
IMPLEMENTACIÓN DE LA EXTRACCIÓN AUTOMÁTICA DE INFORMACIÓN CLÍNICA Y PARACLÍNICA RELEVANTE DEL EXPEDIENTE MÉDICO PARA LA BÚSQUEDA DE PATRONES PRECURSORES A EVENTOS EPILÉPTICOS

Dra. Laura Ivoone Garay Jiménez

lgaray@ipn.mx,

Dra. Blanca Tovar Corona

bltovar@ipn.mx,

SISTEMAS-UPIITA, Instituto Politécnico Nacional

M.T.A. Sunaina Singh Múgica

sakurita_1034@hotmail.com

M.T.A. Carlos Alberto Ramirez Fuentes

ing.carlosaramf@outlook.com

Alumnos SEPI-UPIITA, Instituto Politécnico Nacional

Resumen

Este artículo describe los resultados de la metodología empleada para generar parte de un sistema auxiliar en la toma de decisiones en el área médica. El primer software IdgEEG permite identificar automáticamente tanto los grafoelementos como eventos característicos dentro de los electroencefalogramas (EEG). El segundo programa MAED sistematiza el ingreso de los parámetros relevantes de los expedientes según las guías clínicas y establece su ponderación en el diagnóstico de pacientes previamente diagnosticados con epilepsia. IdgEEG, tuvo una eficiencia del 96.8% en los casos analizados y el evaluador de congruencia de MAED encontró que de los casos analizados, el 95% de los expedientes tuvo datos clínicos y paraclínicos congruentes. Esta información acondicionada e integrada, conforma la base de conocimiento para los sistemas de búsqueda de patrones asociados a los eventos epilépticos.

I. Introducción

En el área médica es importante tener un diagnóstico clínico e información de soporte auxiliar para corroborarlo o descartarlo. La historia médica completa permite establecer el estado de salud del paciente [1,2] y entre más apegada sea su conformación a las recomendaciones de las guías clínicas propuestas en forma consensada en las asociaciones de especialistas y más información adicional relevante sea anexada, más certero será el diagnóstico y tratamiento.

Para el caso del diagnóstico de la epilepsia, una enfermedad neurológica caracterizada por eventos eléctricos de alta energía en el cerebro que alteran la conciencia y/o desencadenan

convulsiones de epilepsia mioclónica, se toman en cuenta dos tipos de información: la información obtenida por el médico a través de una entrevista con el paciente y su exploración física. Así como, al menos el resultado del estudio estándar denominado electroencefalografía, que consiste en el registro de las señales eléctricas generadas por el cerebro siguiendo un protocolo de registro. Debido a que los eventos epilépticos no son predecibles, el protocolo de registro del EEG requiere que sea tomado sin la sintomatología y que el paciente sea expuesto a estímulos desencadenantes de los eventos tales como la privación de sueño, estímulo luminoso, auditivo, cognitivo e hiperventilación [1,3]. En caso de sospecha de lesión física en el cerebro, se puede recurrir a diferentes técnicas de imagenología tales como resonancia magnética craneal (MRI), tomografía por emisión de positrones (PET) y tomografía computarizada con emisión fotónica (SPECT), debido a su costo no es un procedimiento de rutina.

Es importante recalcar que un estudio de EEG nunca podrá establecer o descartar por sí mismo el diagnóstico de epilepsia, ya que no siempre hay repercusiones clínicas de la enfermedad asociados a los eventos epilépticos. Pero puede usarse como una herramienta de monitoreo o análisis de la enfermedad una vez diagnosticada siempre complementándose con las observaciones de los médicos.

Por otro lado, desde el punto de vista biomédico, es deseable auxiliar a los médicos en el análisis de la gran cantidad de información y en su entrenamiento para el diagnóstico, es por ello que existen algunas propuestas que se basan en sistemas expertos los cuales se caracterizan por ser una herramienta computacional especializada en resolver un problema dado, utilizando el análisis de las bases de conocimiento relevante en la solución del problema, así como las reglas de inferencia asociadas a las distintas combinaciones de los parámetros en la toma de decisión. En este campo en especial, el análisis de expedientes clínicos, la base de conocimiento se consideró conveniente dividir la información en dos grandes rubros: El primero es la información clínica proporcionada por el paciente, sus familiares y el médico tratante, la cual puede ser cualitativa y/o cuantitativa y puede presentar incertidumbre o incluso falta parcial de ella. En el segundo caso, la información extraída a partir del registro EEG que es cuantitativa pero que presenta muchas formas y orígenes posibles, por lo que se considera una señal aleatoria compleja resultante de la actividad de un conjunto de cientos de neuronas que se encuentran debajo de cada zona de análisis. En el estado del arte se observa una intensa actividad en este rubro con el objetivo final de detectar, y extraer información de los picos epilépticos tales como su conformación espacial y su frecuencia [1]. La presencia de eventos precursores en los periodos interictales, podría ser un elemento de análisis interesante para la epiletogénesis o para identificar nuevas terapias de tratamiento y monitoreo [4]. Otro enfoque interesante a nivel clínico es que las enfermedades son el producto de n-parámetros no siempre considerados por el médico porque no tienen una relación evidente o visualmente detectable y en ocasiones es muy difícil su interpretación. La generación de sistemas expertos para abordar esta metodología de manera sistemática ha sido utilizada en varias enfermedades, pero hasta el momento no se tiene conocimiento de un sistema diseñado para

asistir en esta patología en especial que combine la información de ambos orígenes [1,4,5]. Por lo que en este trabajo se analizaron los dos tipos de información y se adecuaron para dejarlos disponibles para su aplicación en el estudio y en su caso diagnóstico de la epilepsia a través del diseño, implementación y evaluación de las técnicas disponibles en el campo de la inteligencia artificial y soft-computing [6].

II. Diseño del software de análisis e inferencia de la base de conocimiento

Ya que el diagnóstico médico contempla ambos tipos de información (clínica y paraclínica) y son las guías clínicas el estándar donde se concentra el conocimiento médico y uniformizan los criterios tanto para el diagnóstico como tratamiento de la epilepsia, fue necesario hacer un estudio a profundidad de ellas, para establecer los parámetros cuantitativos y cualitativos más predominantes en el diagnóstico, auxiliados por un médico especialista. A partir del levantamiento de requerimientos del sistema, se propuso probar métodos que permitieran buscar bajo el mismo criterio y de forma automática los patrones y secuencias de elementos identificados como relevantes en el diagnóstico de la epilepsia, eliminando así la subjetividad.

La solución se dividió en tres grandes rubros:

- a) Los datos cualitativos y los cuantitativos proporcionados por los pacientes, familiares y el médico que pueden estar o no presentes en el expediente y que no siempre están descritos de forma precisa
- b) El análisis sistemático de la señal electroencefalográfica bajo un protocolo clínico de búsqueda de detonadores de crisis, basado en los parámetros cuantitativos siempre presentes durante todo el registro y que el médico evalúa como parte de la interpretación de la información y c) La evaluación de la congruencia de la información con lo esperado para esta enfermedad (Figura 1).

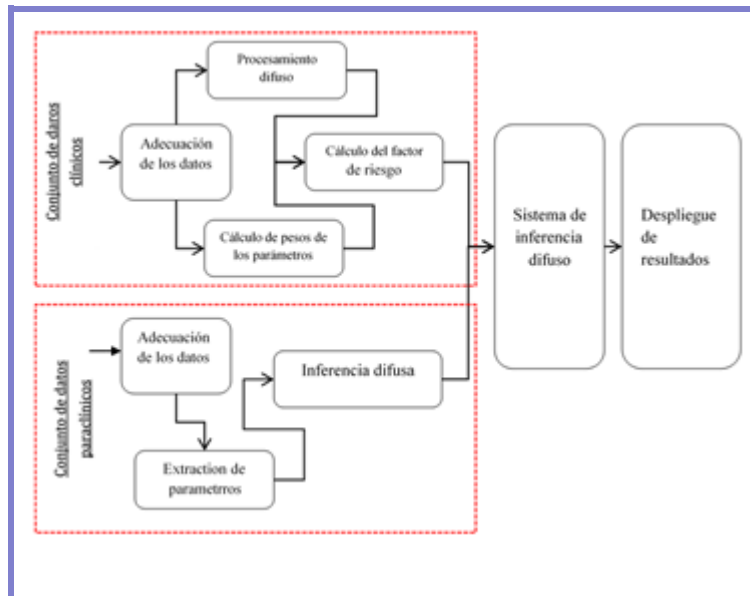


Figura 1. Diseño propuesto para la identificación de parámetros relevantes en EEG tanto clínicos como paraclínicos

A continuación se describen las condiciones y diseño de cada etapa considerada, así como las técnicas de evaluación utilizadas para el desempeño del sistema total.

a. Módulo de Extracción de parámetros clínicos

Como parte del trabajo realizado se diseñó e implementó un software que primero solicita en forma sistemática la información necesaria para realizar el diagnóstico diferencial de la enfermedad según las guías clínicas a través de una interfaz gráfica. Una vez que el médico ingresa la información (Tabla 1), el programa hace la transformación difusa de la información que lo requiera o las pondera de acuerdo a su relevancia en el diagnóstico final.

La lista de los parámetros considerados y las variables asociadas, se encuentran resumidas en la Tabla 1 y como se puede observar son 46 variables que pueden ser continuas o discretas, dicotómicas, policotómicas, nominales, cualitativas o cuantitativas. En la Figura 3 se muestra en la parte superior izquierda el menú de cuestionarios y en la pantalla derecha, un ejemplo de uno de ellos. Este módulo permite uniformizar el procedimiento para introducir los datos clínicos y manejar la incertidumbre o falta de información utilizando la lógica difusa como técnica de inferencia y de interpretación lingüística.

b. Módulo de Extracción de parámetros paraclínicos

Por otro lado, los registros de las señales EEG obtenidas en los protocolos de registro se subdividieron en sueño y vigilia. Posteriormente, se utilizó el clasificador IdgEEG ©, el cual identifica de forma automática, en segmentos de un segundo, los patrones de acuerdo a sus frecuencias y amplitudes en normal y anormal (Figura 2, bloque violeta) y los etiqueta, lo cual permite calcular la prevalencia de cada comportamiento por electrodo dentro del segmento analizado, así como la duración de este comportamiento con respecto al total del tiempo registrado.

Otro de los parámetros de interés para el neurólogo es como cambian espacialmente los voltajes y frecuencias de la señal a lo largo del cerebro, ya sea por tipos de ensambles o por zonas funcionales. Para ello se calculan los gradientes de voltaje y frecuencia. En la Figura 2 los bloques azules resumen los vectores de información considerados.

Tabla 1. Parámetros clínicos con el número de variables asociadas a cada uno que se consideraron en la evaluación de congruencia

Grupo	Parámetro	Variables	Numero de variables	Tipo de variable		
1	Datos del paciente	Sexo	1	cualitativa / nominal		
		Edad	1	cuantitativa / continua		
2	Historia familiar	Directa / Indirecta	2	cualitativa / policotómica		
		Riesgo de aborto, nacimiento prematuro, cesárea, presencia de polihidramnios, oligodramnios, calificación de placenta, pre-eclampsia, eclampsia, diabetes gestacional	12	cualitativa / dicotómica		
	Condición perinatal	Edad gestacional	1	Cualitativa/continua		
		Peso al nacer (Kg)	1	cuantitativa / continua		
		sufriamiento fetal, ventilación asistida, apnea	3	cualitativa / dicotómica		
	Desarrollo Neurológico	APGAR	1	cualitativa / ordinal		
		Motor, lenguaje, atención, impulsividad, agresividad y desinhibición	6	(Normal/Deficiente)	cualitativa/dicotómica	
		Nivel escolar, Desempeño escolar	2			
		Historial de eventos	traumatismo craneal, neuro-infecciones, hospitalizaciones previas, otras condiciones	4	cualitativa / dicotómica	
			Tipo de crisis	Motor, lado del cerebro afectado, inconciencia durante el evento, afección en el sistema sensoriomotor, autonómico, gástricos, uncimatos, psicológicos, otros síntomas relevantes.	10	cualitativa / dicotómico
	Duración (min)		1	Quantitative / continuo		
	Frecuencia de las crisis		1	cualitativo/dicotómico		

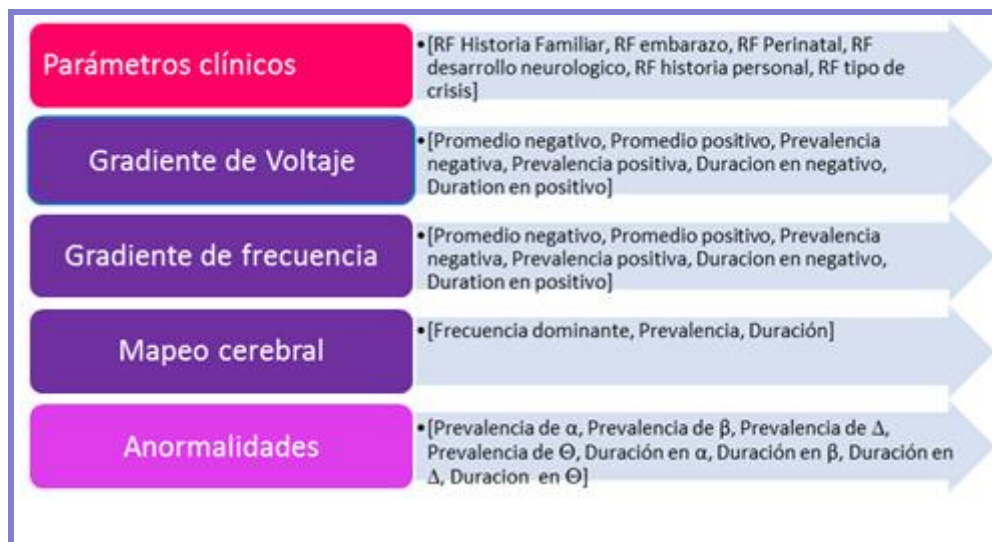


Figura 2. Parámetros clínicos y paraclínicos considerados para la evaluación del caso clínico

Para establecer la congruencia entre los dos tipos de información y que a su vez permitirá tomar decisiones en etapas posteriores, se integraron las bases de datos con la información de la Figura 2, es decir cada caso clínico fue representado por un vector de información con ese formato. La interfaz empleada para armar la base de datos se observa en la Figura 3, la cual contiene un menú donde se selecciona el cuestionario por parámetro (A), una vez seleccionado se abre cada cuestionario donde se recopila la información y se pondera de acuerdo al tipo de dato, su universo y su relevancia y se obtiene un factor de riesgo(B). Al terminar de ingresar la información se habilita la página de análisis de la señal Electroencefalografica y se obtienen todos los parámetros asociados a la etapa de vigilia y sueño del vector (C). Finalmente el programa evalúa la congruencia y pone a disposición del usuario la base de conocimiento(D).



Figura 3. Interfaz del sistema experto de ponderación y evaluación de congruencia entre tipos de información, parte de MAED©.(A) menú principal de la captura del historial clínico del paciente, (B) ejemplo de cuestionario de los parámetros clínicos y resultado mostrado en el ovalo rojo,(C) pantalla de solicitud de análisis del EEG y (D) resultado arrojado del análisis de congruencia y botón de solicitud de acceso a base de conocimiento

c. Módulo de inferencia de la congruencia entre información del expediente

Los parámetros de entrada a este módulo del sistema se encuentran condensados en la Figura 2, donde se muestra el tipo de información contenida en cada vector de información. El primer paso en el módulo de inferencia consistió en fusificar la información de la prevalencia de acuerdo a los niveles considerados comúnmente en las guías clínicas. Posteriormente se agruparon los comportamientos siguiendo el criterio propuesto por el estándar ACNS [7] (Tabla 2).

Es importante mencionar que las combinaciones de todas las variables que evalúan el caso clínico generaron N reglas difusas tipo Mamdani.

$$N = \prod_{i=2}^k n_i$$

donde: i es el número de variables consideradas para la inferencia, n es el número de posibles niveles de la variable i y k es el número máximo de variables.

Tabla 2. Tabla de inferencia sobre las prevalencias y duración de las variables acorde con las directrices de interpretación por zonas

Prevalencia				
Duración		Ocasional	Frecuente	Muy frecuente
	Muy corto	C	PC	NC
	Corto	C	PC	NC
	Intermedio	PC	PC	NC
	Largo	NC	NC	NC

Este número de reglas permiten inferir la congruencia entre la información de los dos orígenes, para este primer caso se obtuvieron 12 reglas.

Una vez que se llevó a cabo el análisis para determinar los parámetros por zona funcional definidas en las guías como parte del estándar, se procedió a realizar la evaluación de la corteza cerebral completa denominada evaluación global. Siguiendo las mismas directrices del anterior y proponiendo un sistema de inferencia Mamdani se realizó con las siguientes características:

- Número de entradas requeridas: 5 que son Frontal, Central, Temporal, Parietal y Occipital, cada uno con 3 variables fusificadas posibles (controlado, Parcialmente controlado y no controlado).

- Número de salidas posibles: 1, asociadas al Control global y tiene tres posibles valores: controlado, Parcialmente controlado y no controlado).

- Número de reglas generadas: 244.

Con este sistema se tiene una referencia sobre qué tan controlada esta la crisis en función del número de eventos anormales, su duración y prevalencia en el EEG y de los factores de riesgo calculados basados en la información clínica.

Como siguiente proceso, se analizaron los resultados previos para definir la congruencia utilizando el siguiente sistema de inferencia difusa:

- Número de entradas: 2, resultado de la evaluación global de los datos clínicos y paraclínicos (controlado, parcialmente controlado y no controlado).

- Número de salidas: 1, Congruencia (congruente y no congruente), la cual es defusificada utilizando el método del centroide para obtener un valor certero.

- Número de reglas: 9.

III. Implementación y validación del software

Para el sistema clasificador que identifica y etiqueta los elementos básicos de la señal EEG, se tomó una muestra de 192 segmentos de un segundo de manera aleatoria a partir de 8 registros de pacientes diagnosticados con epilepsia parcial. Se calcularon los elementos de la tabla de confusión, donde se tienen los verdaderos positivos (VP), falsos positivos (FP), falsos negativos (FN) y verdaderos negativos (VN) y con ellos se obtuvieron los parámetros de desempeño, tales como la eficiencia, sensibilidad, especificidad, valor predictivo positivo(VPP), valor predictivo negativo(VPN) y valor Global(VG) definidos como:

$$Eficiencia = \frac{VP}{N}$$

$$Sensibilidad = \frac{VP}{VP + FP}$$

$$Especificidad = \frac{VN}{VN + FP}$$

$$VPP = \frac{VP}{VP + FP}$$

$$VPN = \frac{VN}{VN + FN}$$

$$VG = \frac{VP + VN}{VP + VN + FP + FN}$$

Finalmente, toda esta información sirve para validar con la metodología ROC al clasificador multi-clase, partiendo de la relación de lo esperado con lo obtenido [4].

El sistema general que evalúa la congruencia de los datos se implementó en Matlab versión R2012b y para probar su eficiencia se utilizó una muestra de 22 pacientes con diagnóstico de epilepsia parcial y 3 pacientes con hiperactividad y se contrastó con lo reportado por el médico para evaluar la eficiencia utilizando la tabla de confusión.

IV. Resultados

Con las pruebas realizadas al clasificador se observó un eficiencia del 96%, que se encuentra dentro de los rangos reportados por otros autores [4], la diferencia principal con respecto a ellos es que esta técnica tiene necesidades mínimas de tiempo y computacionales, debido a que los parámetros se restringen a voltaje y frecuencias de la señal, secuencias de características básicas y se modulan de acuerdo a los demás parámetros considerados (Tabla 3). En la Figura 4, se muestra un ejemplo de la información analizada por el software IdgEEG ©, es un tramo de señal de 8 segundos donde se calcula el parámetro frecuencia dominante en segmentos de un segundo.

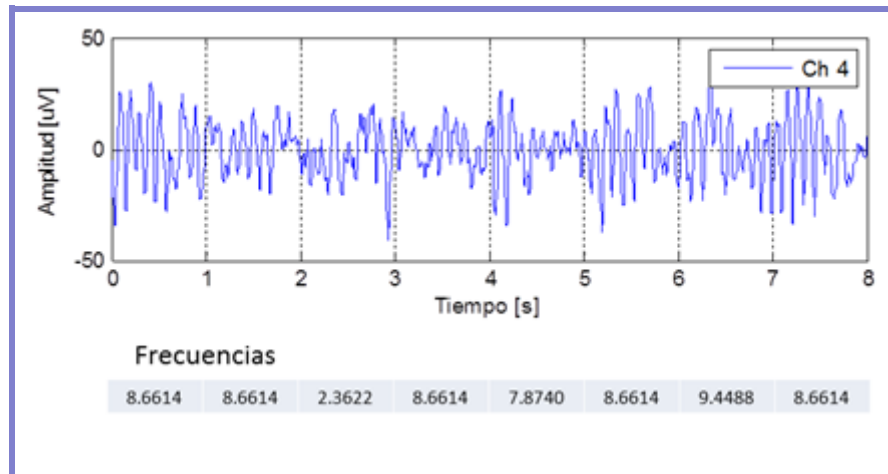


Figura 4. Segmento de señal con los valores de frecuencias dominantes por cada segundo, que se clasifican con IddEEG ©

Tabla 3. Resultados de la evaluación del clasificador IdgEEG ©

Eficiencia	0.968	
Sensibilidad	0.968	
Especificidad	0.963	
Valor Predictivo Positivo (VPP)	0.968	
Valor Predictivo Negativo (VPN)	0.963	
Valor Global (VG)	0.966	

En la tabla 4, se observan los resultados de la etapa de ponderación de datos clínicos para un paciente diagnosticado con epilepsia parcial. En la Tabla 5 se presentan los parámetros paraclínicos de otro paciente, utilizando la prevalencia, duración y frecuencia dominante para cada zona funcional del cerebro [1].

Tabla 4. Resultado del análisis de los datos clínicos con respecto a sus parámetros para la evaluación de congruencia entre tipos de información.

PACIENTE	Historia Familiar	Embarazo	Condición perinatal	Desarrollo neuronal	Historia personal	Tipo de evento reportado
	Rango {0-28}	Rango {0-11}	Rango {0-7}	Rango {0-15}	Rango {0-4}	Rango {0-8}
5	0.00	0.09	0.60	0.00	0.00	1.00

PACIENTE	FRONTAL			CENTRAL			PARIETAL			TEMPORAL			OCCIPITAL		
	Frec.	%	sec	Frec.	%	Sec	Frec.	%	Sec	Frec.	%	sec	Frec.	%	Sec
9	2.0	37	1.8	9.0	34	1.6	7.4	51	2.6	7.1	42	2.3	7.7	37	2.3

Tabla 5. Resultado del análisis de los datos paraclínicos con respecto a sus parámetros para la evaluación de congruencia entre tipos de información

Tanto el paciente 5 como el paciente 9 tienen un expediente con información congruente y el programa preparó la base de conocimiento y la dejó disponible para los procesos siguientes.

Para el sistema de evaluación de congruencia se obtuvo 1 de los 25 expedientes con clara evidencia de incongruencia entre los datos, lo que indicaría que se deben de revisar esos datos y el diagnóstico. Ese caso en específico fue de un paciente con diagnóstico de epilepsia aunado a otra enfermedad, lo que conllevó que se tuviera una tasa del 95% de expedientes útiles.

V. Conclusiones

Considerando que el 75% de casos clínicos de pacientes neurológicos son diagnosticados con epilepsia dentro del sector salud, y que se requiere alta especialización para el manejo y diagnóstico de estos pacientes, resulta importante auxiliar a los médicos en el manejo de la gran cantidad de información. El sistema experto propuesto auxilia al médico a sistematizar el manejo de la información clínica así como le proporciona la información que se requiere para hacer el diagnóstico, verificando que ambos tipos de información sean congruentes. La eficiencia de cada uno de los elementos condiciona a la eficiencia total del sistema por lo cual se maneja una eficiencia del 96% asociada a la automatización del clasificador, la última etapa donde se evalúa la información es generada directamente utilizando los rangos e información extraída de las guías estándar que rigen el quehacer médico, por lo cual la evaluación fue realizada por el médico especialista a través del seguimiento de la implementación y se cubrieron los parámetros más relevantes para el diagnóstico de epilepsia.

Este software genera la extracción automática y evaluación de parámetros para etapas posteriores de análisis que permitirán aportar herramientas clínicas tanto a los estudiantes de neurología, así como auxiliares de detección temprana en instancias de primer contacto que permitan que un médico general redirija al paciente a las instancias correspondientes, ya que la epilepsia detectada a tiempo puede llegar a ser controlada de manera eficiente a través de tratamientos adecuados.

VI. Referencias

- [1] Singh, S., 2015. Development of an intelligent system to assist the diagnosis of epileptic events in children. Tesis de Maestría para obtener el grado de Maestría en Tecnología Avanzada, UPIITA, IPN.
- [2] Nuwer, M. R., Comi, G., Emerson, R., Fuglsang-Frederiksen, A., Guerit, J. M., Hinrichs, H., ... & Rappelsberger, P. (1999). IFCN standards for digital recording of clinical EEG. The International Federation of Clinical Neurophysiology. *Electroencephalography and clinical neurophysiology. Supplement*, 52, 11.
- [3] Manford, M., Hart, Y. M., Sander, J. W., & Shorvon, S. D. (1992). The National General Practice Study of Epilepsy: the syndromic classification of the International League Against Epilepsy applied to epilepsy in a general population. *Archives of Neurology*, 49(8), 801.
- [4] Ramirez C., 2016. Estudio de la señal EEG para la identificación de patrones precursores de crisis epilépticas, Tesis para obtener el grado de Maestría en Tecnología Avanzada, UPIITA.
- [5] M. O. B. O. O.W. Samuel, A web based decision support system driven by fuzzy logic for the diagnosis of typhoid fever, *Expert Systems with Applications*, pp. 1-8, 2013.
- [6] Jang S. R. C.-T. S. a. E. M. Jang., 1997, *Neuro-fuzzy and soft computing; a computational approach to learning and machine intelligence.*, Prentice Hall.
- [7] A. C. N. S. (ACNS), American Clinical Neurophysiology Society, 2001-2015. [Online]. Available: <http://www.acns.org/practice/guidelines>. [Last login: 2014]