentos de 1 segundo (178 muestras) no solapados para incrementar el número de señales de 500 (100 por cada tipo) a 11500 señales (2300 por cada tipo). Los autores diseñaron 14 modelos de clasificación binaria donde combinaron las señales A, B, C, D y la designaron como la clase normal contra la clase ictal representada por las señales E.

Para el caso de tres clases diseñaron cinco clasificadores, donde cuatro de los clasificadores tomaron dos de las clases entre los conjuntos A, B, C y D y la tercera clase correspondió al conjunto E. El quinto clasificador de tres clases correspondió a la combinación de los conjuntos AB como la primera clase, los conjuntos CD como la segunda clase, y el conjunto E como la tercera clase. Para el clasificador de cinco clases cada

conjunto A, B, C, D y E representa una clase. Todas las redes convolucionales se entrenaron con validación cruzada de 10 conjuntos de datos. Para el caso del entrenamiento de tres clases con la combinación de los conjuntos AB, CD y E, la exactitud promedio del entrenamiento con validación cruzada fue de 96,97% [26].

Ramakrishnan y colaboradores [15] proponen una red convolucional de una dimensión, pero donde la señal de entrada no corresponde a la señal temporal sino a un conjunto de trece rasgos temporales y frecuenciales extraídos de la misma. Las bases de datos utilizadas en este trabajo son la base de datos de la Universidad de Bonn [3] y la base de datos CHB-MIT EEG del Hospital de Niños de Boston [17]. Los rasgos extraídos de la señal temporal son: Media, mediana, desviación estándar, skewness, kurtosis, varianza, mínimo, máximo, energía y la raíz del valor medio cuadrático. Los últimos tres rasgos los obtuvieron de la transformada Wavelet de la señal: Frecuencia pico, energía de la banda espectral, y la densidad espectral de potencia de Welsh.

La arquitectura de la red convolucional está compuesta de 3 capas convolucionales, 2 capas de maxpooling, una capa densamente conectada y la capa softmax de salida. Los clasificadores diseñados fueron: a) Binario A-E, AB-E y AC-E, b) Tres clases A-D-E y AB-CD-E, para la base de datos de la Universidad de Bonn [3]. El clasificador binario que se basó en la base de datos CHB-MIT EEG [17] discriminaba entre la clase normal y epiléptica. Los datos se dividieron en 50% para entrenamiento y el 50% para pruebas. La exactitud reportada para el problema de tres clases AB-CD-E fue de 98,37% y 98,34% para el problema binario con la base de datos CHB-MIT-EEG.

Agrawal y colaboradores [2] proponen una red convolucional para clasificar las señales en epilépticas y no epilépticas. La novedad de este trabajo es que la red se basa en la transferencia de aprendizaje, y para ello ensayan con tres redes convolucionales pre-entrenadas muy conocidas: googlenet [20], resnet101 [11] y vgg19 [19].

La base de datos utilizada es la de la Universidad de Bonn [3], cada señal es segmentada en 23 segmentos de 1 segundo (178 muestras). Las

señales de 1 segundo son convertidas a imágenes RGB de  $224 \times 224$  píxeles utilizando el siguiente algoritmo: Conjunto E (clase epiléptica), de las 2300 señales de 1 segundo de duración toma 2292 señales y forma una matriz de  $2292 \times 178$ , luego esta matriz es convertida en un vector de 407976 elementos. Este vector es redimensionado en 8 matrices de  $224 \times 224$  elementos, descartando el resto de los elementos del vector. Luego convierte esta matriz a imagen RGB utilizando una función especial de MATLAB [22].

El mismo proceso es aplicado a las señales no epilépticas (no especificadas en el artículo), pero generando solo 8 imágenes RGB de la clase no epiléptica. El número de imágenes para entrenar la red es dieciséis, de las cuales tomaron 11 imágenes (70%) para entrenar y 5 (30%) para prueba. Los autores reemplazaron la última capa para sustituirla por una de dos clases, y sustituyeron las capas densamente conectadas. Sus resultados muestran que la mejor exactitud fue lograda con la vgg19 (99,8%), pero el menor tiempo de entrenamiento lo logró la red preentrenada con googlenet (exactitud de 98,55%) de 41 segundos, mucho menor que los 10,45 minutos que tardó en entrenarse la red basada en la vgg19.

Hussein y colaboradores [10] proponen una red de cuatro capas, la primera capa es una LSTM (Long Short-Term Memory), la segunda es una capa densamente conectada, estas dos capas extraen los rasgos más relevantes [10] que representen las clases de las señales EEG, que son las entradas de la tercera capa average pooling de una dimensión, y la última capa softmax de salida que estima la clase de la señal de entrada. La base de datos que utilizaron en este trabajo es de la Universidad de Bonn [3], la cual expandieron (aumentaron) agregando ruido blanco y artefactos como movimiento muscular y de los ojos artificialmente, obtenidos a través de modelos. Ellos plantearon cuatro modelos clasificadores: a) Dos binarios con las clases normal y epiléptica (A vs E y ABCD vs E); b) Uno de tres clases con las clases normal (A), interictal (C) y epiléptica (E) y; c) Cinco clases, A vs B vs C vs D vs E. Los autores propusieron dos estrategias de entrenamiento: 1) Dividieron a los

datos en dos conjuntos, uno para entrenamiento y otro para prueba, con varios porcentajes de división, y 2) Validación cruzada con tres, cinco y diez conjuntos. Cada una de las señales de los conjuntos A al E (4097) se redimensionaron a  $2048 \times 2$ .

Para las dos estrategias de evaluación, para todos problemas (dos, tres y cinco clases), y señales sin ruido ni artefacto añadidos, tanto la exactitud como la sensibilidad y la especificidad fueron del 100%. Para los clasificadores con artefactos musculares y movimiento ocular, y ruido blanco, la exactitud es afectada por la relación señal-ruido (RSR), pero para ciertos niveles de la RSR la exactitud de todos los clasificadores fue del 100%.

### 3. Fundamentación teórica

#### 3.1. Transformada *synchrosqueezing* basada en la transformada corta de Fourier acolchada

La transformada *synchrosqueezing* (SST) fue utilizada por primera vez para la identificación de voz humana, donde Daubechies y colaboradores propusieron un método basado en la transformada Wavelets y un modelo del sistema del nervio auditivo humano [7], cuyo objetivo fue obtener una representación tiempo-frecuencia bien definida de los distintas componentes espectrales de la señal [6]. La transformada consiste en una reasignación de componentes en el plano tiempo-escala al plano tiempo-frecuencia [7], obtenidos mediante el análisis de la señal con la transformada Wavelet continua, para obtener una representación tiempo-frecuencia concentrada, de donde se extraen las amplitudes y frecuencias instantáneas bien separadas [6, 5].

Thakur y Wu [21] propusieron una transformada SST basada en la transformada corta de Fourier (STFT de sus siglas en inglés), pero que no permite la reconstrucción de los modos [14]. Oberlin y colaboradores [14] propusieron la transformada SST basada en la transformada corta de Fourier donde desarrollan un teorema de aproximación similar al propuesto por Daubechies

y colaboradores [6], que permite la reconstrucción de los modos.

La principal debilidad que presenta la transformada SST basada en Wavelets o la STFT es que la resolución tiempo-frecuencia está limitada por la función de la ventana, que pudiera no exhibir la riqueza tiempo-frecuencia de la señal con suficiente exactitud [5]. En [4] proponen una variante de la SST basada en la STFT donde ajustan una ventana dependiendo de la región de interés en el plano tiempo frecuencia lo que permite una mayor separabilidad de las frecuencias instantáneas [4]. Esta transformada recibe el nombre de SST basada en la STFT acolchada (del inglés *quilted*) mejor conocida por sus siglas en inglés SST-QSTFT.

Dada una señal  $f$ :

$$f(t) = \sum_{k=1}^K f_k(t), \quad (1)$$

donde  $f_k(t) = A_k(t)e^{2\pi i\phi_k(t)}$  es una descomposición amplitud-fase de  $f$ .  $A_k(t)$  y  $\phi'_k(t) = d\phi_k/dt$ , para  $k = 1, 2, \dots, K$ , son las amplitudes instantáneas (AI) y frecuencias instantáneas (FI) de  $f$ . Las AI y FI son una clase de señales débilmente moduladas si existen  $\epsilon, d > 0$  y se cumple para  $k = 1, 2, \dots, K$  [4]:

- $A_k \in C^1 \cap L^\infty, \phi_k \in C^2, \phi'_k \in L^\infty, \inf_{t \in \mathbb{R}} A_k(t) > 0, \inf_{t \in \mathbb{R}} \phi'_k(t),$
- $\forall t \in \mathbb{R} \mid A'_k(t) \leq \epsilon |\phi'_k(t)| \text{ y } |\phi''_k(t)| \leq |\phi'_k(t)|,$
- Si  $k \geq 2$ , entonces  $\phi'_k(t) - \phi'_{k-1}(t) > d \forall t \in \mathbb{R}.$

La STFT de  $f(t)$  utilizando la ventana  $g \in L^2(\mathbb{R})$  centrada en 0 es:

$$V_g f(t, \xi) = \int_{-\infty}^{+\infty} f(\tau) g(\tau - t) e^{-2\pi i \xi(\tau - t)} d\tau. \quad (2)$$

Esta transformada tiene la limitación de la resolución tiempo-frecuencia; para  $g$  angosta tiene buena resolución temporal pero muy pobre resolución en frecuencia, y para  $g$  ancha tiene buena resolución en frecuencia pero muy pobre resolución en el tiempo.

La transformada SS basada en la STFT [21] de  $f$ , con una tolerancia  $\gamma > 0$  y un parámetro limitante  $\beta > 0$  está dada por:

$$S_{g,\gamma}^\beta f(t, \xi) = \int_{A_{g,\gamma} f(t)} V_g f(t, \eta) \frac{1}{\beta} b\left(\frac{\xi - \xi_g f(t, \eta)}{\beta}\right) d\eta,$$

donde  $b \in C_c^\infty(\mathbb{R})$  es una "función protuberancia" (*bump function*) que satisface  $\hat{b}(0) = 1$ ,  $A_{g,\gamma} f(t) = \{\eta \in \mathbb{R}^+ : |V_g f(t, \eta)| > \gamma\}$ , y  $\xi_g f(t, \eta) = \frac{\partial_t V_g f(t, \eta)}{2\pi i V_g f(t, \eta)}$  es la reasignación de frecuencia basada en la STFT.

La transformada corta de Fourier acolchada (QSTFT: Quilted Short-Time Fourier Transform), definida para una familia de ventanas continuas acolchadas de dos parámetros  $h_{t,\xi} \in L^2(\mathbb{R})$  centrada en 0, y  $h(x, t, \xi) = h_{t,\xi}(x)$ , es [5]:

$$V_h^Q f(t, \xi) = \int_{\mathbb{R}} f(\tau) h_{t,\xi}(\tau - t) e^{-2\pi i \xi(\tau - t)} d\tau. \quad (3)$$

Para garantizar la exactitud de la SST, la familia de ventanas acolchadas  $h_{t,\xi}$  debe satisfacer ciertas condiciones que el interesado puede revisar en [4]. La transformada SS basada en la QSTFT (SST-QSTFT) de  $f$ , con una tolerancia  $\gamma > 0$  y un parámetro limitante  $\beta > 0$  está dada por:

$$S_{h,\gamma}^{Q,\beta} f(t, \xi) = \int_{A_{g,\gamma}^Q f(t)} V_h f(t, \eta) \frac{1}{\beta} b\left(\frac{\xi - \xi_h^Q f(t, \eta)}{\beta}\right) d\eta,$$

donde  $b \in C_c^\infty(\mathbb{R})$  es una función protuberancia que satisface  $\hat{b}(0) = 1$ ,  $A_{g,\gamma}^Q f(t) = \{\eta \in \mathbb{R}^+ : |V_h^Q f(t, \eta)| > \gamma\}$ , y  $\xi_h^Q f(t, \xi) = \frac{\partial_t V_h^Q f(t, \xi)}{2\pi i V_h^Q f(t, \xi)}$  es la reasignación de frecuencia basada en la QSTFT.

### 3.2. Redes neuronales convolucionales

Las redes neuronales convolucionales (de sus siglas en inglés CNN: *Convolutional Neural Networks*) son modelos inspirados en la estructura fisiológica de la corteza visual del gato [9], en la cual existen regiones específicas que excitan

neuronas particulares (campos receptivos). El *Neocognitron* [8] fue la primera arquitectura basada en la corteza visual de los mamíferos para procesar imágenes, y fue el precursor de la primera red neuronal convolucional, conocida como *LeNet-5* [12]. Las CNN consisten de la interconexión de dos tipos de capas, la primera se conoce como capa de convolución [16] (de la cual estas redes reciben su nombre), y la segunda como capa de *pooling* o submuestreo espacial.

La capa de convolución recibe la imagen de entrada de cierto ancho (píxeles), alto (píxeles) y profundidad (canales de color) y se convoluciona con un filtro (*kernel*), de mucho menor dimensión que la imagen (menores alto y ancho pero con un mayor número de canales), sobre regiones localizadas o campos receptivos de la imagen, finalmente esta capa está seguida de una función de activación no lineal, produciendo una salida conocida como mapa de rasgos [16]. La capa de submuestreo espacial reemplaza cierta región (campo receptivo) del mapa de rasgos de entrada (salida de la capa de convolución-función de activación) con un valor que puede ser el valor máximo (*max pooling*) o el valor promedio (*average pooling*) de la región del mapa de rasgos en consideración; esta capa recibe el nombre de submuestreo porque su salida tiene ancho y alto menores que las dimensiones del mapa de rasgos de entrada.

Las CNN terminan en una capa de clasificación que consiste de una capa densamente conectada (*fully connected*) y la capa de salida, que corresponden a un perceptrón multicapa [16]. Las capas de convolución menos profundas (las cercanas a la entrada de la red) capturan formas como líneas verticales u horizontales u otras formas primitivas o simples, mientras que las capas más profundas o cercanas a la salida capturan formas más complejas [16].

La función de activación no lineal proporciona características no lineales a la red; la función de activación no lineal más utilizada es la unidad lineal rectificadora, ReLU (de sus siglas en inglés *Rectified Linear Unit*). La capa de submuestreo tiene la propiedad de ser invariante a pequeños desplazamientos en el espacio, además de reducir el número de parámetros [26, 16].

### 3.3. Máquinas de vectores de soporte

Otro algoritmo de inteligencia artificial para clasificación binaria, con soporte matemático bien sólido, son las máquinas de vectores de soporte (SVM de sus siglas en inglés), cuyo objetivo de optimización es maximizar el margen entre dos clases de ejemplos o muestras, en vez de minimizar el error de clasificación para el caso de las redes neuronales artificiales [16]. El margen es la distancia entre el hiperplano separador o frontera de decisión lineal y los ejemplos de entrenamiento más cercanos a la frontera de decisión, estos ejemplos más cercanos al hiperplano separador se conocen como vectores de soporte [16]. Las SVM se basan sobre una transformación de los datos a un espacio de mayor dimensión, que el espacio donde yacen los datos, en el cual los ejemplos pertenecientes a las dos clases distintas pueden ser separados linealmente.

Debido a que el espacio a donde se proyectan los datos es de mayor dimensión que el espacio de entrada el plano separador de las clases, recibe el nombre de hiperplano separador. Cuando la distancia o margen del hiperplano separador es máxima entre los ejemplos de ambas clases denominados vectores de soporte, recibe el nombre de hiperplano separador óptimo. La función de transformación del espacio de los datos originales al hiperespacio se conoce como función *kernel*, entre las cuales se puede mencionar la función de base radial o Gaussiana (RBF: *Radial Basis Function*) que posee un solo parámetro de ajuste (ancho de la Gaussiana), y es ampliamente usada en problemas de clasificación. Las SVM son por definición un clasificador binario, pero para los casos con más de dos clases ( $K > 2$  clases) existe un enfoque conocido uno contra todos (*one-versus-rest*), donde se entrenan  $K$  clasificadores binarios como  $K$  clases existan en el problema.

Este enfoque toma una clase de datos como la clase positiva y el resto ( $K - 1$  clases) como la clase negativa y se entrena un clasificador, y así para el resto de las clases. Para determinar la clase a la que pertenece una muestra se evalúan todos los clasificadores con dicha muestra, y se

le asigna la clase cuya evaluación de la función de decisión de cada clasificador sea la más alta. Uno de los algoritmos de SVM se conoce como las SVM de margen suave o  $C$ -SVM, donde a través del parámetro de regularización  $C$  se ajusta el margen del hiperplano separador óptimo en presencia de datos con ruido.

## 4. Metodología

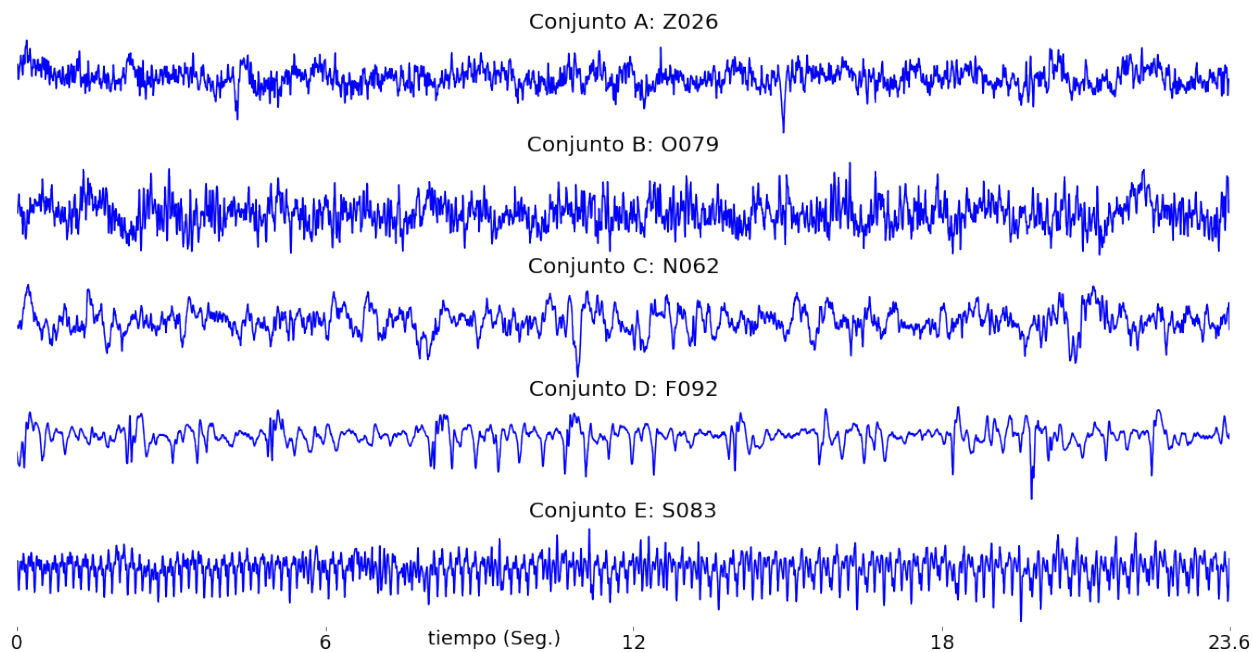
### 4.1. Base de datos

La base de datos utilizada en este trabajo corresponde a las señales EEG de la Universidad de Bonn descrita en [3], está conformada por cinco conjuntos A, B, C, D y E, cada uno con 100 señales electroencefalográficas (EEG) monocanal de 23,6 segundos de duración cada una. Estas señales EEG están libres de artefactos debido a la actividad muscular o movimientos de los ojos. Los conjuntos A y B consisten en segmentos tomados de los registros EEG superficiales obtenidos de cinco voluntarios sanos usando un esquema de colocación de los electrodos estandarizada, conocida como sistema 10-20. Los conjuntos A y B corresponden a voluntarios despiertos, relajados y con los ojos abiertos (A) y los ojos cerrados (B). Los conjuntos C, D y E, de pacientes diagnosticados con epilepsia, corresponden a EEG profundos o intracraneales.

Las señales en el conjunto C fueron obtenidas de la formación hipocampal del cerebro. El conjunto D se obtuvo dentro de la zona epileptogénica. Los conjuntos C y D solo contienen registros de la actividad cerebral medida durante los intervalos sin crisis epilépticas (interictal). El conjunto E contiene registros durante la actividad convulsiva, o periodo ictal.

Todos estos segmentos EEG se registraron con un amplificador de 128 canales, un convertidor analógico-digital de 12 bits a una frecuencia de muestreo de 173,61 Hz, y se les aplicó un filtro pasabanda con ajustes de 0,53 Hz y 40 Hz [3]. El número de registros en total es 500, con 4097 muestras cada uno. La Figura 1 presenta una muestra de cada una de las cinco señales por cada conjunto, la unidad de los ejes verticales está en microvoltios ( $\mu V$ ).





**Fig. 1.** Muestras de señales de los conjuntos A, B, C, D y E

#### 4.2. Clasificación de la señal EEG

Las tres clases de interés son Normal, Preictal e Ictal. Los conjuntos de la clase Normal lo conforman los conjuntos A y B, la clase Preictal la conforma los conjunto C y D, y la clase Ictal es el conjunto E. Se mantiene el nombre de la clase preictal en vez de interictal para ser consistente con los trabajos citados en la sección 2.

El diagrama de bloques general para la clasificación de la señales EEG se muestra en la Figura 2. La señal EEG sin pre-procesar entra al bloque que calcula su SST-QSTFT y genera una imagen, a escala de grises con dicha representación, de 496 pixeles de ancho y 368 pixeles de alto, luego esta imagen se reescala a 31 pixeles de ancho y 23 pixeles de alto, que finalmente entra a la red neuronal convolucional (CNN), que clasifica la señal en una de las tres clases definidas anteriormente.

La arquitectura de la red se muestra en la Tabla 1. El tipo de capa referido como Conv2D (filtro de  $3 \times 3$ ) y MaxPooling2D (filtro de  $2 \times 2$ ) realiza la operación de convolución y submuestreo, respectivamente. La capa ReLU es la función de

activación que sigue a la capa de convolución, la capa Flatten lo que hace es convertir en un vector unidimensional lo que llega a su entrada, la capa Dense y Softmax corresponden al perceptrón multicapa que efectúa la clasificación.

Los dos conjuntos de capas Conv2D-ReLU-MaxPooling2D son básicamente extractores de rasgos para el perceptrón multicapa. La capa Softmax de salida posee tres (3) neuronas, cada una correspondiendo a cada clase, cuyos rangos varían entre cero (0) y uno (1), y su sumatoria es igual uno (1).

Esta arquitectura no contempla un esquema de regularización basado en la técnica dropout, que consiste en la eliminación al azar de algunas neuronas de la red.

La división de los datos fue de 90% (450 señales) para la fase de entrenamiento y 10% (50 señales) para pruebas. Del conjunto de entrenamiento se tomó el 30% (135 señales) para validación en cada época de entrenamiento y el 70% (315 señales) para entrenar.

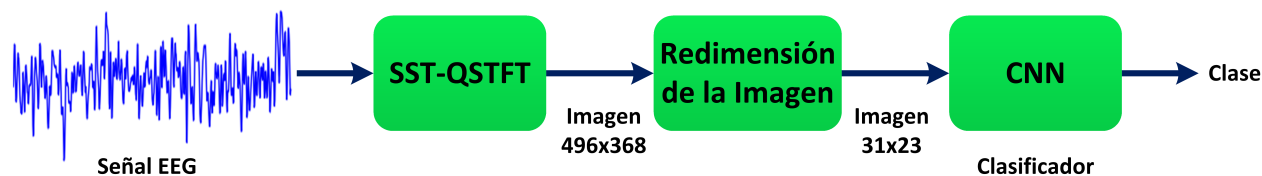


Fig. 2. Proceso de clasificación de la señal EEG

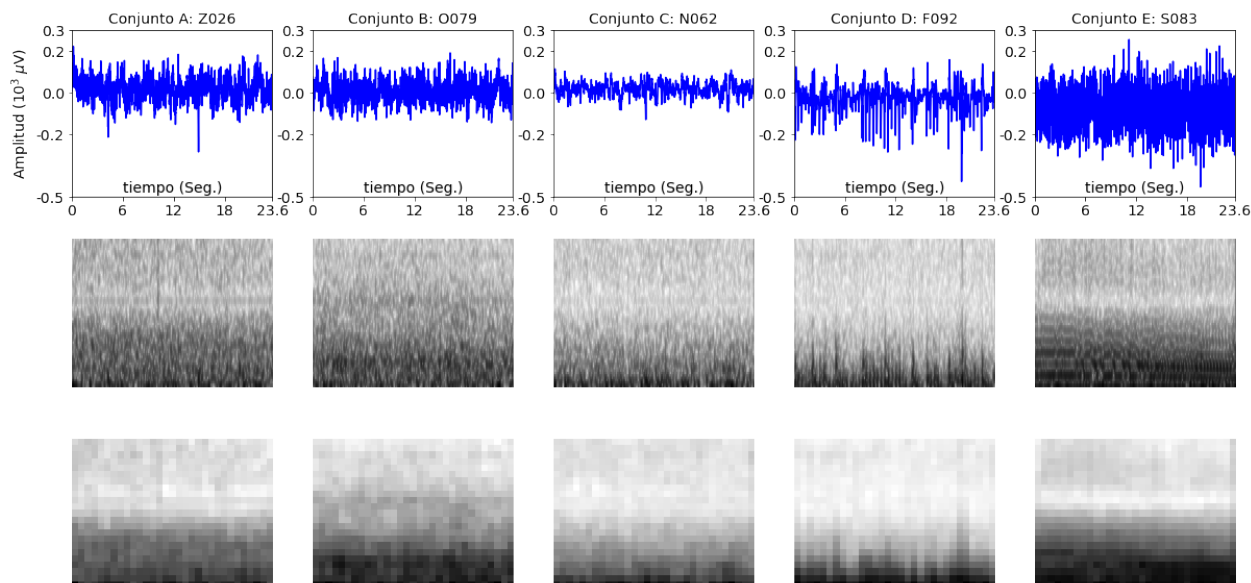


Fig. 3. Superior: Muestras de señales de los conjuntos A, B, C, D y E, Medio: Transformada SST-QSTFT con resolución de 496×368 píxeles, Inferior: Transformada SST-QSTFT con resolución de 31×23 píxeles

#### 4.2.1. Clasificación con la SST-QSTFT de la señal completa con CNN

La Figura 3 muestra una señal de cada conjunto (parte superior), tomadas al azar, y su respectiva transformada SST-QSTFT (parte media) con resolución 496 píxeles de ancho y 368 píxeles de alto, la parte inferior de la Figura 3 muestra la transformada SST-QSTFT reescalada a 31 píxeles de ancho y 23 píxeles de alto.

Los indicadores de desempeño para medir la calidad de la clasificación son la sensibilidad ( $Sen$ ), la especificidad ( $Esp$ ) y la exactitud ( $Exac$ ), las cuales se definen a continuación:

$$Sen = \frac{TP}{TP + FN}, \quad (4)$$

$$Esp = \frac{TN}{TN + FP}, \quad (5)$$

$$Exac = \frac{TP + tn}{TP + TN + PF + FN}, \quad (6)$$

donde  $TP$ ,  $TN$ ,  $FP$  y  $FN$  son los verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos, respectivamente. La sensibilidad mide la capacidad de detectar la clase positiva, la especificidad mide la capacidad de detectar la clase negativa, y la exactitud mide la capacidad de detectar las clases positivas y negativas correctamente. La Tabla 2 muestra la matriz de confusión del entrenamiento, donde se plasman los valores, promediados sobre los diez modelos de CNN, de los indicadores de desempeño para el problema de tres clases. A pesar de tener exactitudes por clase mayor al 90% se observa

Tabla 1. Arquitectura del autoencoder

Tipo de capa	Filtro			Salida
	Dimensión	Profundidad	Paso	Dimensión
Entrada	31×23	1	-	-
Conv2D	3×3	4	1	29×21×4
ReLU	-	-	-	29×21×4
MaxPooling2D	2×2	-	2	14×10×4
Conv2D	3×3	8	1	14×10×8
ReLU	-	-	-	14×10×8
MaxPooling2D	2×2	-	2	7×5×8
Flatten	-	-	-	280×1
Dense	-	-	-	64×1
Softmax	-	-	-	3×1

como la sensibilidad de la clase Ictal es menor al 90%, y la menor de las sensibilidades.

La Tabla 3 muestra la matriz de confusión, promediada sobre los diez modelos de CNN, obtenida con los datos de prueba (50 registros EEG). Se observa un buen desempeño general, incluyendo una buena sensibilidad de la clase Ictal.

Se modificó la arquitectura de la red convolucional al agregarle una capa de dropout (20%) después de la segunda capa de MaxPooling2D, y se realizó el entrenamiento con validación cruzada de 10 conjuntos.

Los resultados del entrenamiento promediados sobre los diez modelos se muestran en la Tabla 4, y la Tabla 5 muestra la evaluación del entrenamiento con el conjunto de pruebas promediado sobre los 10 modelos. Se observa el desempeño superior al mejorar la capacidad de generalización del nuevo modelo al incluir la capa de dropout.

#### 4.2.2. Clasificación con la SST-QSTFT de la señal segmentada con CNN

Para aumentar el conjunto de entrenamiento se segmentaron las señales en cinco segmentos de 5 segundos; los cuatro primeros segmentos no están solapados, y el quinto segmento se solapa con el

cuarto segmento en un 28% (entre los segundos 18,6 y 20). Las figuras 4, 5, 6, 7 y 8 muestran las señales Z006, O079, N062, F092 y S083, correspondientes a las señales de los conjuntos A, B, C, D y E (ver la Figura 1), segmentadas y con su correspondiente transformada SST-QSTFT.

Para esta parte, el la división de los datos para entrenamiento, validación y prueba se hizo de la misma forma que en el desarrollo del primer modelo, es decir, 90% (2250 señales) para la fase de entrenamiento y 10% (250 señales) para pruebas. Del conjunto de entrenamiento se tomó el 30% (675 señales) para validación en cada época de entrenamiento y el 70% (1575 señales) para entrenar.

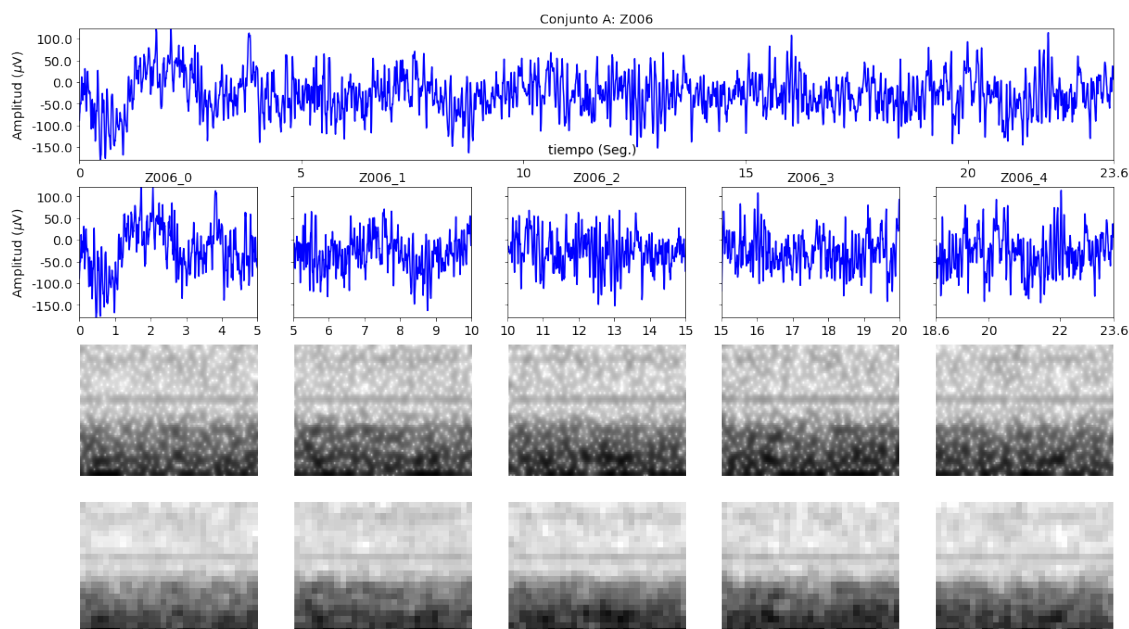
La arquitectura de la red fue idéntica a la red del modelo anterior con dropout, que elimina aleatoriamente el 20% de los pesos de la segunda capa de MaxPooling2D. El entrenamiento se hizo con validación cruzada de diez conjuntos, cuyos resultados de desempeño se ven en las matrices de confusión mostradas en las tablas 6 y 7. La Tabla 6 corresponde a la matriz de confusión de entrenamiento promedio evaluada sobre los diez modelos, y la Tabla 7 corresponde a la matriz de confusión de prueba promedio evaluada sobre los diez modelos. Es evidente, que los índices de exactitud, sensibilidad y especificidad, con los

**Tabla 2.** Matriz de confusión promedio de entrenamiento con validación cruzada sin *dropout*

		Predicción			Indicadores de desempeño (%)		
		Normal	Preictal	Ictal	Exactitud	Sensibilidad	Especificidad
Real	Normal	160,6	19,3	0,1	95,33	89,22	99,41
	Preictal	1,0	177,0	2,0	92,60	98,33	88,74
	Ictal	0,6	11,1	78,3	96,93	87,00	99,42

**Tabla 3.** Matriz de confusión promedio de prueba con validación cruzada

		Predicción			Indicadores de desempeño (%)		
		Normal	Preictal	Ictal	Exactitud	Sensibilidad	Especificidad
Real	Normal	17,8	2,1	0,1	95,60	89,00	100,00
	Preictal	0,0	19,2	0,8	92,20	96,00	89,67
	Ictal	0,0	1,0	9,0	96,20	90,00	97,75

**Fig. 4.** Señal Z006, del conjunto A, segmentada con su respectiva transformada SST-QSTFT con resolución de 496×368 píxeles y con resolución de 31×23 píxeles (inferior).

conjuntos aumentados con la segmentación de las señales, son mayores a los obtenidos con el mejor modelo (con *dropout*) entrenado con las señales completas o sin segmentar.

#### 4.2.3. Clasificación con la SST-QSTFT de la señal segmentada con *C-SVM*

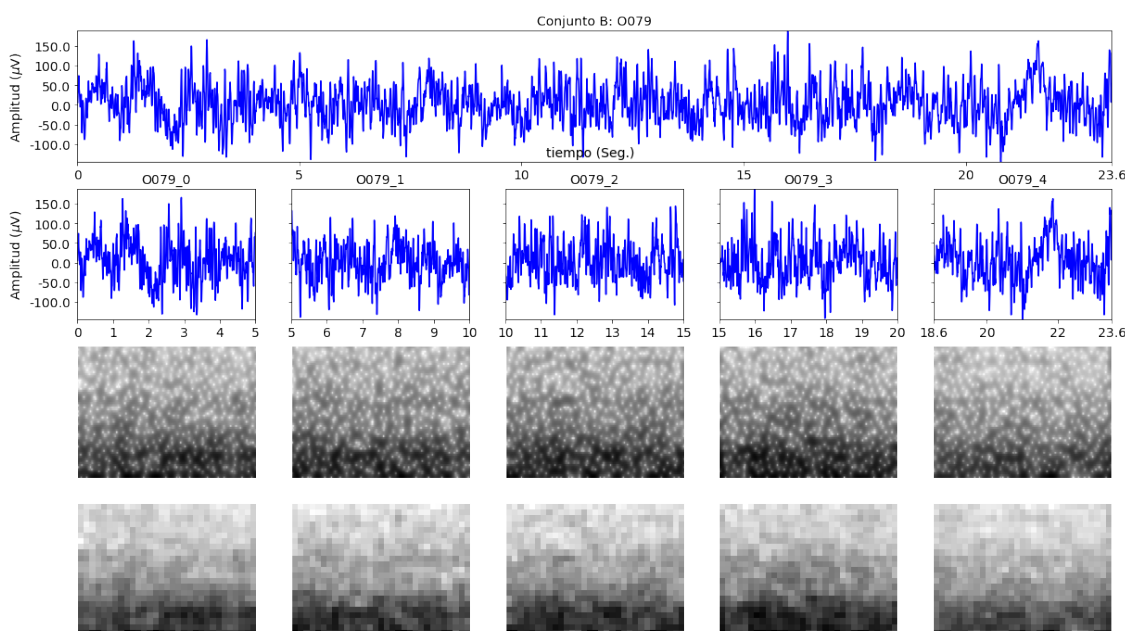
En esta fase de experimentos, se tomaron los mejores modelos CNN con *dropout* (estimadores),

**Tabla 4.** Matriz de confusión promedio de entrenamiento con validación cruzada y dropout

		Predicción			Indicadores de desempeño (%)		
		Normal	Preictal	Ictal	Exactitud	Sensibilidad	Especificidad
Real	Normal	179,5	0,3	0,2	99,33	99,72	99,10
	Preictal	1,7	176,7	1,6	98,36	98,17	98,48
	Ictal	0,8	3,8	85,4	98,58	94,89	99,50

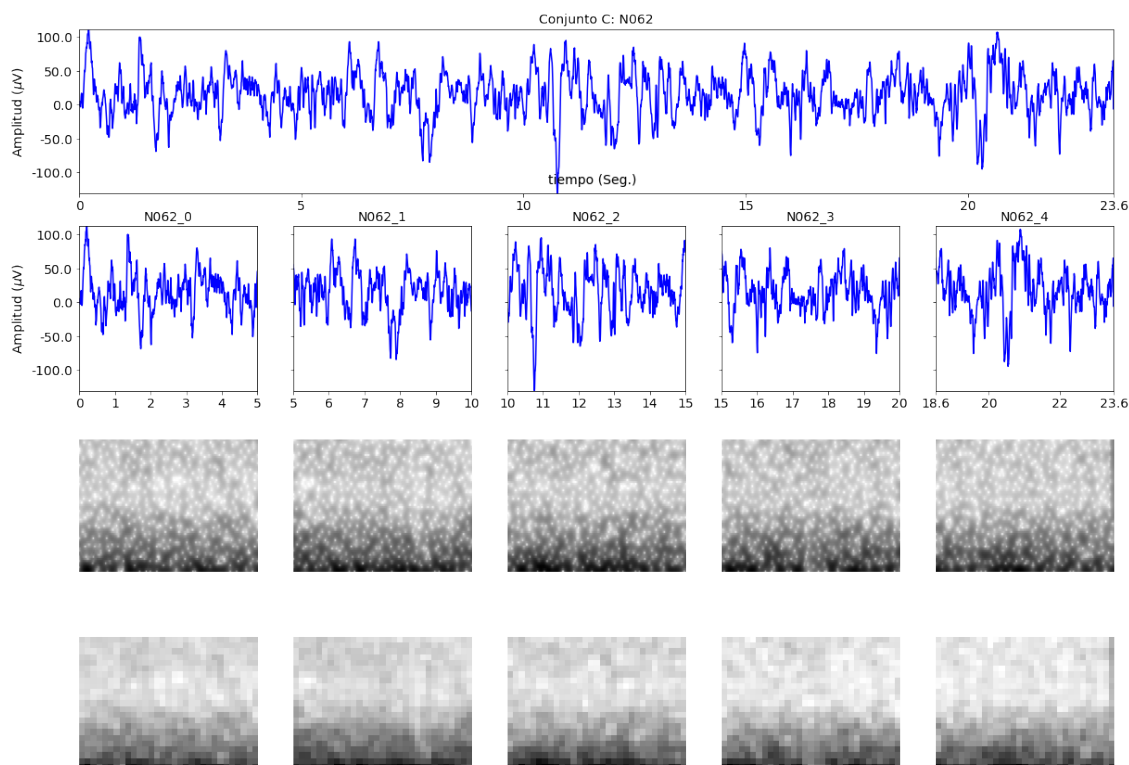
**Tabla 5.** Matriz de confusión promedio de prueba con validación cruzada y dropout

		Predicción			Indicadores de desempeño (%)		
		Normal	Preictal	Ictal	Exactitud	Sensibilidad	Especificidad
Real	Normal	20,0	0,0	0,0	99,60	100,00	99,33
	Preictal	0,2	19,5	0,3	99,00	97,50	100,00
	Ictal	0,0	0,0	10,0	99,40	100,00	99,25

**Fig. 5.** Señal O079, del conjunto B, segmentada con su respectiva transformada SST-QSTFT con resolución de  $496 \times 368$  píxeles y con resolución de  $31 \times 23$  píxeles (inferior).

entrenados con las señales completas (no segmentadas) y con las señales segmentadas, obtenidos por validación cruzada como se describió en las secciones 4.2.1 y 4.2.2, respectivamente. A estos modelos CNN se les suprimió la última capa (*Softmax*), para utilizarlos como extractores de

rasgos. La salida de estas dos redes modificadas es ahora una matriz de rasgos de  $m \times 64$  correspondiente a la capa densamente conectada (capa *Dense*, ver Tabla 1, donde  $m$  es el número de muestras (de entrenamiento o de prueba). Las imágenes de entrenamiento son 450 para



**Fig. 6.** Señal N062, del conjunto C, segmentada con su respectiva transformada SST-QSTFT con resolución de 496×368 píxeles y con resolución de 31×23 píxeles (inferior).

las señales no segmentadas y 2250 para las señales segmentadas.

Las imágenes de prueba son 50 para las señales no segmentadas y 250 para las señales segmentadas. Se procedió a realizar el entrenamiento de la  $C$ -SVM con validación cruzada de 10 conjuntos, utilizando el kernel Gaussiano (RBF) con ajuste de los parámetros de regularización ( $C$ ) y el ancho  $\gamma$  (gamma) de la Gaussiana.

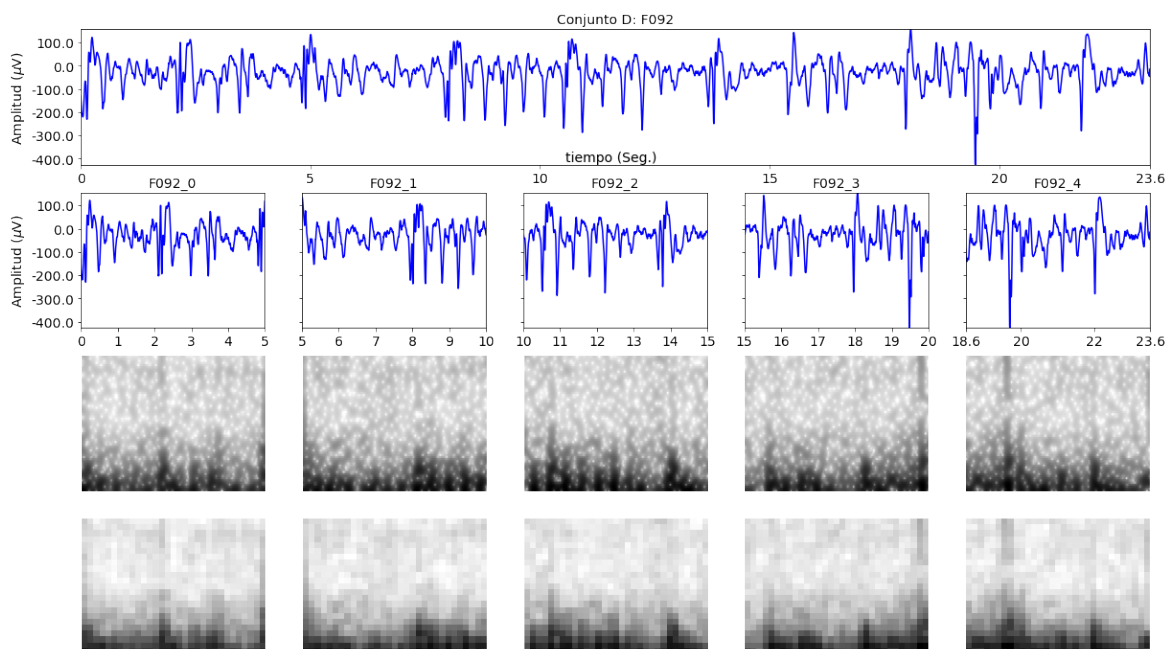
Los valores de  $C$  y  $\gamma$  ensayados por cada fase de validación cruzada fueron  $C = 100, 10, 1, 0, 1, 0, 01$ , y  $\gamma = 10, 1, 0, 1, 0, 01, 0, 001$ . La Tabla 8 muestra los mejores parámetros para cada conjunto de datos (no segmentados y segmentados). La estrategia utilizada para la clasificación multiclase fue *uno-contra-todos* (*one-vs-rest*).

La exactitud, la sensibilidad y la especificidad fueron todas del 100% para ambos clasificadores SVM, tanto para los datos de entrenamiento como para los datos de prueba. Una diferencia

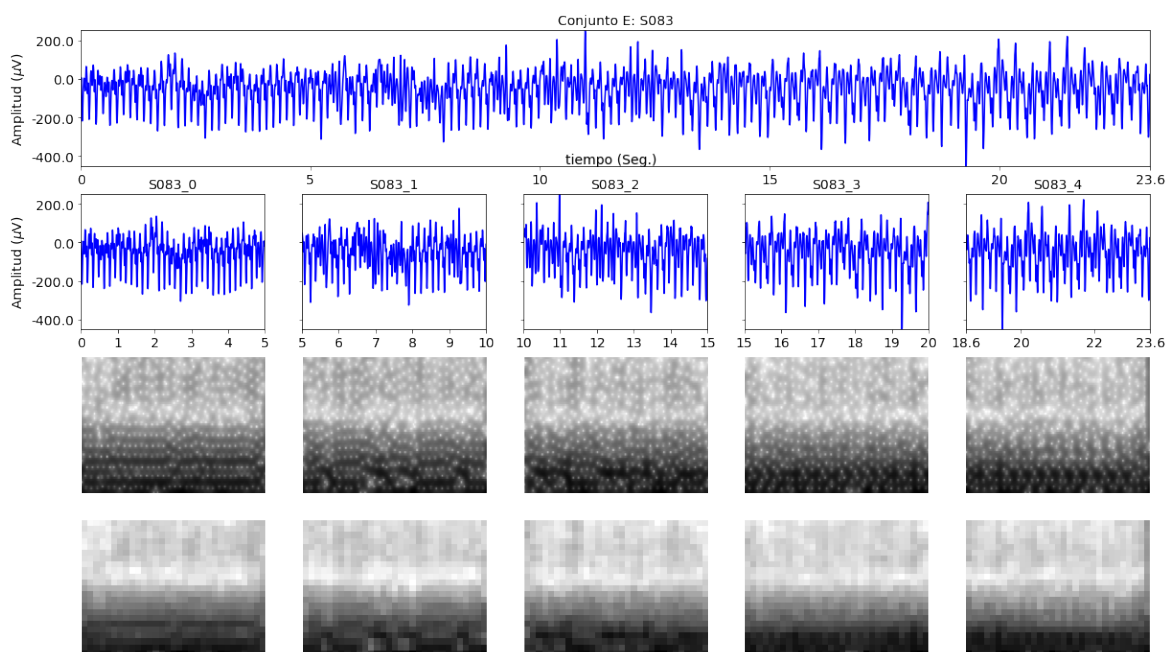
que hay que resaltar entre ambos clasificadores SVM es que el modelo obtenido con las señales segmentadas tiene menor número de vectores de soporte con menor  $\gamma$  (ancho de la Gaussiana mayor), lo que significa que es óptimo desde este punto de vista. No obstante, en el modelo SVM obtenido con las señales completas (no segmentadas) no se necesita segmentar las señales, lo que implica un menor uso del recurso computacional tanto para procesarlas como para obtener los rasgos de entrada a la misma.

## 5. Discusión

Se puede verificar, de los resultados obtenidos, que las imágenes de las señales obtenidas con la transformada SST-QSTFT, condujeron al desarrollo de modelos clasificadores basados en redes convolucionales, con muy alto desempeño



**Fig. 7.** Señal F092, del conjunto D, segmentada con su respectiva transformada SST-QSTFT con resolución de 496×368 píxeles y con resolución de 31×23 píxeles (inferior).



**Fig. 8.** Señal S083, del conjunto E, segmentada con su respectiva transformada SST-QSTFT con resolución de 496×368 píxeles y con resolución de 31×23 píxeles (inferior).

**Tabla 6.** Matriz de confusión promedio de entrenamiento con validación cruzada y dropout, señales segmentadas

		Predicción			Indicadores de desempeño (%)		
		Normal	Preictal	Ictal	Exactitud	Sensibilidad	Especificidad
Real	Normal	898,8	1,	0,1	99,69	99,87	99,58
	Preictal	5,1	893,1	1,8	99,56	99,23	99,78
	Ictal	0,6	1,9	447,5	99,80	99,44	99,89

**Tabla 7.** Matriz de confusión promedio de prueba con validación cruzada y dropout, señales segmentadas

		Predicción			Indicadores de desempeño (%)		
		Normal	Preictal	Ictal	Exactitud	Sensibilidad	Especificidad
Real	Normal	100,0	0,0	0,0	99,36	100,00	98,93
	Preictal	1,6	97,3	1,1	99,92	97,30	98,60
	Ictal	0,0	0,0	50,0	99,56	100,00	99,45

**Tabla 8.** Arquitectura del autoencoder

Tipo de Dato	Parámetros SVM		Número de Vectores de Soporte			
	$C$	$\gamma$	AB	CD	E	Total
Señal completa	100	1.0	74	64	54	195
Señal Segmentada	100	0.1	28	19	34	81

discriminatorio entre clases de señales epilépticas, incluso para el caso más desfavorable, para el entrenamiento, donde la señal completa se utiliza para obtener la transformada SST-QSTFT, ya que se obtiene un conjunto de entrenamiento de pocas muestras.

El aumento de los datos, obtenidos por segmentación, condujo a un mayor número de muestras disponible (cinco veces a los conjuntos originales) para entrenamiento, lo que aunado a la transformación tiempo frecuencia basado en la SST-QSTFT resultó en un modelo con un altísimo desempeño de discriminación de las clases.

El desempeño de los clasificadores basados en las SVM fue superior al desempeño de la CNN, pero hay que resaltar que los rasgos de entrada a la SVM fueron obtenidos con las CNN entrenadas previamente, lo que confirma que las CNN extraen los rasgos más relevantes de los datos.

La Tabla 9 muestra el excelente desempeño logrado por las propuestas de este trabajo, cuando se compara con otros trabajos previos que utilizaron la misma base de datos y definieron las mismas clases de interés, e incluso superan el desempeño de las propuestas previas.

## 6. Conclusiones

Este trabajo presentó la propuesta de un clasificador multiclase para clasificar señales EEG en tres clases (Normal, Preictal e Ictal), que se basa en una transformación tiempo-frecuencia SST-QSTFT para convertir la señal temporal en una imagen, que utiliza el clasificador CNN para clasificar la señal en una de las tres clases definidas.

Se plantearon cuatro enfoques de clasificadores, dos basados en la CNN, y dos basados



**Tabla 9.** Comparación de diferentes métodos de clasificación de tres clases

Método	Clasificador	Exac.	Sens.	Espec.
Türk y Özerdem (2019) [23]	CWT + CNN	98,01	96,92	98,59
Acharya et al. (2018) [1]	1D-CNN	88,70	95,00	90,00
Ullah et al. (2018) [24]	1D-CNN	99,10	94,60	96,84
Zhao et al. (2020) [26]	1D-CNN	96,97	-	-
Ramakrishnan et al. (2019) [15]	1D-CNN	98,37	-	-
Agrawal et al. (2019) [2]	Transferencia de Aprendizaje	100	-	-
Hussein et al. (2019) [10]	LSTM	100	100	100
<b>Propuesta de este Estudio</b>	SST-QSTFT + CNN + Señal Completa	94,67	91,67	95,81
<b>Propuesta de este Estudio</b>	SST-QSTFT + CNN + Señal Segmentada	99,61	99,10	98,99
<b>Propuesta de este Estudio</b>	SST-QSTFT + CNN + SVM	100	100	100

en la CNN como extractor de rasgos más un clasificador SVM. Dos clasificadores (CNN y SVM) se entrenaron con las señales completas, y los otros clasificadores (CNN y SVM) se entrenaron con las señales segmentadas, para evaluar el efecto del aumento del conjunto de entrenamiento sobre el desempeño de los clasificadores. Los indicadores de exactitud, sensibilidad y especificidad mostraron un mejor desempeño de los clasificadores entrenados con el conjunto de datos aumentados mediante la segmentación de la señal.

Los resultados mostraron que la transformada STT-QSTFT resultó ser un pre-proceso clave de la señal temporal para convertirla en imagen y lograr un clasificador con alto desempeño en la estimación de la clase de la señal EEG.

Se verificó la fortaleza de las CNN como un extractor de los rasgos más relevantes para entrenar una SVM con desempeño altísimo en la predicción de la clase de la señal EEG, y además comparable con los resultados que han obtenidos otros estudios similares.

## Referencias

1. Acharya, U., Oh, S., Hagiwara, Y., Tan, H., J.H. and Adeli (2018). Deep convolutional neural network for the automated detection and diagnosis of seizure using eeg signals. *Computers in Biology and Medicine*, Vol. 100, pp. 270–278. DOI: 10.1016/j.combiomed.2017.09.017.
2. Agrawal, A., Gupta, C., Prachi, G. (2019). A deep transfer learning approach for seizure detection using rgb features of epileptic electroencephalogram signals. *2019 IEEE International Conference on Cloud Computing Technology and Science (CloudCom)*, pp. 367–373. DOI: 10.1109/CloudCom.2019.00063.
3. Andrzejak, R., Lehnertz, K., Mormann, F., Rieke, C., David, P., Elger, C. (2001). Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E*, Vol. 64, No. 061907, pp. 061907–1–061907–8. DOI: 10.1103/physreve.64.061907.
4. Berrian, A., Leung, J., Saito, N. (2015). Time-frequency feature extraction via synchrosqueezing transform and its application to data sonification. *Proceedings of the Annual Conference of the Japan Society for Industrial and Applied Mathematics*, pp. 336–337.
5. Berrian, A., Saito, N. (2017). Adaptive synchrosqueezing based on a quilted short-time fourier transform. *Proc. SPIE 10394, Wavelets and Sparsity XVII*, pp. 1039420. DOI: 10.1117/12.2271186.
6. Daubechies, I., Lu, J., Wu, H.-T. (2011). Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool. *Applied and Computational*

- Harmonic Analysis, Vol. 30, No. 2, pp. 243–261. DOI: 10.1016/j.acha.2010.08.002.
7. **Daubechies, I., Maes, S. (1996).** A nonlinear squeezing of the continuous wavelet transform based on auditory nerve models. In **Aldroubi, A., Unser, M.**, editors, *Wavelets in Medicine and Biology*, chapter 20. CRC Press, pp. 527–546.
  8. **Fukushima, K. (1980).** Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, Vol. 36, pp. 193–202.
  9. **Hubel, D., Wiesel, T. (1959).** Receptive fields of single neurons in the cat's striate cortex. *The Journal of Physiology*, Vol. 124, No. 3, pp. 574–591.
  10. **Hussein, R., Palangi, H., Ward and, R. K., Wang, Z. J. (2019).** Optimized deep neural network architecture for robust detection of epileptic seizures using eeg signals. *Clinical Neurophysiology*, Vol. 130, pp. 25–37. DOI: 10.1016/j.clinph.2018.10.010.
  11. **Kaiming, H., Zhang, X., Ren, S., Sun, J. (2016).** Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.
  12. **LeCun, Y. Bottou, L., Bengio, Y., Haffner, P. (1998).** Gradient-based learning applied to document recognition. Vol. 86, No. 11, pp. 2278–2324.
  13. **Neligan, A., Sander, J. (2014).** Epidemiology of seizures and epilepsy. In **Miller, W., John, Goodkin, P., Howard**, editors, *Epilepsy*, chapter 4. Wiley-Blackwell, pp. 28–32.
  14. **Oberlin, T., Meignen, S., Valérie, P. (2014).** The fourier-based synchrosqueezing transform. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 315–319. DOI: 10.1109/ICASSP.2014.6853609.
  15. **Ramakrishnan, S., Muthanatha, A., Saravanan, P. (2019).** Epileptic eeg signal classification using multi-class convolutional neural network. 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN).
  16. **Raschka, S., Mirjalili, V. (2017).** Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow. Packt Publishing Ltd, 2nd edition. Birmingham, UK. ISBN 978-1-78712-593-3.
  17. **Shoeb, A. H. (2009).** Application of machine learning to epileptic seizure onset detection and treatment.
  18. **Shorvon, S. (2005).** Handbook of epilepsy treatment. Forms, causes and therapy in children and adults. Blackwell Publishing.
  19. **Simonyan, K., Zisserman, A. (2014).** Very deep convolutional networks for large-scale image recognition. ArXiv preprint arXiv:1409.1556.
  20. **Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A. (2015).** Going deeper with convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9.
  21. **Thakur, G., Wu, H.-T. (2011).** Synchrosqueezing-based recovery of instantaneous frequency from nonuniform samples. *SIAM Journal on Mathematical Analysis*, Vol. 43, No. 5, pp. 2078–2095. DOI: 10.1137/100798818.
  22. **The MathWorks, I., .** Matlab 2019b. Natick, Massachusetts, United States.
  23. **Türk, O., Özerdem, M. (2019).** Epilepsy detection by using scalogram based convolutional neural network from eeg signals. *Brain Sciences*, Vol. 9, No. 5, pp. 115. DOI: 10.3390/brainsci9050115.
  24. **Ullah, I., Hussain, M., Aboalsamh, H. (2018).** An automated system for epilepsy detection using eeg brain signals based on deep learning approach. *Expert Systems with Applications*, Vol. 107, No. 1, pp. 61–71.
  25. **World Health Organization (2019).** Epilepsy. Available from: <http://www.who.int/en/news-room/factsheets/detail/epilepsy>.
  26. **Zhao, W., Zhao, W., Wang, W., Jiang, X., Zhang, X., Peng, Y., Zhang, B., Zhang, G. (2020).** A novel deep neural network for robust detection of seizures using eeg signals. *Computational and Mathematical Methods in Medicine*, Vol. 2020, pp. 1–9. DOI: 10.1155/2020/9689821.

*Article received on 05/08/2020; accepted on 25/03/2021.  
Corresponding author is Guillermo Montilla.*