

## DETEKSI KEJANG *EPILEPSY* DENGAN MENGUNAKAN PEMILIHAN FITUR *INFORMATIOAN GAIN* DAN PEMBELAJARAN *ENSEMBLE RANDOM FOREST*

Mulaab<sup>1)</sup>

<sup>1</sup>Prodi Teknik Informatika, Fakultas Teknik, Universitas Trunojoyo

Jl. Raya Telang, PO BOX 2 Kamal, Bangkalan

E-mail : [mulaab@trunojoyo.ac.id](mailto:mulaab@trunojoyo.ac.id)

### ABSTRAK

Epilepsi adalah kondisi neurologis kronis yang ditandai dengan kejang yang tidak diketahui penyebabnya karena pelepasan neuron yang abnormal (*Seizures inturn*). *Epilepsi* pada manusia ditandai dengan kejang berulang yang mengakibatkan konsekuensi *neurologis, kognitif, psikologis* dan sosial. Kejang karena *epilepsi* akan berbeda dengan kejang yang disebabkan oleh kelainan pelepasan *neuronal* dari peristiwa *non epilepsi*, seperti kejang *psikogenik*. Deteksi kejang *epilepsi* dapat diintegrasikan ke dalam perangkat *implan* yang dimaksudkan untuk mendeteksi timbulnya kejang dan memberikan pengobatan pada pasien sehingga dapat menekan perkembangan kejang pada pasien. *Electroencephalogram* (EEG) merekam fluktuasi tegangan dari beberapa elektroda yang ditempatkan pada kulit kepala subjek selama periode waktu tertentu untuk mendiagnosa kelainan sindrom afektif dan organik pada manusia. Kemampuan mendeteksi *epilepsy* berdasarkan ekstraksi sinyal EEG untuk proses diagnosa bergantung kepada proses yang dilakukan yaitu ekstraksi fitur dari sinyal, dan pemilihan fitur penting dari sinyal serta metode pengklasifikasi dari sinyal tersebut. Pada penelitian ini bertujuan melakukan deteksi epilepsi berdasarkan proses seleksi fitur *information gain* dan pembelajaran *ensemble random forest*. Metode yang digunakan pada seleksi fitur menggunakan *information gain* dan pengklasifikasi pembelajaran *ensemble random forest*. Berdasarkan hasil percobaan didapatkan akurasi identifikasi kejang *epilepsy* dengan pemilihan fitur tertentu pada beberapa subjek pasien dengan menggunakan pengklasifikasi pembelajaran *ensemble random forest* telah dihasilkan akurasi diatas 0.99

**Kata kunci :** *Epilepsy Seizure Detection, EEG, Information Gain, Random Forest, Ensemble Learning*

### ABSTRACT

*Epilepsy is a chronic neurological condition characterized by seizures of unknown cause due to abnormal neuronal discharge (Seizures inturn). Epilepsy in humans is characterized by recurrent seizures that result in neurological, cognitive, psychological and social consequences. Seizures due to epilepsy will differ from seizures caused by abnormal neuronal discharge from non-epileptic events, such as psychogenic seizures. Epileptic seizure detection can be integrated into implantable devices that are intended to detect the onset of seizures and provide treatment to the patient so as to suppress the development of seizures in the patient. and organic in humans. The ability to detect epilepsy based on the extraction of EEG signals for the diagnostic process depends on the process carried out, namely the extraction of features from the signal, and the selection of important features of the signal and the method of classifying the signal. This study aims to detect epilepsy based on the information gain feature selection process and random forest ensemble learning. The method used in feature selection uses information gain and classifier of random forest ensemble learning. Based on the experimental results, it was found that the accuracy of the identification of epileptic seizures by selecting certain features in some patient subjects using a classifier of random forest ensemble learning has resulted in an accuracy above 0.99*

**Keywords:** *Epilepsy Seizure Detection, EEG, Information Gain, Random Forest, Ensemble Learning*

## PENDAHULUAN

Epilepsi adalah kondisi neurologis kronis yang ditandai dengan kejang yang tidak diketahui penyebabnya karena pelepasan neuron yang abnormal (*Seizures inturn*)[1] Secara nasional, epilepsi mempengaruhi di mana saja dari 0,5% hingga 3% dari populasi. Insiden tertinggi terjadi pada anak-anak dan orang tua

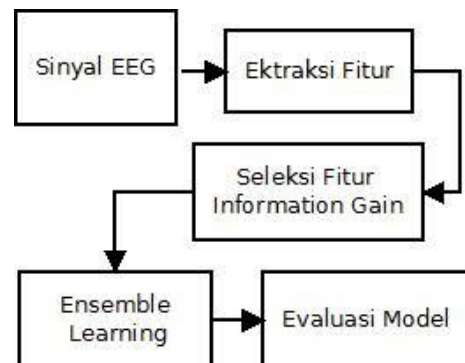
Kejang merupakan perubahan *paroksismal* dari fungsi *neurologis* yang disebabkan karena *hipersinkronik* yang berlebihan dari pelepasan *neuron* di otak, yang akan menyebabkan [2] perubahan perilaku, gerakan berhenti sesaat, perubahan perasaan, kehilangan kesadaran, kejang otot dan koma [3]. Kejang karena epilepsi akan berbeda dengan kejang yang disebabkan oleh kelainan pelepasan *neuronal* dari peristiwa *nonepilepsi*, seperti kejang *psikogenik*. Pada epilepsi disfungsi dari otak yang ditandai dengan kejang berulang-ulang. Epilepsi pada manusia ditandai dengan kejang berulang yang mengakibatkan konsekuensi *neurologis*, *kognitif*, *psikologis* dan sosial. Kejang karena epilepsi akan berbeda dengan kejang yang disebabkan oleh kelainan pelepasan *neuronal* dari peristiwa *nonepilepsi*, seperti kejang *psikogenik*. Deteksi kejang epilepsi dapat diintegrasikan ke dalam perangkat *implan* yang dimaksudkan untuk mendeteksi timbulnya kejang dan memberikan pengobatan pada pasien sehingga dapat menekan perkembangan kejang pada pasien.

*Electroencephalogram* (EEG) adalah salah satu teknologi *non-invasif* tertua untuk mengukur aktivitas saraf otak manusia. EEG merekam fluktuasi tegangan dari beberapa elektroda yang ditempatkan pada kulit kepala subjek selama periode waktu tertentu yang dibuat oleh psikiater Jerman Hans Berger pada tahun 192 [4]. Penelitian berkaitan dengan EEG telah banyak dilakukan dalam untuk mendiagnosa kelainan sindrom *afektif* dan organik pada manusia[5-7]. Deteksi sinyal EEG juga

merupakan salah satu cara terpenting untuk mendiagnosis epilepsi. Upaya untuk menerapkan data EEG kulit kepala dalam deteksi dan prediksi kejang telah membawa hasil yang menarik selama praktik klinis. Beberapa penelitian terkait dengan diagnosa *epilepsy* telah banyak dilakukan yaitu klasifikasi kejang epilepsi yang optimal di EEG menggunakan pengklasifikasi tunggal [8-10]. Pemilihan fitur sinyal EEG pada *epilepsy* telah dilakukan tujuan untuk memperbaiki akurasi identifikasi dengan menggunakan Algoritma Genetika (GA) [11], dan koefisiern *korelasi* Matthews *correlation* dan *Kappa*[12]. Banyak pekerjaan yang telah dilakukan untuk meningkatkan akurasi pada identifikasi kejang epilepsi.[13-15] Pada penelitian ini bertujuan mengimplementasikan metode pembelajaran *ensemble* (*ensemble learning*) dan seleksi fitur *information gain* untuk memperbaiki performansi identifikasi atau deteksi kejang epilepsi.

## METODE

Untuk mengimplementasikan deteksi kejang epilepsi, kami merancang diagram alir algoritma, seperti yang ditunjukkan pada Gambar 1.



Gambar 1. Identifikasi Kejang Epilepsi

Pertama, data sinyal EEG diinputkan. Data signal yang digunakan adalah berasal dari *CHB-MIT Scalp EEG Database* (<https://physionet.org/content/chbmit/1.0.0/>)[16] yang dikumpulkan di Rumah Sakit Anak Boston, terdiri dari rekaman

EEG dari subjek pediatrik dengan kejang yang tidak dikendalikan. Subyek dipantau dalam beberapa hari setelah tidak diberi obat anti-kejang untuk mengamati karakteristik kejang Sinyal EEG diambil sampelnya pada 256 sampel per detik dengan resolusi 16-bit dan direkam secara bersamaan di 21-28 saluran. Pada penelitian ini menggunakan dengan 23 *channel*. Peristiwa kejang diidentifikasi oleh para ahli klinis. Pada penelitian ini memilih tiga subyek untuk diamati juga mempertimbangkan distribusi jenis kelamin dan usia subjek. Demografi subjek yang dipilih ditunjukkan pada tabel 1. Data yang digunakan adalah pada anak perempuan umur 11 (chb01), anak laki umur 11 (chb02) dan anak perempuan umur 14 (chb03)

Tabel.1 Subyek sinyal EEG

Subyek	chb01	chb02	chb03
Jenis Kel.	P	L	P
Umur	11	11	14

Kemudian dari data yang diinputkan dilakukan ekstraksi fitur berdasarkan *Time Domain Features* (TDF) dan *Frequency Domain Features* yaitu *Power Spectral Intensity* dan *Relative Intensity Ratio*, amplitudo *Root Mean Square*, variansi, *kurtosis*, *skewness*, maksimum amplitudo, minimum amplitudo, banyaknya *local extrema* atau *peaks*, banyaknya *zero crossings*, *Higuchi Fractal Dimension*, *Petrosian Fractal Dimension*, *Hurst exponent*, *Spectral Entropy*, *total power*, *median frequency*, *peak frequency*, *Hjorth Parameters (Hjorth mobility dan complexity)*. Dari 22 fitur yang akan telah diekstrak selanjutnya dilakukan pemilihan fitur untuk mendapaftkan fitur dominan sebagai karakteristik dari signal EEG yang ada.

Pemilihan fitur menggunakan *Informatin Gain* yaitu metode pemilihan fitur berbasis filter yang hasilnya akan diperingkat pada berdasarkan nilai *entropy* yang telah dihiutung pada masing masing fitur. Pada tahapan berikutnya dilakukan identifikasi dengan

pengklasifikasi dengan teknik pembelajaran *ensemble* dengan model *majority voting*. Implementaasi pembelajaran *ensemble* menggunakan *Random Forest*.

Identifikasi kejang *epilepsy* pada sinyal dengan sumber data EEG mengandalkan karakterisasi sinyal EEG. Untuk itu, perlu dilakukan ekstraksi fitur ke dalam fitur tertentu. Proses ini dikenal sebagai ekstraksi fitur. Mengekstraksi fitur EEG dapat dilakukan seperti pada analisa deret waktu : kerapatan *spektral* daya dari pemrosesan sinyal, dimensi fraktal dari geometri komputasi, entropi dari teori informasi, dan sebagainya. Beberapa fitur EEG yang dapat diektrak diantaranya adalah *Power Spectral Intensity* dan *Relative Intensity Ratio* [17], amplitudo *Root Mean Square*, variansi, *kurtosis*, *skewness*, maksimum amplitudo, minimum amplitudo, banyaknya *local extrema* atau *peaks*, banyaknya *zero crossings*, *Higuchi Fractal Dimension* [18], *Petrosian Fractal Dimension* [19], *Hurst exponent* [20], *Spectral Entropy* [21], *total power*, *median frequency*, *peak frequency*, *Hjorth Parameters (Hjorth mobility dan complexity)*[22]. Seleksi fitur dapat membantu mengurangi *noise* yang disebabkan oleh fitur-fitur yang tidak relevan. Metode pemilihan fitur juga dapat dikategorikan berdasarkan apa yang mereka evaluasi dan rangking: fitur individual atau *subset* fitur. *Information Gain* adalah salah satu pendekatan metode seleksi fitur dengan mengevaluasi fitur berdasarkan ukuran Entropi dari fitur [23]. Ukuran ini digunakan untuk menentukan peringkat variable atau fitur. *Information Gain* mengevaluasi semua fitur satu per satu dan memberi peringkat. Pemilihan subset fitur dicapai dengan memilih fitur peringkat N tertinggi atau semua fitur dengan nilai di atas t, di mana N dan t adalah batas ambang yang ditentukan pengguna. Diberikan himpunan kelas  $C = \{c_1, \dots, c_k\}$ , mendapatkan informasi dari fitur f,  $IG(f)$  pada persamaan(1), dengan

$$IG(f) = H(C) - H(C|f) \quad (1)$$

$$H(C) = -\sum_{i=1}^k P(c_i) \log P(c_i) \quad (2)$$

$$H(C|f) = -P(f) \sum_{i=1}^k P(c_i|f) \log P(c_i|f) \quad (3)$$

Information gain dari setiap fitur diperoleh dari entropy kelas (persamaan 2) dikurangi dengan entropy setiap kelas terhadap semua fitur (persamaan 3). Perhitungan dilakukan untuk setiap fitur di semua kelas dan kemudian fitur-fitur tersebut diperingkat berdasarkan nilai IG masing masing; semakin tinggi nilainya, semakin informatif fitur tersebut. Setelah proses seleksi fitur dilakukan tahapan berikutnya adalah melakukan pembelajaran jamak (ensemble learning).

Pembelajaran ensemble adalah bentuk sistem pembelajaran hibrida di mana beberapa analitik digabungkan secara cerdas dengan tujuan untuk mendapatkan hasil yang lebih baik (lebih akurat, lebih kuat) daripada yang dapat diberikan oleh satu analitik[24]. Majority Voting merupakan salah satu teknik dari pembelajaran ensemble untuk melakukan klasifikasi dari data baru tidak berlabel berdasarkan dari kelas yang memperoleh jumlah suara terbanyak (Plurality Vote) dari classifier yang digunakan. Pembelajaran Ensemble dengan pendekatan Majority Voting dari suatu kelas data baru x (persamaan 4) dapat dinyatakan dengan

$$\text{class}(x) = \arg \max_{c_i \in \text{dom}(y)} \left( \sum_k g(y_k(x), c_i) \right) \quad (4)$$

di mana  $y_k(x)$  adalah klasifikasi dari pengklasifikasi ke-k dan  $g(y, c)$  adalah fungsi indikator yang didefinisikan sebagai:

$$g(y, c) = \begin{cases} 1 & y = c \\ 0 & y \neq c \end{cases} \quad (5)$$

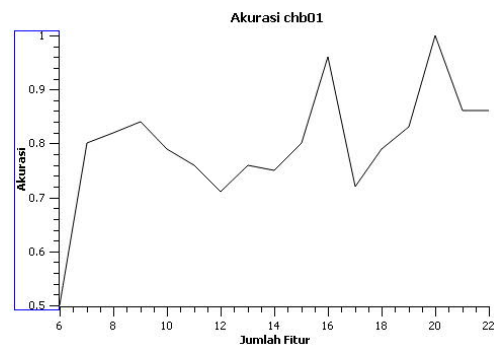
Perhatikan bahwa dalam kasus pengklasifikasi probabilistik, klasifikasi  $y_k(x)$  biasanya diperoleh dengan

$$y_k(x) = \arg \max_{c_i \in \text{dom}(y)} \hat{P}_{M_k}(y = c_i | x) \quad (6)$$

dimana  $M_k$  menunjukkan pengklasifikasi k dan  $\hat{P}_{M_k}(y = c | x)$  menunjukkan probabilitas y sama dengan c yang dari data baru x.

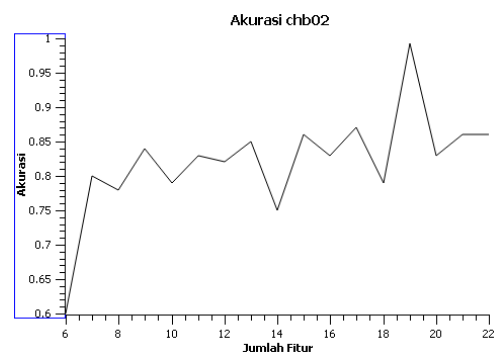
## HASIL DAN PEMBAHASAN

Percobaan yang telah dilakukan terhadap 3 subjek pasien yaitu chb01, dan chb02, chb03 dengan melakukan ekstraksi fitur dan pemilihan fitur serta melakukan pengklasifikasian dengan pembelajaran ensemble random forest. Pada pasien chb01 akurasi didapatkan dengan beberapa skenario pemilihan fitur didapat hasil akurasi maksimal dengan pemilihan fitur sebanyak 20 fitur, dengan membuang 2 fitur yang tidak dominan seperti gambar 2.



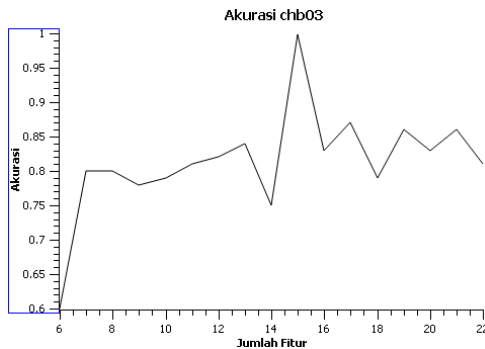
Gambar 2. Akurasi dengan beberapa skenario pemilihan fitur pada pasien chb01

Pada pasien chb02 akurasi pemilihan fitur didapat hasil akurasi maksimal dengan pemilihan fitur sebanyak 20 fitur, dengan membuang 3 fitur yang tidak dominan seperti gambar 3.



Gambar 3. Akurasi dengan beberapa skenario pemilihan fitur pada Chb02

Pada pasien chb03 akurasi pemilihan fitur didapat hasil akurasi maksimal dengan pemilihan fitur sebanyak 20 fitur, dengan membuang 7 fitur yang tidak dominan seperti gambar 4.



Gambar 4. Akurasi dengan beberapa skenario pemilihan fitur pada Chb03

Beberapa variasi jumlah fitur yang dipilih dan menggunakan kombinasi dari klasifikasi dasar dalam menerapkan pembelajaran ensemble pada 3 pasien epilepsi chb01, chb02, chb03 didapatkan hasil seperti pada tabel 2.

Tabel 2. Hasil identifikasi kejang *epilepsy*

Pasien	Fitur maksimal	Akurasi
Chb01	20	0.9994
Chb02	19	0.9993
Chb03	15	0.9990

Pada table 2 subyek **chb01** mempunyai fitur kombinasi adalah *Power Spectral Intensity dan Relative Intensity Ratio*, amplitudo *Root Mean Square*, variansi, *kurtosis*, *skewness*, maksimum amplitudo, minimum amplitudo, banyaknya *local extrema* atau *peaks*, *Higuchi Fractal Dimension*, *Hurst exponent*, *Spectral Entropy*, *total power*, *median frequency*, *peak frequency*, *Hjorth Parameters* (*Hjorth mobility dan complexity*).

Subyek **Chb02** mempunyai fitur kombinasi adalah *Power Spectral Intensity dan Relative Intensity Ratio*, amplitudo *Root Mean Square*, variansi, *kurtosis*, *skewness*, maksimum amplitudo, minimum amplitudo,

banyaknya *local extrema* atau *peaks*, banyaknya *zero crossings*, *Dimension*, *Hurst exponent*, *Spectral Entropy*, *total power*, *median frequency*, *peak frequency*, *Hjorth Parameters* (*Hjorth mobility dan complexity*).

Subyek **Chb03** mempunyai fitur kombinasi adalah *Power Spectral Intensity dan Relative Intensity Ratio*, amplitudo *Root Mean Square*, variansi, *kurtosis*, *skewness*, maksimum amplitudo, minimum amplitudo, banyaknya *local extrema* atau *peaks*, banyaknya *zero crossings*, *Higuchi Fractal Dimension*, *Petrosian Fractal Dimension*, *Hurst exponent*, *Spectral Entropy*, *total power*, *median frequency*, *peak frequency*

Dari hasil yang didapatkan bahwa terdapat variasi fitur penting yang ada pada masing masing sinyal rekaman EEG pada pasien epilepsi yang akan menentukan akurasi identifikasi kejang epilepsi. Dari ketiga pasien terdapat tingkat derajat yang sama yaitu *Power Spectral Intensity dan Relative Intensity Ratio*, amplitudo *Root Mean Square*, variansi, *kurtosis*, *skewness*, maksimum amplitudo yang menjadi 6 fitur dominan pertama

## SIMPULAN

Berdasarkan hasil dan pembahasan bahwa variasi dari jumlah fitur sangat menentukan akurasi dari identifikasi kejang *epilepsy*. Terdapat beberapa fitur penting yang akan menentukan karakteristik kejang *epilepsy*. Pemilihan fitur yang tepat dan membuang fitur yang tidak dominan akan meningkatkan performansi dari identifikasi kejang epilepsy

## SARAN

Pada penelitian ini masih perlu dikembangkan kombinasi beberapa seleksi fitur (*ensemble seleksi fitur*) yang dapat digunakan untuk memilih fitur sehingga diharapkan nantinya memilih fitur yang benar benar dominan

menentukan karakteristik sinyal EEG kejang *epilepsy*.

#### DAFTAR PUSTAKA

- [1] P. Kandula and C. Harden, "Epilepsy," in *Encyclopedia of Neuroscience*, L. R. Squire, Ed., ed Oxford: Academic Press, 2009, pp. 1147-1149.
- [2] H. E. Scharfman, "Chapter 17 - Epilepsy," in *Neurobiology of Brain Disorders*, M. J. Zigmond, L. P. Rowland, and J. T. Coyle, Eds., ed San Diego: Academic Press, 2015, pp. 236-261.
- [3] A. Michael-Titus, P. Revest, and P. Shortland, "13 - EPILEPSY," in *The Nervous System (Second Edition)*, A. Michael-Titus, P. Revest, and P. Shortland, Eds., ed: Churchill Livingstone, 2010, pp. 237-250.
- [4] R. İnce, S. S. Adanır, and F. Sevmez, "The inventor of electroencephalography (EEG): Hans Berger (1873–1941)," *Child's Nervous System*, 2020/03/05 2020.
- [5] J. G. Small, "3 - EEG IN AFFECTIVE DISORDERS," in *EEG and Evoked Potentials in Psychiatry and Behavioral Neurology*, J. R. Hughes and W. P. Wilson, Eds., ed: Butterworth-Heinemann, 1983, pp. 41-54.
- [6] M. Lavanga, J. De Ridder, K. Kotulska, R. Moavero, P. Curatolo, B. Weschke, *et al.*, "Results of quantitative EEG analysis are associated with autism spectrum disorder and development abnormalities in infants with tuberous sclerosis complex," *Biomedical Signal Processing and Control*, vol. 68, p. 102658, 2021/07/01/ 2021.
- [7] E. Minato and K. A. Myers, "Age-related evolution of EEG in Dravet syndrome: Meta-analysis of 155 patients," *Seizure*, vol. 91, pp. 108-111, 2021/10/01/ 2021.
- [8] H. Ocak, "Optimal classification of epileptic seizures in EEG using wavelet analysis and genetic algorithm," *Signal Processing*, vol. 88, pp. 1858-1867, 2008/07/01/ 2008.
- [9] E. D. Übeyli, "Combined neural network model employing wavelet coefficients for EEG signals classification," *Digital Signal Processing*, vol. 19, pp. 297-308, 2009/03/01/ 2009.
- [10] E. D. Übeyli, "Combining recurrent neural networks with eigenvector methods for classification of ECG beats," *Digital Signal Processing*, vol. 19, pp. 320-329, 2009/03/01/ 2009.
- [11] R. Dhiman, J. S. Saini, and Priyanka, "Genetic algorithms tuned expert model for detection of epileptic seizures from EEG signatures," *Applied Soft Computing*, vol. 19, pp. 8-17, 2014/06/01/ 2014.
- [12] S. Yang, B. Li, Y. Zhang, M. Duan, S. Liu, Y. Zhang, *et al.*, "Selection of features for patient-independent detection of seizure events using scalp EEG signals," *Computers in Biology and Medicine*, vol. 119, p. 103671, 2020/04/01/ 2020.
- [13] B. Direito, J. Duarte, C. Teixeira, B. Schelter, M. L. Van Quyen, A. Schulze-Bonhage, *et al.*, "Feature selection in high dimensional EEG features spaces for epileptic seizure prediction," *IFAC Proceedings Volumes*, vol. 44, pp. 6206-6211, 2011/01/01/ 2011.
- [14] V. Harpale and V. Bairagi, "An adaptive method for feature selection and extraction for classification of epileptic EEG signal in significant states," *Journal of King Saud University - Computer and Information Sciences*, 2018/05/24/ 2018.
- [15] E. Pippa, E. I. Zacharaki, I. Mporas, V. Tsirka, M. P. Richardson, M. Koutroumanidis, *et al.*, "Improving classification of epileptic and non-epileptic EEG

- events by feature selection," *Neurocomputing*, vol. 171, pp. 576-585, 2016/01/01/ 2016.
- [16] A. H. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," Massachusetts Institute of Technology, 2009.
- [17] R. Q. Quiroga, S. Blanco, O. Rosso, H. Garcia, and A. Rabinowicz, "Searching for hidden information with Gabor Transform in generalized tonic-clonic seizures," *Electroencephalography and clinical Neurophysiology*, vol. 103, pp. 434-439, 1997.
- [18] T. Higuchi, "Approach to an irregular time series on the basis of the fractal theory," *Physica D: Nonlinear Phenomena*, vol. 31, pp. 277-283, 1988.
- [19] A. Petrosian, "Kolmogorov complexity of finite sequences and recognition of different preictal EEG patterns," in *Proceedings eighth IEEE symposium on computer-based medical systems*, 1995, pp. 212-217.
- [20] T. Balli and R. Palaniappan, "A combined linear & nonlinear approach for classification of epileptic EEG signals," in *2009 4th International IEEE/EMBS Conference on Neural Engineering*, 2009, pp. 714-717.
- [21] T. Inouye, K. Shinosaki, H. Sakamoto, S. Toi, S. Ukai, A. Iyama, *et al.*, "Quantification of EEG irregularity by use of the entropy of the power spectrum," *Electroencephalography and clinical neurophysiology*, vol. 79, pp. 204-210, 1991.
- [22] B. Hjorth, "EEG analysis based on time domain properties," *Electroencephalography and clinical neurophysiology*, vol. 29, pp. 306-310, 1970.
- [23] I. Koprinska, "Feature selection for brain-computer interfaces," in *Pacific-Asia conference on knowledge discovery and data mining*, 2009, pp. 106-117.
- [24] L. Rokach, *Pattern classification using ensemble methods* vol. 75: World Scientific, 2010.