

Classification of Electroencephalographic Signals Using Support Vector Machines with Cross-Training.

Sergio Villazana^{*,a}, Guillermo Montilla^a, César Seijas^a, Antonino Caralli^a, Antonio Eblen^b

^aCentro de Investigación en Bioingeniería, Universidad de Carabobo, Facultad de Ingeniería, Valencia, Venezuela.

^bCentro de Biofísica y Neurociencia de la Universidad de Carabobo. Departamento de Ciencias Fisiológicas, Facultad de Ciencias de la Salud, Valencia, Venezuela

Abstract.-

This paper presents an application of Support Vector Machines (SVM) with cross training to classify electroencephalographic (EEG) records in normal and pathological ones. Pathological EEG records correspond to people with proven epilepsy diagnosis. Cross training has proven to be a technique to selectively remove error support vectors, and reduce the complexity of the underlying function. A low complexity implies a high possibility of classifying samples with less error. In this technique, the optimization process is performed in two stages, 1) the dataset is divided into an initial number of subsets, and an SVM is trained for each subset, 2) the margin (distance) for each sample of entire dataset to each decision function is found, and the average margin is obtained for each sample. Those samples whose average margins are outside from a certain range are eliminated from the dataset. Then, the final SVM is trained with a reduced dataset. The results show a high performance in the classification of EEG patterns, resulting in a high rate greater than 95 % of accuracy.

Keywords: epilepsy; EEG; Support Vector Machines; cross training.

Clasificación de Señales Electroencefalográficas Utilizando Entrenamiento Cruzado de Máquinas de Vectores de Soporte.

Resumen.-

Este artículo presenta una aplicación de entrenamiento cruzado de Máquinas de Vectores Soporte (de sus siglas en inglés Support Vector Machines, SVM) para clasificar registros electroencefalográficos (EEG) en normales y patológicos. Los registros de EEG patológicos corresponden a personas con diagnóstico comprobado de epilepsia. El entrenamiento cruzado ha demostrado ser una técnica para eliminar selectivamente los vectores de soporte de error reduciendo la complejidad de la función subyacente, lo que implica una mayor posibilidad de clasificar las muestras con menor error. En esta técnica, el proceso de optimización se realiza en dos etapas, 1) Se divide el conjunto de datos en un número inicial (predefinido) de subconjuntos, y se entrena una SVM por cada subconjunto, 2) Se halla el margen (distancia) de cada uno de los datos del conjunto de datos total a cada función de decisión, y se obtiene el margen promedio de cada muestra. Las muestras cuyo margen promedio esté fuera de cierto rango se eliminan del conjunto de datos. Luego se entrena la SVM final con el conjunto de datos reducido. Los resultados demuestran un alto desempeño en la clasificación de los patrones electroencefalográficos, obteniéndose una alta tasa de precisión mayor al 95 %.

Palabras clave: epilepsia; EEG; Máquinas de Vectores Soporte; entrenamiento cruzado

Recibido: Febrero 2015

Aceptado: Mayo 2015

1. Introducción.

La epilepsia es un trastorno crónico del cerebro que afecta a las personas de todas las edades en todos los países del mundo y se caracteriza

* Autor para correspondencia

Correo-e: svillaza@uc.edu.ve (Sergio Villazana)

por convulsiones recurrentes. Alrededor de 50 millones de personas a nivel mundial tienen epilepsia [1]. Cerca del 80 % de las personas con epilepsia viven en países en vías de desarrollo [1] y el 5 % (del total mundial) de los pacientes con epilepsia son de Latinoamérica y el Caribe [2]. La epilepsia puede desarrollarse debido a, anomalías en la red neuronal del cerebro, desequilibrio de las sustancias químicas, que participan en la transmisión de información, llamadas neurotransmisores, o alguna combinación de estos factores.

El electroencefalograma (EEG) es la herramienta clínica de información que se utiliza para supervisar, diagnosticar y tratar los trastornos neurológicos relacionados con la epilepsia. El EEG es un registro de las oscilaciones de los potenciales eléctricos generados por fuentes de micro corrientes cerebrales y registrados con electrodos ubicados sobre la superficie del cráneo [3, 4] o directamente sobre la corteza cerebral, en cuyo caso recibe el nombre EEG profundo o electrocortigrama, ECoG. Los rangos de amplitudes pueden llegar aproximadamente a los $100\mu V$ para los EEG superficiales, y alrededor de 1 a $2mV$ para los ECoG. El rango de frecuencias de las ondas cerebrales va desde $0,5Hz$ hasta los $50Hz$ [5]. La medición puede ser de carácter bipolar (entre electrodos) o unipolar, colocando un electrodo adicional de referencia o tomando como referencia el potencial promediado de todos los electrodos. La frecuencia de muestreo debe estar por encima de dos veces la mayor frecuencia que posea el EEG, para garantizar satisfacer el límite inferior de muestreo de Nyquist en la banda de frecuencias de interés.

Las Máquinas de Vectores de Soporte (de sus siglas en inglés *Support Vector Machines*, SVM) [6] son potentes algoritmos de clasificación y regresión, cuyo cálculo y los requisitos de almacenamiento crecen exponencialmente con el número de vectores de entrenamiento. La Programación cuadrática es el núcleo utilizado para determinar los vectores de soporte y los coeficientes del problema. Los algoritmos que resuelven la programación cuadrática tienen un orden de complejidad que escala con el cubo

del número de datos de entrenamiento, $O(n^3)$. Existen varias técnicas para acelerar el algoritmo de programación cuadrática que dividen los datos de entrenamiento en pequeños subgrupos (trozos) y resuelven el problema de optimización iterativa hasta que se alcanza el óptimo global [7, 8]. La optimización mínima secuencial [8] es el algoritmo más utilizado para resolver el problema de programación cuadrática. Esta técnica reduce el tamaño de los datos a 2 vectores y resuelven analíticamente el problema de optimización.

El entrenamiento cruzado es un procedimiento para reducir el número de vectores de soporte no acotados o de error. Con el entrenamiento cruzado es posible crear una distribución separable de ejemplos de entrenamiento con un impacto mínimo en la posición de la función de decisión [9], y romper la complejidad de SVM durante la etapa de entrenamiento.

El clasificador basado en SVM, de ahora en adelante SVC, está basado en extracción de rasgos de espectros de orden superior de las series de tiempo derivadas de los EEG aplicados a pacientes epilépticos y pacientes de control. Los EEG son series de tiempo altamente no lineales y no gaussianas que exhiben espectros de orden superior, los cuales se extraen y se utilizan como rasgos que mejoran el desempeño del SVC [10].

Se entiende como estadísticos de orden superior (HOS, de sus siglas en inglés: “*Higher Order Statistics*”), los momentos y cumulantes de orden mayor al segundo (los momentos de orden 1 y 2 son la media y la varianza). Los momentos de orden superior son generalizaciones naturales de la autocorrelación y los cumulantes son combinaciones lineales específicos de estos momentos. Los cumulantes de orden 3 y 4 son la *asimetría* y el *apuntalamiento* [11]. La *asimetría* es una medida del grado en que una distribución de probabilidad de una variable aleatoria real “se inclina” hacia a un lado de la media. La *asimetría* puede ser positivo o negativa, o incluso no definido. El *apuntalamiento* es una medida del apuntamiento de la distribución de probabilidad de una variable aleatoria real.

La bicoherencia o coherencia biespectral es una versión normalizada al cuadrado del biespectro

de una señal. La bicoherencia es una medida para cuantificar el grado de acoplamiento de fase entre componentes espectrales de una señal. El biespectro o densidad biespectral es un estadístico que se usa para buscar interacciones no lineales entre componentes espectrales de una señal [12]. El biespectro es la transformada de Fourier del cumulante de tercer orden (la *asimetría*). Si los cumulantes de tercer orden de un proceso son iguales a cero, entonces su biespectro es cero, y por lo tanto su bicoherencia también es cero. Si el biespectro no es cero, entonces el proceso es no-Gaussiana, y si el proceso es lineal y no-Gaussiana, la bicoherencia es una constante distinta de cero [11, 13, 14].

Este trabajo propone un entrenamiento cruzado de Máquinas de Vectores de Soporte con estadísticos de orden superior como rasgos de entrada para clasificar segmentos de señales EEG en normales y epilépticas. Este esquema reduce el número de vectores de soporte, mientras que se mantiene el error de generalización en niveles comparables al caso de entrenamiento del SVC utilizando todos los datos de entrenamiento en una sola etapa (una sola SVC).

2. Base de Datos.

La base de datos utilizada en este artículo se describe en [15]. Está conformada por cinco conjuntos A, B, C, D y E, cada uno con 100 señales monocanal de EEG de 23,6 segundos de duración cada uno. 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 de 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 corresponden a EEG profundos. Las señales en el conjunto C se obtuvieron de la formación del hipocámpal (ubicado en el lóbulo temporal medio) del cerebro. El conjunto D se obtuvo dentro de la zona epileptogénica. Los

conjuntos C y D sólo contienen registros de la actividad cerebral medida durante los intervalos sin crisis (interictal). El conjunto E contiene registros durante la actividad convulsiva, o periodo ictal.

Todos estos segmentos de EEG se registraron con un amplificador de 128 canales, un convertidor analógico-digital de 12 bits y, a una frecuencia de muestreo de 173,61 Hz. El número de registros en total es 500, con 4097 muestras cada uno (23,6 segundos de registro por EEG). La Figura 1 muestra cinco segmentos de 10 segundos cada uno identificados con las letras “a” a la “e”. Estas letras indican el conjunto al cual pertenecen los segmentos. La unidad de los ejes verticales de cada uno de los segmentos es μV , pero, no están a la misma escala.

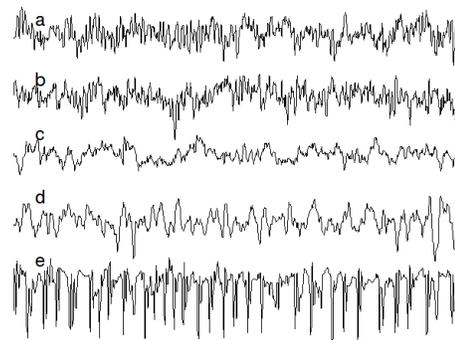


Figura 1: Segmentos de 10 segundos de ejemplo de cada uno de los conjuntos. La letra minúscula a la izquierda indica el conjunto al cual pertenece cada segmento. Cada segmento se representa a escalas diferentes en μV .

3. Revisión Teórica.

3.1. Clasificador SVM de Margen Suave (C-SVM).

Este trabajo está centrado en el clasificador SVM de margen suave (del inglés *soft margin classifier*) conocido como C - SVM [6, 16, 17], donde C es la constante de regularización positiva que controla el balance entre maximizar el margen y minimizar el error. Dado n ejemplos de entrenamiento x_i y sus clases asociadas $y_i = \pm 1$, la

función de decisión SVM, $f(x)$, es:

$$f(x) = \sum_{i=1}^n \alpha_i^* y_i K(x_i, x) + b^* \quad (1)$$

donde α_i^* son los coeficientes de Lagrange, $K(\cdot, \cdot)$ es la matriz kernel, b^* es la distancia de la función de decisión al origen, y x es cualquier vector de prueba, que en general no pertenece al conjunto de entrenamiento.

El coeficiente α_i^* se obtiene resolviendo el problema:

$$\alpha_i^* = \operatorname{argmax}_{\alpha} \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (2)$$

sujeto a las restricciones:

$$\forall i, 0 \leq \alpha_i \leq C \text{ y } \sum_i \alpha_i y_i = 0$$

El problema de programación cuadrática produce tres categorías de ejemplos de entrenamiento que dependen del valor de α_i^* definidas según la ecuación (2), que deben cumplir con las condiciones Karush-Kuhn-Tucker [6, 17]. La Tabla 1 resume las categorías de las muestras de entrenamiento. El margen en la tercera columna de la Tabla 1 representa la distancia de la muestra a la función de decisión multiplicada por su etiqueta o clase correspondiente.

Los vectores de soporte de frontera son todos aquellos vectores ubicados a un margen normalizado 1 de la función de decisión, están bien clasificados, y son los que definen los hiperplanos +1 y -1. Todos aquellos vectores de soporte no acotados o de error tienen margen negativo, lo que significa que están mal clasificados. Los ejemplos con $\alpha_i^* = 0$ están bien clasificados, y tienen margen mayor que uno. Estos ejemplos no contribuyen en la función de decisión SVM (ecuación (1)). Si se reentrena la SVM, manteniendo los parámetros de ajuste (valor de C y los parámetros del kernel, K , utilizado), descartando estos ejemplos se producirá la misma función de decisión (ecuación (1)).

3.2. Entrenamiento Cruzado.

La función decisión f puede ser expresado por múltiples combinaciones lineales de las funciones base $K(x_i; \cdot)$, ya que dichas funciones base $K(x_i; \cdot)$ asociadas a los vectores de soporte x_i no constituyen necesariamente una familia linealmente independiente [9].

Existen métodos de reducción que intentan seleccionar un subconjunto de los vectores de soporte que sea suficiente para expresar la función de decisión SVM. En [18, 19, 20] afirman que el conjunto de vectores de soporte obtenido no es linealmente independiente, por tanto, dicho conjunto se puede reducir para producir la misma función de decisión. En [21] propone calcular la forma reducida o canónica por filas de la matriz kernel y eliminar aquellos vectores de soporte que conducen a filas cero.

Algoritmo para el entrenamiento cruzado. El entrenamiento cruzado de SVM está inspirado en los algoritmos de edición de conjunto de entrenamiento propuestos por Hart [22]. El entrenamiento cruzado comienza dividiendo el conjunto de entrenamiento en S subconjuntos de tamaño r y entrena S SVMs independientes en cada subconjunto s_i . Las funciones de decisión de estas SVMs se utilizan para descartar ciertos ejemplos de entrenamiento. Los ejemplos de entrenamiento desechados son aquellos cuyo margen no cumple cierto criterio. El algoritmo se describe a continuación [22]:

1. Dividir los datos en S subconjuntos de igual tamaño que no se solapan, s_1, \dots, s_S .
2. Entrenar las S SVMs independientes f_1, \dots, f_S utilizando cada uno de los subconjuntos como el conjunto de entrenamiento.
3. Calcular el margen promedio para cada ejemplo de entrenamiento (x_i, y_i) .
4. Eliminar todos los ejemplos de entrenamiento cuyo margen promedio $m_i \notin [0, 1]$ (el que se encuentra fuera del intervalo cerrado).
5. Entrenar la SVM final sobre los ejemplos de entrenamiento restantes.

La Figura 2 muestra un esquema de entrenamiento cruzado en particular, donde $S = 8$. Como

Tabla 1: Categorías de las muestras de entrenamiento en función de α_i^* .

Categoría de la muestra	Condición	Margen
Vectores de Soporte de frontera	$0 < \alpha_i^* < C$	$y_i f(x_i) = 1$
Vectores de Soporte no acotados o de error	$\alpha_i^* = C$	$y_i f(x_i) < 0$
Vectores ordinarios	$\alpha_i^* = 0$	$y_i f(x_i) > 1$

se puede ver en la Figura 2, el entrenamiento cruzado tiene una validación cruzada intrínseca que utiliza S conjuntos, ya que una fracción (s_i) de todo el conjunto de datos (TD) se utilizan para entrenar a una SVM y el conjunto de datos restantes se prueban en esta SVM. La única variación es que el subconjunto s_i se prueba también, porque es necesario para encontrar el margen de cada muestra de entrenamiento en el subconjunto actual.

3.3. Estadísticos de Orden Superior.

La asimetría de una señal se define como:

$$C_{3x}(0, 0) = E \{x^*(n) x(n) x(n)\}$$

El apuntalamiento (otro término es planitud) se define como:

$$C_{4x}(0, 0, 0) = E \{x^*(n) x(n) x(n) x^*(n)\} \\ - 2C_{2x}(0) C_{2x}(0) \\ - M_{2x}^*(0) M_{2x}(0)$$

donde $C_{2x}(0)$ es la varianza de la señal.

El biespectro, $P_{3x}(f_1, f_2)$, y su cumulante está relacionado a la Transformada de Fourier de la señal respectiva por:

$$P_{3x}(f_1, f_2) = E \{X^*(f_1 + f_2) X(f_1) X(f_2)\} \\ = \sum_{k, m=-\infty} C_{3x}(k, m) e^{-2\pi i(f_1 k + f_2 m)}$$

En [9, 11] puede hallar un desarrollo más extenso de los estadísticos de orden superior.

4. Parte Experimental.

Los EEG de la base de datos ya están filtrados con un filtro pasa banda entre 0.53 Hz y 40 Hz

[22]. Se eliminó el ruido de línea con un filtro de ranura de 50Hz. Se obtuvo 2098 subsegmentos de 2000 muestras cada uno solapados 1999 muestras por cada conjunto, en total, se generaron 10490 subsegmentos de EEG. Los subsegmentos de los conjuntos A y B se etiquetaron con -1 (normal) y aquellos pertenecientes a los conjuntos C, D y E se etiquetaron +1 (epiléptico). A cada subsegmento se le removió la media, y se calcularon los siguientes rasgos: potencia de la señal, *asimetría*, *apuntalamiento* y la máxima coherencia. Los rasgos se normalizaron entre ± 1 . El conjunto de datos para entrenar y validar está compuesto de 10490 filas (muestras) por 4 columnas correspondientes a cada rasgo. Se tomó el 80% del conjunto de datos para entrenar y el 20% para validar. Los datos de entrenamiento se dividieron en diez subconjuntos. En todas las SVMs de la primera etapa, ver Figura 2, se utilizó el algoritmo $C - SVM$ de *LIBSVM* [23], con el kernel *RBF* ($\sigma = 20 \rightarrow \gamma = 0,00125$), en las diez SVMs se fijó $C = 10^3$, en la última SVM, C se fijó en ∞ [10], en la práctica se fijó en 10^8 .

Las medidas de desempeño fueron: la *sensibilidad* (fracción de casos verdaderos positivos, VP , clasificados como positivos con respecto a los VP):

$$sensibilidad = \frac{V}{(VP + FN)}$$

donde FN son los falsos negativos, es decir, los casos positivos clasificados como negativos. La *especificidad* (fracción de los casos verdaderos negativos, VN , clasificados como negativos con respecto a los VN):

$$especificidad = \frac{VN}{(VN + FP)}$$

donde FP son los falsos positivos, es decir, los

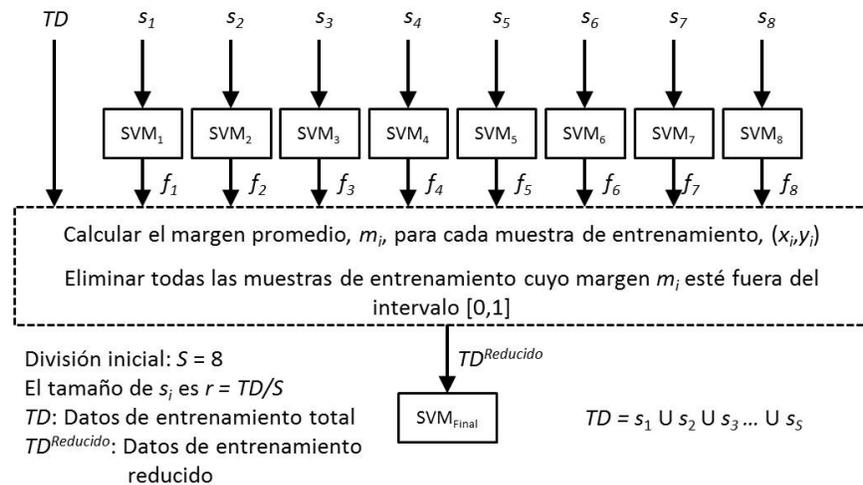
Figura 2: Entrenamiento cruzado de SVM. Caso particular de $S = 8$.

Tabla 2: Medidas de desempeño de los entrenamientos cruzado y convencional (una SVM).

Esquema	Fase	Nº VS	Sensibilidad	Especificidad	Precisión	Valor F	Exactitud
Cruzado	Entrenamiento	25	0,9994	0,9999	0,9999	0,9997	0,9996
	Validación	-	0,9154	0,9440	0,9672	0,9360	0,9270
Una SVM	Entrenamiento	2283	0,9802	0,8973	0,9349	0,9570	0,9471
	Validación	-	0,9808	0,9139	0,9438	0,9619	0,9538

casos negativos clasificados como positivos. La *precisión* (fracción de los casos *VP* con respecto a todos los ejemplos clasificados como positivos, *VP* y falsos positivos):

$$\text{precisión} = \frac{VP}{(VP + FP)}$$

La *exactitud* (fracción de los casos clasificados *VP* y *VN* con respecto a todos los ejemplos):

$$\text{exactitud} = \frac{VP + VN}{VP + VN + FP + FN}$$

Por último, el *Valor F* (*F-score*), esta última medida combina la *precisión* y la *sensibilidad*. El *Valor F* está dado por:

$$\text{ValorF} = 2 \frac{\text{precisión}(\text{sensibilidad})}{\text{precisión} + \text{sensibilidad}}$$

Se hicieron múltiples corridas de entrenamiento cruzado y se tomaron los valores promedios que

se muestran en la Tabla 2. La Tabla 2 también muestra las medidas de desempeño del clasificador obtenido usando el entrenamiento convencional de una sola SVM.

En la Tabla 2 se observa una gran reducción del número de vectores de soporte (de aproximadamente 91 a 1) con altas *sensibilidad* (91,54% de los *VP* fueron clasificados como positivo) y *especificidad* (94,40% de los *VN* fueron clasificados como negativo), también tiene alta *precisión* (el 96,72% de los clasificados como positivo son *VP*, es decir, el 3,28% de los casos *VN* fueron clasificados como positivo). Es evidente el alto *Valor F*, mayor que la *sensibilidad*, ya que la *precisión* fue bastante alta. También se puede ver la Tabla 2 que la *precisión* del clasificador obtenido con el entrenamiento cruzado es mayor que la obtenida con el clasificador obtenido convencionalmente.

5. Conclusiones.

Los resultados demostraron que el clasificador obtenido con entrenamiento cruzado tuvo alta precisión, es decir, más del 95 % de los casos epilépticos fueron clasificados como tal (positivo) y menos del 4 % de los casos normales fueron clasificados epilépticos (falso positivo). El número de vectores de soporte se redujo significativamente (0,30 % y 27,20 % de los datos de entrenamiento con entrenamiento cruzado y convencional, respectivamente) con un mínimo impacto en la exactitud del clasificador. En general, el entrenamiento cruzado demostró ser una metodología potente para minimizar el número de vectores de soporte y mantener bajo el error de generalización.

Referencias

- [1] WHO. Epilepsy, fact sheet N° 999. <http://www.who.int/mediacentre/factsheets/fs999/en/index.html>, Octubre 2008.
- [2] C. Acevedo et al. Informe sobre la epilepsia en latinoamérica. Technical report, Organización Panamericana de la Salud, el Departamento de Salud Mental y Abuso de Sustancias de la Organización Mundial de la Salud, la Liga Internacional Contra la Epilepsia y el Buró Internacional para la Epilepsia, Panama, 2008.
- [3] P. Nunez. Quantitative eeg analysis methods and clinical applications. In Shanbao Tong and Nitish V. Thakor, editors, *Physiological Foundations of Quantitative EEG Analysis*, chapter 1, pages 1–22. Artech House, 685 Canton Street, Norwood, MA 02062, 2009.
- [4] P. Nunez and R. Srinivasan. Electroencephalogram. *Scholarpedia*, 2(2):1348, 2007. revision #91218.
- [5] J. Malmivuo and R. Plonsey. *Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields*. Oxford university press, 1995.
- [6] Vladimir Naumovich Vapnik and Vladimir Vapnik. *Statistical Learning Theory*, volume 1. Wiley New York, 1998.
- [7] Thorsten Joachims. Making large-scale support vector machine learning practical. In *Advances in kernel methods*, pages 169–184. MIT Press, 1999.
- [8] J. Platt et al. Fast training of support vector machines using sequential minimal optimization. *Advances in kernel methods-support vector learning*, 3:1–21, 1999.
- [9] C. Seijas, A. Caralli and S. Villazana. Neuropathology classifier based on higher order spectra. *Journal of Computer and Communications*, 1(04):28, 2013.
- [10] Léon Bottou, Jason Weston and Gökhan H. Bakir. Breaking svm complexity with cross-training. In *Advances in Neural Information Processing Systems*, pages 81–88. Citiseer, 2005.
- [11] A. Swami, C. Mendel, and C. Nikias. Higher-order spectral analysis (hosa) toolbox. *Version*, 2(0):3, 2000.
- [12] Kuang Chua Chua, Vinod Chandran, U. Rajendra Acharya and C. Lim. Application of higher order spectra to identify epileptic eeg. *Journal of Medical Systems*, 35(6):1563–1571, 2011.
- [13] T. Ning and J. D. Bronzino. Autoregressive and bispectral analysis techniques: Eeg applications. *Engineering in Medicine and Biology Magazine, IEEE*, 9(1):47–50, 1990.
- [14] V. Salai Selvam and S. Shenbagadevi. Bispectral analysis of scalp electroencephalograms: Quadratic phasecoupling phenomenon in detecting brain tumor. *American Journal of Applied Sciences*, 10(3):294–306, 2013.
- [15] Ralph G. Andrzejak, Klaus Lehnertz, Florian Mormann, Christoph Rieke, Peter David, and Christian E. Elger. 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*, 64(6):061907, 2001.
- [16] Christopher J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2):121–167, 1998.
- [17] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 2000.
- [18] N. Syed, H. Liu and K. Sung. Incremental learning with support vector machines. In *International Joint Conference on Artificial Intelligence*, 1999.
- [19] Christopher J. C. Burges et al. Simplified support vector decision rules. In *ICML*, volume 96, pages 71–77. citeseer, 1996.
- [20] Christopher J. C. Burges and Bernhard Schölkopf. Improving the accuracy and speed of support vector machines. In *Advances in Neural Information Processing Systems*, pages 375–381. Mit press, 1997.
- [21] T. Downs, K. Gates and A. Masters. Exact simplification of support vector solutions. *The Journal of Machine Learning Research*, 2:293–297, 2002.
- [22] P Hart. The condensed nearest neighbor rule (corresp.). *IEEE Transactions on Information Theory*, 14(3):515–516, 1968.
- [23] Chih-Chung Chang and Chih-Jen Lin. Libsvm: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27, 2011.