PAPER • OPEN ACCESS

Epileptic Seizure Detection in EEGs by Using Random Tree Forest, Naïve Bayes and KNN Classification

To cite this article: Fauzia P. Lestari *et al* 2020 *J. Phys.: Conf. Ser.* **1505** 012055

Epileptic Seizure Detection in EEGs by Using Random Tree Forest, Naïve Bayes and KNN Classification

Fauzia P. Lestari[1], Mohammad Haekal[1], Rizki Edmi Edison[2], Fikry Ravi Fauzy[2], Siti Nurul Khotimah[1] and Freddy Haryanto[1]

[1]Faculty of Mathematics and Natural Sciences, Intitut Teknologi Bandung, Jl. Ganesha No. 10, Bandung, West Java, Indonesia, 40132

[2]Universitas Muhammadiyah Prof. DR. HAMKA, Jl. Limau II, RT.3/RW.3, Kramat Pela, Kec. Kby. Baru, Kota Jakarta Selatan, DKI Jakarta 12130

fauzia@fi.itb.ac.id

Abstract. Epilepsy is a disease that attacks the nerves. To detect epilepsy, it is necessary to analyze the results of an EEG test. In this study, we compared the naive bayes, random tree forest and K-nearest neighbor (KNN) classification algorithms to detect epilepsy. The raw EEG data were pre-processed before doing feature extraction. Then, we have done the training in three algorithms: KNN Classification, naïve bayes classification and random tree forest. The last step was validation of the trained machine learning. Comparing those three classifiers, we calculated accuracy, sensitivity, specificity, and precision. The best trained classifier is KNN classifier (accuracy: 92.7%), rather than random tree forest (accuracy: 86.6%) and naïve bayes classifier (accuracy: 55.6%). Seen from precision performance, KNN Classification also gives the best precision (82.5%) rather than Naïve Bayes classification (25.3%) and random tree forest (68.2%). But, for the sensitivity, Naïve Bayes classification is the best with 80.3% sensitivity, compare to KNN 73.2% and random tree forest (42.2%). For specificity, KNN classification gives 96.7% specificity, then random tree forest 95.9% and Naïve bayes 50.4%. The training time of naïve bayes was 0.166030 sec, while training time of random tree forest was 2.4094sec and KNN was the slower in training that was 4.789 sec. Therefore, KNN Classification gives better performance than naïve bayes and random tree forest classification.

1. Introdunction

Epilepsy is a condition that can cause a person to experience seizures repeatedly. Seizure is often referred as epilepsy. The symptoms of seizures are vary, many patients have more than one type of seizure, and can experience symptoms of other neurological problems. Epilepsy is a disease that attacks the nerves of the brain. Epilepsy causes sufferers to experience recurrent seizures. Epilepsy becomes dangerous if the sufferer experiences sudden seizures in unprepared conditions, for example when driving, cooking, etc. Until now, there is no truly effective treatment to cure epilepsy. Patients are only given drugs to reduce the frequency of seizures, but do not completely cure the cause of epilepsy.

Electroencephalogram (EEG) is one of the tests performed to measure the electrical activity of the brain to detect abnormalities from the brain. This action uses a special sensor that is an electrode mounted on the head and connected through a cable to the computer. EEG plays an important role to

Journal of Physics: Conference Series **1505** (2020) 012055 doi:10.1088/1742-6596/1505/1/012055

detect epilepsy. To diagnose epilepsy, a doctor needs to check the results of a very long and complicated EEG test. The results of this reading are relatively dependent on the examining physician. An inexperienced doctor may give different results from the results of an expert doctor's examination. This method also takes a lot of time. To reduce this error, we need to make an algorithm that can automatically detect epilepsy accurately and quickly.

Machine learning is an application branch of Artificial Intelligence that focuses on developing a system that is capable of learning "on its own" without having to be repeatedly programmed by humans. Machine learning applications require data as learning material (training) before issuing outputs.

One of the machine learning algorithms that has been used to classify epilepsy is Suport Vector Machine (SVM). In this study, we used the KNN classification algorithm, naive bayes and random tree forest to detect epilepsy.

2. Methods

The EEG data was from chb-mit database, downloaded from https://physionet.org/pn6/chbmit/. These EEG data collected at the Children's Hospital Boston. The data consists of EEG recordings from pediatric subjects with intractable seizures. Subjects were monitored for up to several days following withdrawal of anti-seizure medication in order to characterize their seizures and assess their candidacy for surgical intervention. There were 11 patients (4 males and 7 females, aged 1.5-22 years old) EEG recording data. The .edf files contain exactly one hour of digitized EEG signals. All signals were sampled at 256 samples per second with 16-bit resolution. The files contain 23 EEG signals. The International 10-20 system of EEG electrode positions and nomenclature was used for these recordings.

Those data were grouped into training data and validation data. The training data consist of 71 Seizure files and 339 Non seizure files, while validation data consist of 12 Seizure files and 10 non seizure files.

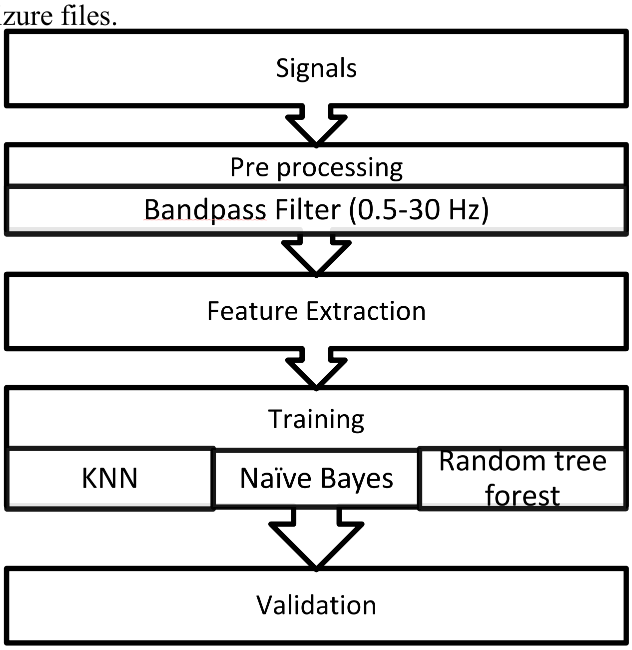

Figure 1. Overall flowchart of the methods

Table 1. Gender and age of patients who recorded in this study

Case	Gender	Age (years)
Chb01	F	11
Chb02	M	11
Chb03	F	14
Chb04	M	22
Chb05	F	7
Chb07	F	1.5
Chb08	F	14.5
Chb09	M	3.5
Chb09	F	10
Chb10	M	3
Chb23	F	6

Figure 1 shows the overall flowchart methods in this study. The raw EEG data were pre-processed before conducting feature extraction. Then, we have done the training in three algorithms: KNN Classification, naïve bayes classification and random tree forest. The last step was validation of the trained machine learning. To compare those three classifiers, we calculated accuracy, sensitivity, specificity, and precision defined in table 2

Table 2. Classifier performance

Accuracy	$ACC = \frac{TP + FN}{TP + TN + FP + FN}$
Sensitivity (True Positive Rate)	$TPR = \frac{TP}{TP + FN}$
Specificity (True Negative Rate)	$TNR = \frac{TN}{TN + FP}$
Precision (Positive Prediction Value)	$PPV = \frac{TP}{TP + FP}$
Negative Predictive value	$NPV = \frac{TN}{TN + FN}$

Features extracted in this study consist of temporal features and spectral features. Temporal features itself consists of mean, root mean square (RMS) and standard deviation (STD), while spectral features consist of spectral peaks and spectral power features. Spectral peaks were chosen 6 peaks location, both frequency and PSD of the peaks. Spectral power features were defined as the sum of the PSD in each brain frequency range, delta frequency (0.5-4 Hz), theta frequency (4-7 Hz), alpha frequency (7-15 Hz) and beta frequency (15-30 Hz) (see figure 2(b)). There were 437 features in total.

3. Results and Discussion

In the pre-processing step, band pass signals were applied for 0.5-30Hz only, because this range is the brain frequency signals. The peak of power spectral density is around delta frequency range for non-seizures signals. But, for seizure signals the PSD peaks is shifting to theta frequency.

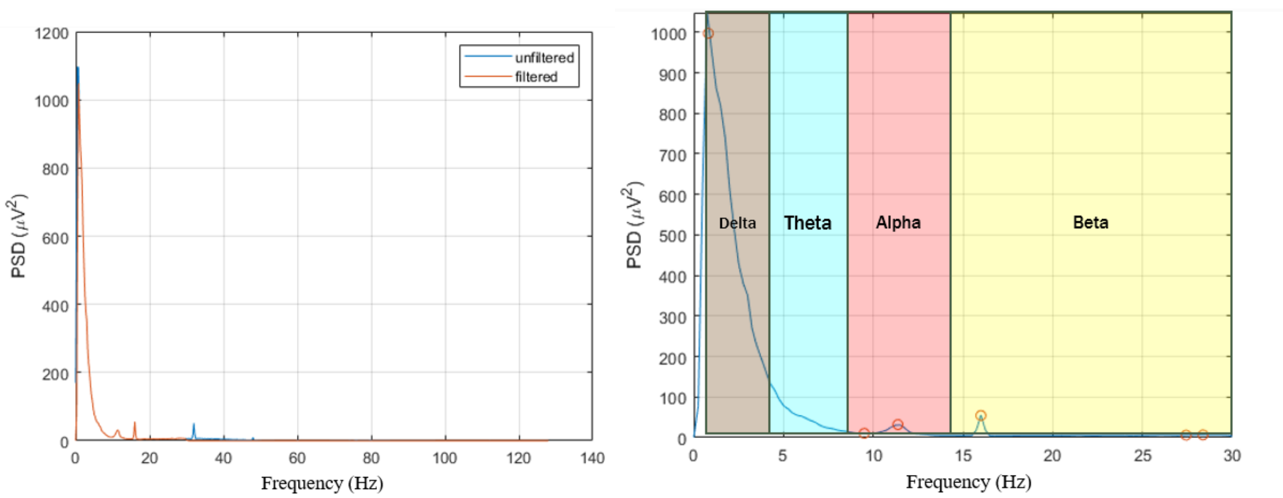

Figure 2. (a) Unfiltered and filtered power spectral density **(b)** Spectral features: 6 peaks locations showed in red bullet and brain frequency range (delta, theta, alpha, beta)

Table 3 shows the results of the training for KNN Classification, naïve bayes classification and random tree forest classification. Table 4 shows the classifier performance comparison. KNN classifier is the best classifier based on accuracy, precision, and specificity. KNN classification also give the highest AUC in true positive rates vs false positive rates. An excellent classifier has AUC near to the 1 which means it has good measure of separability. A poor classifier has AUC near to the 0 which means it has worst measure of separability. Therefore, KNN classification has good measure of separability.

But for sensitivity, naïve bayes is the best classifier. Because sensitivity and specificity are inversely proportional to each other. If we increase sensitivity, specificity decreases and vice versa. We should choose between specificity or sensitivity.

Table 3. Training results

Parameter	KNN Classification	Naïve bayes Classification	Random Tree forest
True Positive	52	57	30
True Negative	328	171	325
False Positive	11	14	14
False Negative	19	168	41

Table 4. Classifier performance

Parameter	KNN Classification	Naïve bayes Classification	Random Tree forest
Accuracy	92.7%	55.6%	86.6%
Precision	82.5%	25.3%	68.2%
Sensitivity	73.2%	80.3%	42.2%
Specificity	96.7%	50.4%	95.9%
AUC[a]	0.81	0.7788	0.72

[a]*Area Under Curve*

4. Conclusion

Machine learning gives good performance to classify seizure and non-seizure EEG data. Between KNN classification, Naïve bayes classification and random tree forest, the best trained classifier is KNN classifier (accuracy: 92.7%, precision: 82.5%, sensitivity 73.2% and specificity: 96.7%), rather than random tree forest (accuracy: 86.6%, precision 68.2%, sensitivity: 42.2%, and specificity: 96.7%) and naïve bayes classifier (accuracy: 55.6%, precision: 25.3%, sensitivity: 80.3%, and specificity: 50.4%). The training time of naïve bayes was 0.166030 sec, while the training time of random tree forest was 2.4094 sec and KNN was the slowest in training that was 4.789 sec.

Acknowledgement

This study was funded by the Penelitian Dasar Unggulan Perguruan Tinggi (PDUPT) 2019, Ministry of research Technology and Higher Education of the Republic of Indonesia2/E1/KP.PTNBH/2019 and 1170f/I1.C01/PL/2019. The authors give special thanks for helping to participate in SEACOMP 2019.

References

[1] Ali Shoeb, John Guttag. Application of Machine Learning to Epileptic Seizure Onset Detection. *27th International Conference on Machine Learning (ICML)*, June 21-24, 2010, Haifa, Israel.
[2] Ali Shoeb, Herman Edwards, Jack Connolly, Blaise Bourgeois, S. Ted Treves, John

Guttag. Patient-Specific Seizure Onset Detection. *Epilepsy and Behavior*. August 2004, 5(4): 483-498. [doi:10.1016/j.yebeh.2004.05.005]

[3] Acharya U. R., Oh S. L., Hagiwara Y., Tan J. H., Adeli H. (2017). Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. Comput Biol. Med. 100, 270–278. 10.1016/j.compbiomed.2017.09.017

[4] Ahmadi A., Behroozi M., Shalchyan V., Daliri M. R. (2018). Classification of Epileptic EEG Signals by Wavelet based CFC. Electr. Electro. Biomed. Eng. Comput. Sci. 10.1109/EBBT.2018.8391471

[5] Ahmedt-Aristizabal D., Fookes C., Nguyen K., Denman S., Sridharan S., Dionisio S. (2018). Deep facial analysis: a new phase I epilepsy evaluation using computer vision. Epilepsy Behav. 82, 17–24. 10.1016/j.yebeh.2018.02.010

[6] Chen G., Xie W., Bui T. D., Krzyżak A. (2017). Automatic epileptic seizure detection in EEG using nonsubsampled wavelet–fourier features. J. Med. Biol. Eng. 37, 123–131. 10.1007/s40846-016-0214-0

[7] Chen W., Lam Y. Y., Shen C. P., Sung H. Y., Chiu M. J., Lai F. (2013). Ultra-fast Epileptic seizure detection using EMD based on multichannel electroencephalogram, in IEEE International Conference on Bioinformatics and Bioengineering (Chania:), 1–4.

[8] Cun Y. L. (1995). Convolutional networks for images, speech, and time series, in Handbook of Brain Theory & Neural Networks, ed Arbid A., editor. (Cambridge, MA: MIT Press;).

[9] De Lucia M., Fritschy J., Dayan P., Holder D. S. (2008). A novel method for automated classification of epileptiform activity in the human electroencephalogram-based on independent component analysis. Med. Biol. Eng. Comput. 46, 263–272. 10.1007/s11517-007-0289-4

[10] Rehan Akbani, Stephen Kwek, and Nathalie Japkowicz. Applying Support- Vector Machines to Imbalanced Datasets. Springer, 2004.

[11] Al-Thaddeus Avestruz, Wesley Santa, Dave Carlson, Randy Jensen, Scott Stanslaski, Alan Helfenstine, and Tim Denison. A 5yw/channel spectral analysis ic for chronic bidirectional brain-machine interfaces. IEEE Journal Of Solid-State Circuits, 43(12):3006-3024, 2008

[12] E. Ben-Menachem. Vagus nerve stimulation for the treatment of epilepsy. Lancet Neurology, 1(8):477-482, 2002.

[13] P. Boon, K. Vonck, and P. Van Walleghem et al. Programmed and magnetinduced vagus nerve stimulation for refractory epilepsy. Journal Of Clinical Neurophysiology, 18(5):402-407, 2001.

[14] Martin J Brodie, Steven C Schachter, and Patrick Kwan. Fast Facts: Epilepsy. Health Press, third edition, 2005.

[15] M. J. Islam, Q. M. J. Wu, M. Ahmadi and M. A. Sid-Ahmed, "Investigating the Performance of Naive-Bayes Classifiers and K- Nearest Neighbor Classifiers," 2007 International Conference on Convergence Information Technology (ICCIT 2007), Gyeongju, 2007, pp. 1541-1546. doi: 10.1109/ICCIT.2007.148

[16] L. Jiang, Z. Cai, D. Wang and S. Jiang, "Survey of Improving K-Nearest-Neighbor for Classification," Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007), Haikou, 2007, pp. 679-683. doi: 10.1109/FSKD.2007.552

[17] Shweta Taneja, Charu Gupta, Kratika Goyal, Dharna Gureja, "An Enhanced K-Nearest Neighbor Algorithm Using Information Gain and Clustering", *Advanced Computing & Communication Technologies (ACCT) 2014 Fourth International Conference on*, pp. 325-329, 2014.

[18] Shweta Taneja, Charu Gupta, Sakshi Aggarwal, Veni Jindal, "MFZ-KNN — A modified fuzzy based K nearest neighbor algorithm", *Cognitive Computing and Information Processing (CCIP) 2015 International Conference on*, pp. 1-5, 2015.

[19] Zahraa Said Abdallah, Mohamed Medhat Gaber, "KB-CB-N classification: Towards unsupervised approach for supervised learning", *Computational Intelligence and Data Mining (CIDM) 2011 IEEE Symposium on*, pp. 283-290, 2011.

[20] Caio Soares, Lacey Montgomery, Kenneth Rouse, Juan E. Gilbert, "Automating Microarray Classification Using General Regression Neural Networks", *Machine Learning and Applications 2008. ICMLA '08. Seventh International Conference on*, pp. 508-513, 2008.