Conference Paper

Deep Learning Classification of Epileptic Magnetoencephalogram

Andreas Stylianou[1], Lefteris Koumakis[2], Maria Hadjinicolaou[3], Adam Adamopoulos[1,*] and Alkinoos Athanasiou[4]

[1] Medical Physics Laboratory, Department of Medicine, Democritus University of Thrace, Alexandroupolis, Greece.
[2] Foundation for Research and Technology—Hellas (FORTH), Institute of Computer Science, Bioinformatics Laboratory, Science and Technology Park of Crete, Greece.
[3] School of Science and Technology, Hellenic Open University, Patras, Greece.
[4] Lab of Medical Physics, School of Medicine, Aristotle University of Thessaloniki, Greece.

[*] Corresponding Author Email: adam@med.duth.gr

ABSTRACT

In this research, we study several statistical methods for feature extraction from Magnetoencephalography (MEG) Signals and classification of these signals into two classes: epileptic and healthy, based on the extracted features. We, then, apply automated feature extraction techniques by means of deep learning using several Artificial Neural Network (ANN) models. Our goal is to try various methods and models for MEG Signal classification and draw some conclusions about their functionality and effectiveness. We base our study on our theoretical knowledge of the neurology of epilepsy, previous studies of epileptic seizure imaging and recognition using MEG and Electroencephalogram (EEG) as well as the Signal Processing Theory and techniques. We apply several advanced classification methods with the use of ANN like Feed-Forward ANN, Convolutional Neural Networks (Convolutional NN), and Inception V3. The results of this study are very encouraging and can be a base for future research on the subject of epileptic seizure recognition, prediction, and prevention.

Keywords—*Magnetoencephalography, Epilepsy, Deep learning signal classification, Artificial Neural Networks, Feed-Forward Neural Networks, Convolutional Neural Networks, Inception V3.*

INTRODUCTION

Epilepsy is one of the most common neurological disorders with tens of millions of patients all over the world. Epileptic patients suffer from seizures which are the most important symptom. A patient is said to suffer from epilepsy after two or more unprovoked seizures separated by at least 24 hours.[1] Seizures vary in type, duration, and severity, but in any case, they are an unpleasant, even painful experience for the patient. Ranging from a short lack of consciousness to strong muscular spasms, seizures may also pose a significant danger to the patient. Injuries may be caused by falling to the ground or while handling dangerous objects or machinery. The abnormal firing of neurons (up to 500 times per second) may also damage the brain cells, especially during prolonged seizures or ones that appear in succession (status epilepticus).[2] A number of previous works[3] stress the need for implementing automated methods for the detection of epileptic activity as well as automated diagnosis and automatic prediction of epileptic seizures. Such methods should help reduce human errors by specialized personnel due to the fatigue after long hours of tracing tiny differences among dozens of recorded MEG images, playing a critical role in preventing epileptic seizures or enabling continuous machine monitoring of epileptic patients in critical condition. Several advanced models regarding EEG classification have been published previously.[3] Significantly fewer publications address the issue of MEG classification using mostly advanced Artificial Neural Networks (ANN)[3] but none of them thoroughly examine basic ANN models. Therefore, our target was to test, compare, and evaluate some basic models to build a solid understanding of the characteristics of our data and gain insights into the model's behavior. This would serve as a preliminary study leading to a larger project to investigate more sophisticated, optimized, fine-tuned models. The results of our study are very encouraging and can constitute a basis for future research on epileptic seizure recognition, prediction, and prevention.

METHODS

Magnetoencephalography is a neuroimaging technique that utilizes an array of sensors placed slightly above the scalp. The use of Superconducting Quantum Interference Devices (SQUID) makes MEG very sensitive to the microscopic alterations of the magnetic field produced by brain cell electrical activity. Thus, it achieves a very good spatial resolution (up to 5 mm) as well as a great time resolution, at the scale of one millisecond or even better, which makes MEG a great tool for tracing real-time changes in brain activity and state. It can be used in combination with other imaging techniques (MRI, fMRI, PET, PET-CT) to give a detailed 3D imaging of brain activity in specific areas. It is non-invasive and it is completely safe, causing no discomfort. Moreover, it can detect epileptic activity and spot epileptic foci in the normal brain activity of the patient, without inducing unpleasant and even painful seizures to the patient.[4]

Epileptic activity appears in EEG and MEG as irregular patterns, in the form of spikes, spikes-and-slow waves, or sharp waves (Figure 1). The morphology of spikes and sharp waves in EEG was thoroughly analyzed by Gortman and these waves can be used for epilepsy diagnosis.[5] Although studies are being carried out[6], there still is no formal definition of epileptic spikes in MEG. However, even if it seems an oxymoron, compared to EEG signals, "MEG spike yield and localization are superior to EEG".[7] Epileptic signals in MEG have different morphological characteristics (duration, shape, and sharpness) from those in EEG. This can be explained by the small affection on the MEG signal from the interference from the skull and scalp. Furthermore, muscular activity and eye movement cause much less effect on MEG.[8] We should note that there is far less research that applies Deep Learning Models to MEG than to EEG. SQUID is a very expensive device with an even more expensive installation requiring a Faraday cage to isolate the super-sensitive SQUID from magnetic interference. This raises the total cost to a few million Euros. The MEG signals used in our study were recorded in the MEG Unit of the Laboratory of Medical Physics, Department of Medicine, Democritus University of Thrace, placed at Alexandroupolis, Greece, from patients who had been diagnosed with epilepsy by specialized neurologists and were referred for MEG evaluation. MEG signals were recorded with patients in a rest state and with eyes closed. In the present study, we worked on MEG signals recorded from 122 points of the patients' brains, with a sampling frequency of 256 Hz and 9 sec duration. A Low-Pass Filter with a cutoff frequency of 30 Hz was applied on all channels. Some channels contain out-of-limit values due to noise and artifacts and are thus rejected. The remaining signals were segmented into 5436 items of 1 channel—1 sec each, therefore each item contains 256 samples. Each item is classified individually by specialized neurologists as a signal-carrying epileptic activity (2059 items) or not (3377 items). Despite the absence of healthy patients, we have an adequate number of non-epileptic items in our dataset (62.1%). Our models were built and tested on a Hewlett-Packard Elite 800 G9 machine with an Intel i5 10500 3.1 GHz 6-core processor and 16 GB RAM, using MATLAB R2018a as the programming environment. In this study, we test a simple 3 and 4-layer CNN on colored images representing the heatmap of the signal spectrum as well as on black-and-white images of the signal. We also test a partially pre-trained Inception V3 on colored images of the signal spectrum. We then compare the results to those of a simple Feed-Forward Neural Networks (FFNN) with 3 and 4 hidden layers applied to the signal values of the same signals.

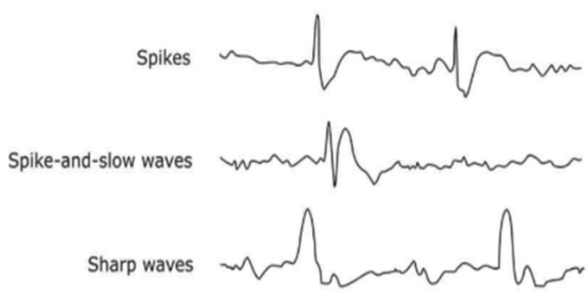

FIGURE 1. Spikes, Spike-and-slow waves, and Sharp waves.

Since Convolutional Neural Network Models are especially effective in image classification, we wanted to test the two models (3-4 layer CNN and Inception V3) on images of the MEG signal and images of the spectrum of the same signal. Similar work was carried out giving impressive results with images obtained from EEG.[9,10] We use a black-and-white bitmap image file, 256 × 256 pixels, to create images from the signal segments, 256 samples or 1 second long. Samples will appear as white dots on a black background and the displacement of dots from the middle of the image will be proportional to the value of the corresponding sample (Figure 2). To create images from the signal spectrum we apply Short-Time Fourier Transform on the signal with a shifting window 128 samples wide producing a heatmap-like RGB image, 30 × 32 pixels, and 3 color channels (Figure 3). Convolutional Neural Network is a Deep Learning model for processing data with grid pattern-like images. It's inspired by the optical cortex of animals, and it's designed to automatically learn and adapt to spatial feature hierarchies, beginning from low and moving towards higher-level patterns. Typically, it consists of three types of layers: convolution layer, pooling layer, and fully connected layer. The first two operate as feature extractors while the third maps the extracted features to the final output, performing classification. The role of the convolution layer is fundamental. Pixel values in digital images are stored in a two-dimensional grid, a matrix. An optimized feature extractor, called the kernel, is applied to each position of the image. Every layer's output is the input of the next layer, so the extracted features may progressively become more complex. The parameters of the kernels are optimized by training performed using the backpropagation algorithm—gradient descent.[11] For processing the black and white 256 × 256 signal images, we create a CNN of 3 convolutional layers. For processing the RGB 30 × 32 spectrum images, we create a CNN of 4 convolutional layers. Inception V3 is a Deep Learning model based on Convolutional Networks, used in Image Classification. It is an improved version of Inception V1, published as GoogLeNet in 2014, with 4 major modifications: Factorization into Smaller Convolutions, Spatial Factorization into Asymmetric Convolutions, Utility of Auxiliary Classifiers, and Efficient Grid Size Reduction[12], which was developed by a Google Team. Inception V3 is made of 42 layers, a few more than in V1 and V2. However, the effectiveness of the model is impressively boosted. As expected, Inception V3 has greater accuracy and a smaller computational cost compared to the previous versions. It even has lower error rates compared to previous and newer image classification models.[13] We download the pre-trained Inception V3 Network with all the necessary libraries and data from the MATLAB command line. Inception V3 is pre-trained on more than a million images of the ImageNet database. It has 316 layers in total and can classify images into 1000 object categories.[14] Unfortunately, we lack the computational power, memory, and time to perform full training on our own dataset. Instead, we keep the first 198 layers frozen by using the FreezeWeights function, leaving the other layers' parameters free to adapt during the training. We estimate that the first layers of the network form the simple, low-level patterns common to all kinds of images. The higher-level patterns built in the last levels, on the other hand, are more important and are the ones that give the differences in the images, so they need to be formed freely during the training. For better visibility of our images by the network, we augment the pixel range by using the imageDataAugmenter tool. Last, but not least, we conducted the simplest experiment of all. We fed our raw data in a relatively small FFNN—three hidden layers of sizes 64, 32, and 16. The Levenberg-Marqurdt algorithm is used to train the network and we use Mean Squared Error as the error measure. Each input item consists of 256 numerical sample values of our signal, one second long (1s). We based this last experiment on the estimate that the Neural Network has the flexibility to extract the correct relations by calculating the appropriate weights that minimize the error. Also, we know that the sample values contain all the information in the signal.

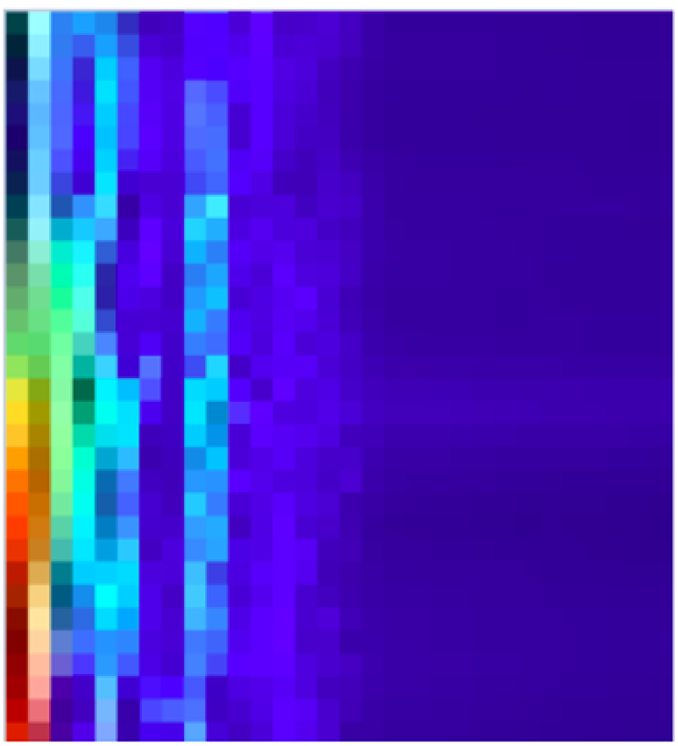

FIGURE 2. Color bitmap image representing the item's spectral heatmap.

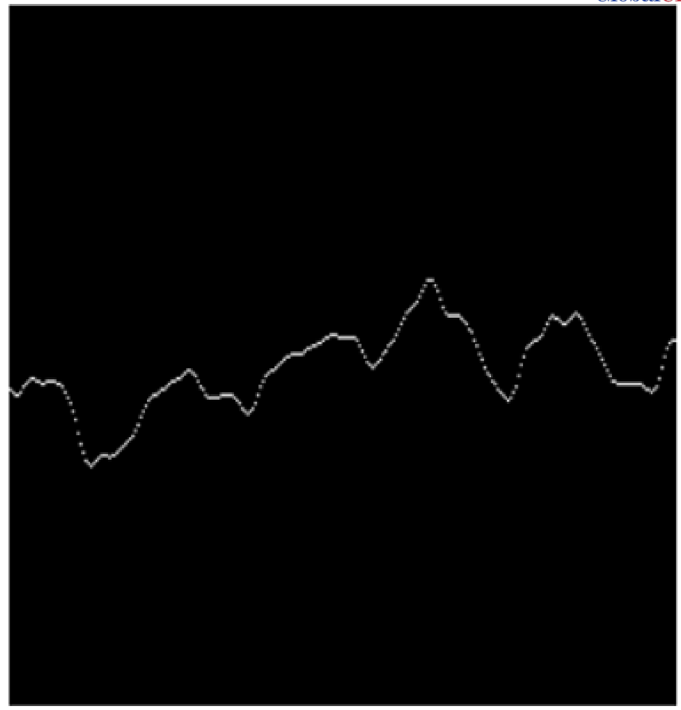

FIGURE 3. Black-and-white bitmap image produced from the item's signal.

RESULTS

We split our data into train and test-set randomly, 70% train-set and 30% test-set. All our models automatically use an independent part of the train set for validation. For the evaluation of our results, we focus mainly on the metric of accuracy (sensitivity and specificity were also measured). Accuracy is a qualitative metric of performance that gives the proximity of a measurement to the actual value.[15] We rely on this metric because our study consists of binary classification where the numbers of items of the two classes do not differ significantly. Also, accuracy includes other metrics (precision and trueness) and is a function of these metrics. We performed the training process of a three-layer CNN several times. In the best case, we have 85% accuracy while the average accuracy is 78.22% after approximately 7 minutes of training. Although far from perfect, these results are surprisingly good. Nobody expected a simple three-layer CNN could trace patterns and relations among them in a black picture with 256 small white dots. These results must be further studied and may help create optimized hybrid MEG Classification models. We follow the same procedure with the spectra images. This time, all experiments gave us very good results, averaging 90.15% accuracy (91.40% in the best case) within just a few seconds. We train the Inception V3 network with a maximum epoch number 6 and an initial learning rate of 10^{-4}. The training process lasts more than 50 hours. Unfortunately, this time the results are not so encouraging, considering the magnitude and complexity of the network as well as the time and effort spent. The prediction accuracy is only 74% and we can observe that only the first training epoch raises the accuracy slightly higher than 70%, while further improvement is very slow. The cause of this poor performance lies in the fact that the network was trained on an enormous set of common images that vastly differ from MEG images, and thus, even partially, creating low-level patterns that are useless in our case. Unfortunately, our resources and time did not allow us to design improved experiments with the Inception V3 network. The accuracy achieved by the FFNN when values of the signal samples are fed directly to it, exceeded all expectations, since it is, by far, the best result in all our experiments. In our third effort, after 47 minutes of training, we got a prediction with 95.4% accuracy which is impressive, considering the simplicity of the model and data. Driven by these encouraging results, we conducted a final experiment adding another layer of 32 neurons before the last 16-neuron layer. After approximately 2 hours of training, we obtain a trained network capable of slightly more accurate classification capability (96.2% accuracy). Tables 1 and 2 present the comparative results given by the three methods of classification we tested in our experiments.

TABLE 1. CNN and Inception V3 results (accuracy %).

Experiment	CNN on Signal	CNN on Spectrum	Inception V3
1	85.00	90.50	73.90
2	77.30	89.70	
3	82.30	90.60	
4	62.20	89.10	
5	84.30	89.60	
6	78.22	91.40	
Average	78.22	90.15	73.90

TABLE 2. FFNN results (accuracy % and network structure).

Experiment	FFNN on signal
1	87.60 [64 32 16]
2	88.50 [64 32 16]
3	95.40 [64 32 16]
4	90.90 [64 32 16]
5	96.20 [64 32 32 16]
Average	91.72

DISCUSSION

The obtained results, summarized in Table 1 for CNN and Inception V3 and in Table 2 for FFNN, are very promising since the FFNN on the signal values achieves impressive results (Accuracy = 96.2%) despite the simplicity of the model. Also, the CNN on the spectrum heatmap images shows good results with accuracy reaching 90%. These results underline the significance of the MEG as a powerful tool for obtaining high-resolution and high signal-to-noise ratio brain signals. The epileptic spikes appear sharper and are easier to observe in MEG than in EEG. The fact that epileptic spikes are more clearly observed in MEG than in EEG can be verified by the good results (Accuracy = 96.2%) of a simple FFNN. Even less sophisticated models perform well, showing the power of MEG as a diagnostic tool for epilepsy. A basic FFNN with 3 hidden layers is capable of successfully classifying MEG signals with an accuracy of up to 95.4% and a slightly larger FFNN with 4 layers can classify MEG Signals with 96.2% accuracy. A three-layer FFNN is equivalent to the extraction of third-order statistics from our data. The 256 raw sample values of our signal contain all the available information in the signal. Our three-layer network shows the ability to mine the information hidden in these values by creating the correct weighted combination of signal sample values. It is obvious that this simple model needs further investigation and experimentation. Unfortunately, our dataset lacks recordings of healthy subjects. It is essential to test our models on such a dataset. We need to test whether the binary classification models will suffice, considering the variability of the signals. Such a case might require more than binary classification (more than two classes). Although not as impressive, we also have a satisfactory performance of the CNN on the Spectra images, achieving 91.4% accuracy. It is observed that signals with epileptic activity contain high-frequency components, so we should try running experiments using a Low-Pass Filter with a higher cut-off frequency to avoid losing this information. Conclusions Based on the results and the discussion above, we need to focus on the following points in our future work: Obtain MEG recordings using a higher sampling rate and higher threshold on the Low-Pass Filter, more sophisticated filtering of the artifacts and noise, so that gain more information that lies within epileptic signals and an even clearer distinction to the non-epileptiform activity. Also, the inclusion of recordings from healthy subjects in the datasets is necessary to create more reliable models. Although we do not believe this would change the models' performance, since there are no serious deviations in the values of sensitivity and specificity, it is necessary to verify our models on a dataset nearer to real-world conditions, where epileptic signals are significantly fewer. Also, we should try to improve the quality of the images fed to CNN. A different colormap and a better analysis could potentially improve the performance of the CNN on the spectra image classification. Also, we need to try a different representation for the signal images so that the signal curve is clearly depicted. An Inception V3 Network with no previous training could prove more effective in such a case. If the untrained Inception V3 performs significantly better than the FFNN, we then can try to reduce the training time and cost by introducing some pre-trained layers. Area Under Curve (AUC) metric should be used for model validation, a step we omitted in our research due to the lack of time and resources. Metaheuristic searching algorithms, like Genetic Algorithms, should be used for FFNN structure optimization. Finally, we plan to develop an optimized hybrid NN Model by combining elements and layers from the best-performing basic models.

CONCLUSION

Based on the results and the discussion above, we need to focus on the following points in our future work: Obtain MEG recordings using a higher sampling rate and higher threshold on the Low-Pass Filter, more sophisticated filtering of the artifacts and noise, so that gain more information that lies within epileptic signals and an even clearer distinction to the non-epileptiform activity. Also, the inclusion of recordings from healthy subjects in the datasets is necessary to create more reliable models. Although we do not believe this would change the models' performance, since there are no serious deviations in the values of sensitivity and specificity, it is necessary to verify our models on a dataset nearer to real-world conditions, where epileptic signals are significantly fewer. Also, we should try to improve the quality of the images fed to CNN. A different colormap and a better analysis could potentially improve the performance of the CNN on the spectra image classification. Also, we need to try a different representation for the signal images so that the signal curve is clearly depicted. An Inception V3 Network with no previous training could prove more effective in such a case. If the untrained Inception V3 performs significantly better than the FFNN, we then can try to reduce the training time and cost by introducing some pre-trained layers. Area Under Curve (AUC) metric should be used for model validation, a step we omitted in our research due to the lack of time and resources. Metaheuristic searching algorithms, like Genetic Algorithms, should be used for FFNN structure optimization. Finally, we plan to develop an optimized hybrid NN Model by combining elements and layers from the best-performing basic models.

REFERENCES

1. Sidiropoulou, K. Basic Principles of the Nervous System Functioning. Open Academic Editions: National Technological University of Athens, Greece, 2015. Available online: https://openstax.org/books/anatomy-and-physiology-2e/pages/12-1-basic-structure-and-function-of-the-nervous-system.
2. Health Information. Available online: https://www.nih.gov/health-information.
3. Boonyakitanont, P., Lek-uthai, A., Chomtho, K., et. al. A review of feature extraction and performance evaluation in epileptic seizure detection using EEG. *Biomed Signal Process Control*. 2019;57(4):101702. https://doi.org/10.1016/j.bspc.2019.101702.
4. Hansen, P., Kringelbach, M., Salmelin, R. *MEG: An Introduction to Methods*. Oxford University Press: Oxford, United Kingdom; 2010; pp. 1–448. Available online: https://ora.ox.ac.uk/objects/uuid:9cd78f0a-f9e7-41d6-b25d-08b0c4d0f349.
5. Kim, T., Nguyen, P., Pham, N., et. al. Epileptic Seizure Detection and Experimental Treatment: A Review. *Front Neurol*. 2020;11:701. https://doi.org/10.3389/fneur.2020.00701.
6. Nowak, R., Santiuste, M., Russi, A. Toward a definition of MEG spike: Parametric description of spikes recorded simultaneously by MEG and depth electrodes. *Seizure*. 2009;18(9):652–655. https://doi.org/10.1016/j.seizure.2009.07.002.
7. Ossenblok, P., De Munck, J.C., Colon, A., et. al. Magnetoencephalography Is More Successful for Screening and Localizing Frontal Lobe Epilepsy than Electroencephalography. *Epilepsia*. 2007;48(11):2139–2149. https://doi.org/10.1111/j.1528-1167.2007.01223.x.
8. Khalid, M.I., Otaiby, T., Aldosari, S., et. al. Epileptic MEG Spikes Detection Using Amplitude Thresholding and Dynamic Time Warping. *IEEE Access.* 2017;5:11658–11667. https://doi.org/10.1109/ACCESS.2017.2718044.
9. Hirano, R., Emura, T., Nakata, O., et. al. Fully-Automated Spike Detection and Dipole Analysis of Epileptic MEG Using Deep Learning. *IEEE Trans Med Imaging*. 2022;41(10):2879–2890. https://doi.org/10.1109/TMI.2022.3173743.
10. Gao, Y.Y., Gao B., Chen, Q., et. al. Deep Convolutional Neural Network-Based Epileptic Electroencephalogram (EEG) Signal Classification. *Front Neurol*. 2020;11:375. https://doi.org/10.3389/fneur.2020.00375.
11. Yamashita, R., Nishio, M., Do, R.K.G., et. al. Convolutional neural networks: an overview and application in radiology. *Insights Imaging*. 2018;9(4):611–629. https://doi.org/10.1007/s13244-018-0639-9.
12. Szegedy, C., Liu, W., Jia, Y.Q., et. al. Going deeper with convolutions. In *Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* IEEE Xplore: Boston, MA, USA; 07–12 June 2015. https://doi.org/10.1109/CVPR.2015.7298594.
13. Deep Learning: Understanding the Inception Module. Available online: https://towardsdatascience.com/deep-learning-understand-the-inception-module-56146866e652.
14. Deep Learning Toolbox Model for Inception V3 Network. Available online: https://www.mathworks.com/matlabcentral/fileexchange/65679-deep-learning-toolbox-model-for-inception-v3-network.
15. Menditto, A., Patriarca, M., Magnusson, B., et. al. Understanding the meaning of accuracy, trueness and precision, Accreditation and Quality Assurance. *Accredit Qual Assur*. 2007;12(1): 45–47. https://doi.org/10.1007/s00769-006-0191-z.