

# Epileptic Seizure Prediction over EEG Data using Hybrid CNN-SVM Model with Edge Computing Services

Punjal Agarwal<sup>1</sup>, Hwang-Cheng Wang<sup>2,\*</sup>, and Kathiravan Srinivasan<sup>2</sup>

<sup>1</sup> LNM Institute of Information Technology, Jaipur, India

<sup>2</sup> National Ilan University, Yilan County, Taiwan

**Abstract.** Epilepsy is one of the most common neurological disorders, which is characterized by unpredictable brain seizure. About 30% of the patients are not even aware that they have epilepsy and many have to undergo surgeries to relieve the pain. Therefore, developing a robust brain-computer interface for seizure prediction can help epileptic patients significantly. In this paper, we propose a hybrid CNN-SVM model for better epileptic seizure prediction. A convolutional neural network (CNN) consists of a multilayer structure, which can be adapted and modified according to the requirement of different applications. A support vector machine is a discriminative classifier which can be described by a separating optimal hyperplane used for categorizing new samples. The combination of CNN and SVM is found to provide an effective way for epileptic prediction. Furthermore, the resulting model is made autonomous using edge computing services and is shown to be a viable seizure prediction method. The results can be beneficial in real-life support of epilepsy patients.

**Keywords:** Autonomous Edge Computing, Deep Learning, CNN, SVM, EEG, Epilepsy, Seizure Prediction, Brain-Computer Interface, Brain-Health Treatment

---

\* Corresponding author: [hcwang@niu.edu.tw](mailto:hcwang@niu.edu.tw)

# 1. Introduction

Epilepsy is an incessant brain disorder, in which patients suffer from the occurrence of spontaneous seizures. It is so common that it occurs to about 70 people out of 100,000 people per year [1] and if we look at the age-adjusted incidence of epilepsy, then we will find that there are 44 cases out of 100,000 every year [2]. In about 30% of the people who have epilepsy, it is still not cured even after they undergo surgeries, which is a considerable amount [3]. Anxiety in patients due to the possibility of a seizure occurring can drastically affect a patient’s life. According to previous studies, it has been found that about 1% of the world’s whole population suffers from epilepsy [4]. Since, Brain-Computer Interface acts as a middleware for the interaction between the brain signals and computer signals, therefore the interface needs to be improved in order for better interfacing. If there is direct communication, then obtaining and processing of data becomes very easy, which can overall increase the outcome of seizure prediction. If the prediction of seizures increases, then it would significantly improve the living expectations of over 50 million patients who are suffering from the ictal events. Schulze-Bonhage et al. [5], have shown some of the advantages of avoidance of injuries, increasing the feeling of security, driving without fear, and reduction of anxiety over the epileptic patients. Additionally, when the patients asked for some emergency help and early medications, these all helped them [6]. Therefore, many medications such as delivering fast-acting antiepileptic drugs, electric stimulation of vagus nerve or even the deep brain stimulation can be applied to overcome the seizures, which was analyzed by Schelter et al. [7].

Epilepsy causes abnormalities in EEG readings. After finding the suitable features, these are then used to classify into preictal and non-preictal states. The extracted features can be univariate (from a single channel), bivariate (from a pair of channels), or multivariate measures (from multiple channels simultaneously). Traditionally, epileptic seizure prediction was challenged either by using the necessary frequency filters, which merely applied a threshold to a given measure extracted from the EEG as employed by Schelter et al. [8], or using non-linear analysis as shown by Le Van Quyen et al. [9]. Recently, researchers started experimenting the classification process based on high-dimensional features spaces, which were used to detect the preictal states. The seizure prediction is tackled as a binary classification problem between the preictal and non-preictal states, where preictal states occur just preictal can cover from several seconds up to several hours before the seizure. The non-preictal class covers the three states of ictal, postictal, and interictal. The seizure happens in the time-period of ictal state. The postictal state encompasses the moments after seizure onset. Interictal state, during which the patient enjoys a normal brain activity, is the interval beginning right after the postictal state of a seizure and ending before the preictal state of the next seizure. before the seizure and need to be predicted to detect a preceding seizure. Depending on the starting time, the

A support vector machine is a discriminative classifier that is characterized by a separating optimal hyperplane, which is used for categorizing new samples. The SVM classifier established by Vapnik et al. [10] was employed for categorizing the preictal and non-preictal states based on the extracted time and

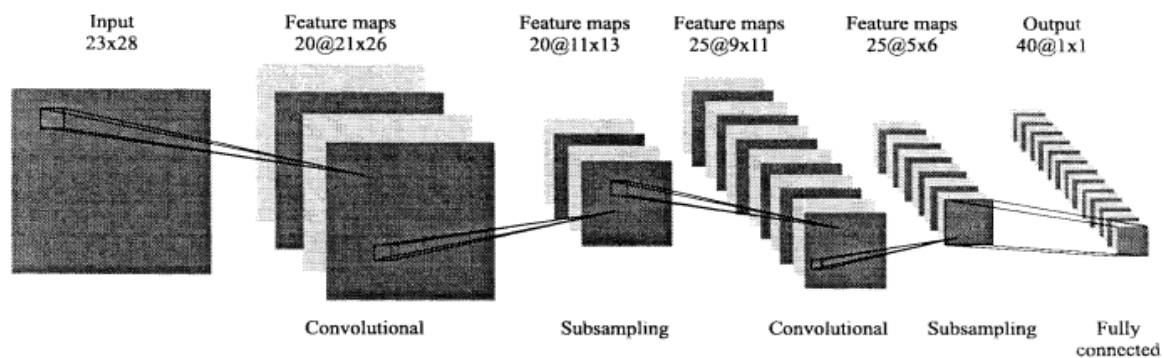


Fig. 1. Basic CNN architecture.

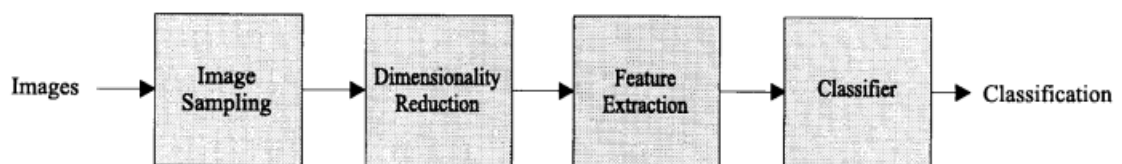


Fig. 2. A high-level block diagram of the system used by the CNN to classify images.

Almost every seizure prediction approach tries first to extract the main features from the pre-processed EEG signals, as these are the best to diagnose epilepsy.

frequency features and the features extracted from the deep learning method as the input for it. Subsequently, the SVM classifies the EEG data by determining the

optimal hyperplane which divides the two classes with the highest margin. The following equation can be used to express the obtained hyperplane:

$$f(\mathbf{x}) = \text{sgn} \left( \sum_{i=1}^N \alpha_i d_i k(\mathbf{x}_i, \mathbf{x}) + b \right) \quad (1)$$

where  $N$  is the number of support vectors,  $d_i$  exemplifies the respective class membership, i.e., binary labels from  $\{-1,1\}$ ,  $\mathbf{x}_i$  denotes the  $i^{\text{th}}$  support vector,  $\mathbf{x}$  specifies the feature vectors,  $k(\mathbf{x}_i, \mathbf{x})$  indicates the kernel function, and  $\alpha_i$  is a slack variable introduced for combining the constraint functions together as a new Lagrangian function. The identification of preictal state is accomplished using the computational intelligence SVM classifier and its output regularization by firing power (FP) as implemented and explained by Teixeira et al. [11].

In our study, we have also used CNN or Convolutional Neural Networks, which are a particular class of deep-networks based on the type of feedforward artificial neural networks. These are based on the 3D neuronal arrangements, local connectivity between neurons of adjacent layers and shared weight vectors. These properties help in better generalization with lower memory which is appropriate for feature representations in the big-data problems correlating with seizure prediction in epilepsy. CNN consists of multilayer structures, which can be developed and modified according to use. The first layer defines the input dimensions, followed by rectified linear units and max-pooling layers [12]. The basic structure of a CNN is shown in Fig. 1 and Fig. 2 [13]. The last layers are the fully connected layers. Basically, these are the layers where the classification is done based on the features provided by the last convolutional layer. More details of CNN will be described in Section 3.3.3 when high-level feature extraction using deep learning is discussed.

The remainder of the paper is organized as follows, Section 2 provides the autonomic edge computing infrastructure. Section 3 describes our proposed novel approach for epileptic seizure prediction. Results are presented and discussed in Section 4. Finally, Section 5 examines future work that can be done.

## 2. Autonomic Edge Computing

Computer systems and applications have seen explosive growth over the years regarding diversity, scale, and complexity. All the conventional approaches to system management, which requires extensive

human intervention, are deemed to be inefficient to cope with the ever-increasing and expanding computer and network systems. Therefore, a new notion of computing was developed known as Autonomic Computing (AC). An autonomic computing system is characterized by the following attributes [14] [15]:

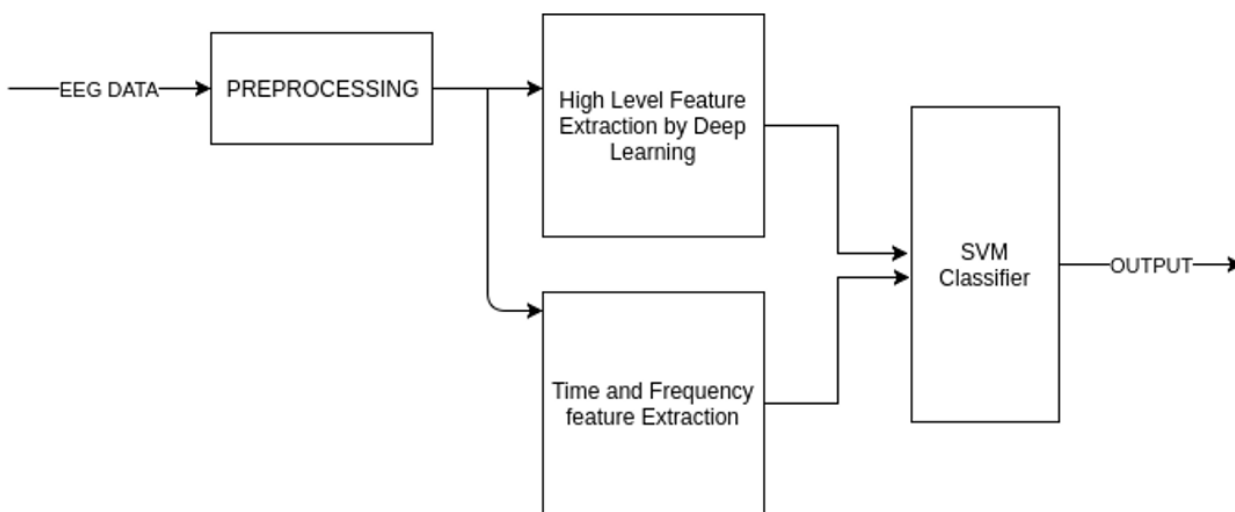
- Self-awareness: An Autonomic Computing system should exhibit a sense of self, its state and behavior.
- Self-configuration: An AC system should be able to configure itself by the environment adaptively.
- Self-healing: An AC system should be fault-tolerant and able to recover from problems.
- Self-optimizing: An AC system should be able to adjust its parameters and behavior to optimize its performance in response to the changing conditions.
- Self-protecting: potential threats and be able to protect itself from attacks and maintain integrity.
- Context awareness: An AC system should be aware of the environment and adapt to changes in the execution environment.
- Openness: An AC system should be able to function appropriately across heterogeneous hardware and software platforms, and therefore should be based on open standards rather than proprietary components.
- Anticipatory power: An AC system should possess the ability to anticipate proposed changes and proactively take necessary measures.

Mobile communications have undergone four generations of evolution in just a few decades, bringing tremendous convenience and profound social and economic impact. The type of data transmitted by mobile communication also has also shifted from voice to video and multimedia. The next generation (5G) mobile communication system needs to carry more diversified services, such the emerging Internet of Things and driverless vehicles, exerting enormous pressure on the system. Therefore, 5G has taken many new approaches in technology development, such as new radio (NR), large antennas, novel channel coding technologies, migration to higher frequency bands for higher bandwidth, differentiating application scenarios to meet different requirements. Also, 5G embraces software and virtualization technologies to reduce cost and provide greater flexibility.

Cloud computing has become a trend in recent years, providing a wide range of services and also using its flexible resource allocation methods to solve many practical problems. The development of 5G

### 3. Proposed Approach

Our proposed approach firstly tries to pre-process the data using fast Fourier transform to extract the



**Fig. 3.** Proposed approach.

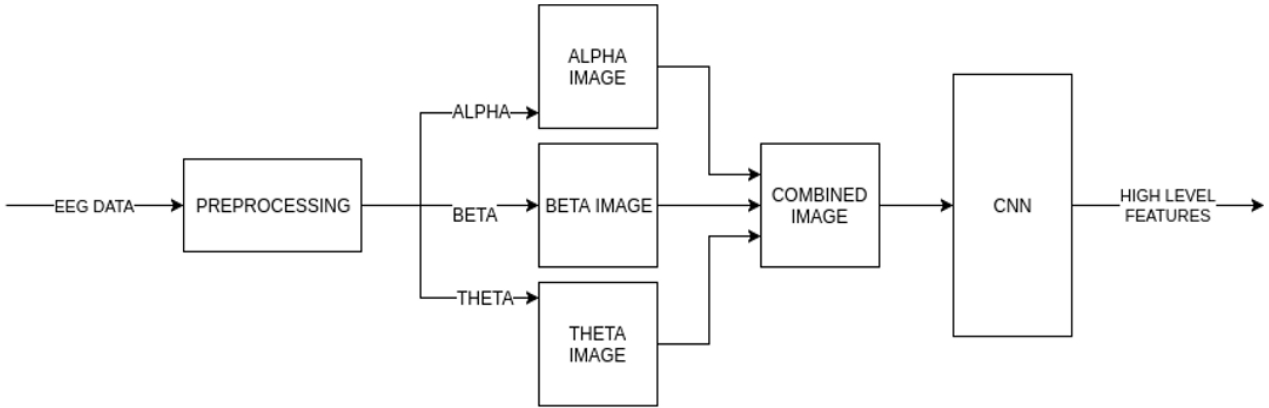
mobile communications has led to the increasing popularity of resource-demanding and delay-sensitive mobile applications. Recently, the computing facilities and storage capabilities of cloud data centers have been moved closer to the end users, ushering in the era of mobile edge computing (MEC). As MEC servers are close to the end users, their services offer the advantages of low latency and high energy efficiency.

These desirable features can be critical to many applications in the medical areas, in particular, in the realization of real internet [16]. Indeed, edge computing has been considered to play an essential role in the realization of ultra-reliable, low-latency communications (URLLC) in 5G [17]. The low-latency property has been exploited in [18]. However, the diverse applications and services that are foreseen to be offloaded to MEC servers will render the management and resource allocation a highly complicated issue. Moreover, the environment is likely to be dynamic and changes rapidly over time. Autonomic computing is thus key to the successful implementation of edge computing. In this paper, we tackle the critical issue of epileptic seizure prediction using a novel hybrid deep-learning approach based on an autonomic edge computer infrastructure.

time and frequency features. Simultaneously, the data is also pre-processed to extract the different information stored in different frequency bands, namely, Alpha, Beta, and Theta. This information is then further used to separately generate EEG 2D images which are then fed to a CNN model to classify between the preictal and non-preictal images. The high-level features are gathered from the last layer of the trained CNN model. These high-level features with time-frequency features together are fed to an SVM which is trained over per-patient data to see the output. Block diagram of the whole process is shown in Fig. 3 and the individual steps are explained in the sequel.

#### 3.1. Pre-processing

The whole process of pre-processing consists of many parts as there are different types of features being extracted. For the time and frequency feature extraction, the EEG data pre-processing is straightforward. The fast Fourier transform is applied to each 1-second clip, where the frequency is taken in the range 1-47Hz, and phase



**Fig. 4.** High-level feature extraction process.

information is discarded. Moreover, for creating EEG images, the FFT is applied over the EEG data separately, for different frequency bands: Alpha (8-13 Hz), Beta (13-30 Hz) and Theta (4-7 Hz), as all the oscillatory cortical activity related to memory operations primarily exists in these three frequency bands.

**3.2. Time and Frequency based feature extraction**

Schindler et al. [19] analyzed and showed that correlation coefficients and their corresponding eigenvalues in the time domain are great features for efficient seizure prediction. By experimenting, we found that if these are computed for the frequency domain as well, they can contribute significantly towards the seizure prediction. Therefore, correlation coefficients and their eigenvalues are found by mathematical calculation in the time and frequency domains, as outlined below.

- Simple FFT extracted features for the range of 1-47 Hz.

Here the FFT is used over each clip for all EEG channels. Then a log<sub>10</sub> is taken, and phase information is removed.

- Correlation coefficients and their respective eigenvalues for the time domain.

The output from the above procedures is then normalized, and a correlation coefficients matrix is generated by treating each single EEG channel as one, subtracting the mean and dividing the standard deviation, as shown in (2) and (3) [20], where *n* is number of subjects, *p* is the number of variables. Eq. (2) is the Pearson correlation coefficient between variables *x<sub>j</sub>* and *x<sub>k</sub>*, and Eq. (3) is the correlation matrix. Then, real eigenvalues are generated by just considering the magnitude of the complex values and discarding the imaginary one.

$$r_{jk} = \frac{S_{jk}}{S_j S_k} = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)}{\sqrt{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2} \sqrt{\sum_{i=1}^n (x_{ik} - \bar{x}_k)^2}}, \tag{2}$$

$$R = \begin{pmatrix} 1 & r_{12} & r_{13} & \dots & r_{1p} \\ r_{21} & 1 & r_{23} & \dots & r_{2p} \\ r_{31} & r_{32} & 1 & \dots & r_{3p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & r_{p3} & \dots & 1 \end{pmatrix}. \tag{3}$$

- Correlation coefficients and their respective eigenvalues for the frequency domain. A similar approach is taken for the frequency domain as well, as explained before for the time domain.

Thus, eventually there will be totally three types of time and frequency features in the feature subset.

**3.3. High-level feature extraction via deep learning**

As introduced before, CNN's tend to have a hierarchical structure. The first few layers are combined with Convolutional layers, ReLU layers, and Max pooling Layers. Subsequently, the end layers consist of 2 or fully connected layers which can also be added or discarded as per the requirement. We start the high-level feature extraction by converting the EEG Time-series data into images and then running a CNN model developed by us over these images. Finally, the CNN model generates the features, which are taken from the last layer of the CNN model and appended to the primary feature dataset. Refer to Fig. 4 for details.

### 3.3.1. Converting EEG Time-series data to images

As the FFT is performed and the different band information are gathered. This information is then further squared and summed for each frequency bands to create a separate measurement for each electrode. Rather than aggregating the obtained information to form a feature vector, the information is transformed into a 2D image, which preserves the spatial information. Pouya Bashivan et al. [21] have implemented this method to correctly store the relative distance between neighboring electrodes as well as their information is correctly told. Azimuthal equidistant projection process was employed in this process to obtain the relative distance between electrodes and correctly projecting them over 3D space. After the 3D projection, a 2D projection can be easily obtained which can represent the position of each electrode over 2D space, and the magnitude of the intensity at that point over the 2D space would represent the information encoded. Further, this was done for each frequency band of importance, and hence three types of topographical activity maps (each for Alpha, Beta, and Theta frequency bands) are obtained which are combined to form one image containing all three bands and all different electrodes values.

### 3.3.2. CNN Model

The multichannel image generated is given as the input to the CNN model. Our CNN model is similar to VGG. As VGG requires relatively less number of epochs to train, we used a similar model to perform more experimentations. It consists of four stacks, with the first stack having four layers, the second stack also having four layers, the third stack having two layers, and the last stack having one layer. Every layer in a stack is a combination of convolutional layer followed by the softmax layer. A max-pooling layer follows each stack. All convolutional layers use small receptive fields of size  $3 \times 3$  and stride of 1 pixel with ReLU activation function. Max-pooling is performed over a  $2 \times 2$  window with a stride of 2 pixels. The number of kernels within each convolution layer increases by a factor of two for layers located in deeper stacks. Simonyan et al. [22] experimented and analyzed that stacking of multiple CNN layers is leading to the active adoption of higher dimensions and also requires very few parameters.

### 3.3.3. Training and Testing Protocol

For training, we used Stochastic Gradient Descent (SGD) as the optimizer as it has shown the better results in most of the previous works as compared to other optimizers. SGD can easily show significant results using even slower learning rates, which makes it the best option. Since CNN uses backpropagation, there can be a considerable difference among the weights of the first and the last layers. We used the cross-entropy as the loss function to evaluate the model. The training was run on all of the samples obtained from the dataset. The learning rate was kept at 0.01,

decay was set to  $1e-6$ , and momentum is equal to 0.9. Further experimentation proved that playing with these hyper-parameters would increase/decrease the accuracy by a factor of 0.1-0.5.

For the first ten epochs, the model showed 95.09% best test accuracy and 93.05% best validation accuracy. For the first 30 epochs, the model's best test accuracy rose to 96.43% and 94.20% for best validation accuracy. Dropout regularization has also played a significant role in the model's accuracy as these are best used for reducing overfitting in deep convolutional neural networks [23]. No data augmentation and manipulation techniques were used to preserve the distinct spatial interpretations of direction and location in EEG images. The network started to converge after 30 epochs, i.e., 3600 iterations (120 iterations per epoch).

### 3.4. Processing the extracted features

After serious experimentation, it was decided to use these features in the model. Many earlier works had shown that EEG data could be analyzed correctly to obtain significant time and frequency features. Furthermore, the convolutional neural network's great work over the images leads us to choose and experiment with these features. The set of features, being extracted in the given section are supplied to a non-linear SVM with a Gaussian Radial Basis Function (GRBF) kernel, as SVM uses a unique transformation of the feature space [24] into a higher order space where linear boundaries may eventually separate the data into two classes. These transformations just linearize the space which is implemented using the kernel functions. The following equation can be used to express the GRBF kernel function:

$$k(\mathbf{x}, \mathbf{x}') = \exp(-\|\mathbf{x} - \mathbf{x}'\|^2 / 2\sigma^2), \quad (4)$$

where  $\mathbf{x}$  and  $\mathbf{x}'$  are the feature vectors in the input space.

For improving the results, GRBF kernel parameters are optimized by increasing a classical class separability criterion as the trace of the scatter ratio. The SVM was trained over four feature sets appended one after the other. Also, per-patient SVM classifier was trained where we kept the high-level deep-learning features same for every patient but changed the time and frequency features for different patients.

### 3.5. Autonomic Edge Computing Analysis

Since there were no edge servers accessible at the time, we used a simple analytical approach to analyze the time it took for the brain-computer interface. The goal is to compare cloud computer network (CCN) and edge computing node (ECN). Assume that the CCN is approximately 300 Km away from the source, and an ECN is just 1 Km away. The link speed between the source and CCN is 2 Mbps due to the presence of intermediate nodes and that for ECN is 100 Mbps. The TUH-EEG corpus, which contains a high proportion of epilepsy-related disorders, has a median file size of 4.1

Mbytes in gzip compressed format for each patient [25]. This file size will be used in subsequent analysis.

## 4. Results and Discussion

The ECoG dataset of eight epilepsy patients was used to experiment on, which was developed jointly by the Mayo Clinic and University of Pennsylvania. The dataset is also present on the Kaggle website and sponsored by the American Epilepsy Society [26]. For cross-validation, heavy experimentation is required. Thus, firstly the dataset was divided according to the whole seizures, i.e., if a ratio of 0.5 is assumed and there are four whole seizures, then two would be in one set and the other two in another. Many other ratios have been chosen as well but this turns out to be the best scenario with the highest accuracy along with excellent sensitivity and specificity. The proposed CNN-SVM model was simulated for training and testing work on a GPU. For ten epochs, high-level deep learning features, the SVM showed a  $(97.07\pm 0.5)\%$  accuracy, but it did not rise significantly for 30 epochs high-level deep learning features. The accuracy just rose to  $(97.86\pm 0.5)\%$ , whereas sensitivity  $(96.47\pm 0.5\%)$  and specificity  $(98.81\pm 0.5\%)$  were almost similar in both cases.

Since this whole process was carried out with the intention of creating a better brain-computer interface system for epileptic seizure prediction, the latency using edge computing service was also found. Assume an EEG file of size 4.1 Mbytes was transmitted from the source to compare the results between CCN and ECN. Provided the queuing delay and processing delay at the intermediate nodes are ignored, the total delay for CCN and ECN is the summation of transmission delay and propagation delay. Based on the analytical model described in Section 3.5, the one-way transmission and propagation delays for CCN are 16.4 s and 1000  $\mu$ s, and those for ECN are merely 328 ms and 3.33  $\mu$ s, respectively. Thus, through simple analysis, we can see that the total delay incurred by ECN is much smaller than that incurred by CCN. Here, for simplicity, the processing time has not been taken into account, which will be investigated in future research.

## 5. Conclusion and Future Work

Based on our research, it can be concluded that the latency and RTT measured over the network showed that edge computing services could be used in the future for creating a better brain-computer interface for epileptic seizure prediction. The whole system can be deployed over edge computing servers where the data packets containing the EEG signals in the compressed form can be sent, and processing can be done. The final results can be obtained which makes this brain-computer infrastructure faster and better as compared to the present state-of-the-art systems. Further, this work can be expanded over to make an end-to-end system where doctors can directly use the system to detect and localize the area of epilepsy in the patient's brain.

The current approaches involve high-complexity methods, which can take a long time to process and show the results. A system can be developed which not only analyzes EEG signal data but can also detect and analyze other bio-signals such as ECG, EMG, MMg, EOG, and so on, which require heavy computations and processing. A variety of models can be developed to evaluate different kinds of biological disorders which require quite complicated diagnosing. Moreover, a whole medical diagnosing brain-computer infrastructure using edge computing services could be developed. We are currently working on a few of these models to significantly improve the autonomic diagnosis. Finally, it is well known that CCN is highly computational efficient as compared to ECN. In this article, only data propagation and transmission delays have been considered in the comparison between cloud and edge computing. This is far from a thorough comparison. In the future, the computation complexity of the proposed approach will be analyzed, followed by a comprehensive comparison of performance when it is implemented in cloud and edge computing settings.

This research was conducted under the auspices of the Ministry of Education, Taiwan, which provided partial financial support through the TEEP@India 2018 project.

## References

1. J. Christensen, M. Vestergaard, M. G. Pedersen, C. B. Pedersen, J. Olsen, P. Sidenius, Incidence and prevalence of epilepsy in Denmark, *Epilepsy Research*, **76**:60–65 (2007)
2. W. A. Hauser, J. F. Annegers, L. T. Kurland Incidence of epilepsy and unprovoked seizures in *Epilepsia* Rochester, Minnesota: 1935–1984, **34**:453–458 (1993)
3. M.-P. Hosseini, M.-R. Nazem-Zadeh, D. Pompili, K. Jafari-Khouzani, K. Elisevich, H. Soltanian-Zadeh, Comparative performance evaluation of automated segmentation methods of hippocampus from magnetic resonance images of temporal lobe epilepsy patients, *Medical Physics*, **43**(1): 538–553, 2016.
4. World Health Organization, Epilepsy. [Online]. Available: <http://www.who.int/mediacentre/factsheets/fs999/en/>
5. A. Schulze-Bonhage et al. Patients' views on EEG-based seizure prediction devices, *Epilepsia* **51** (2010): 21.
6. A. Shoeb et al. Vagus nerve stimulation triggered by machine learning Based seizure, in *Epilepsy*: 385-395, CRC Press (2011)
7. B. Schelter et al. Seizure prediction: the impact of long prediction horizons, *Epilepsy Research* **73**(2): 213-217 (2007).
8. B. Schelter et al., Testing statistical significance of multivariate time series analysis techniques for epileptic seizure prediction, *Chaos: An Interdisciplinary Journal of Nonlinear Science* **16**, 013108:1-10 (2006).

9. M. Le Van Quyen et al., Anticipation of epileptic seizures from standard EEG recordings, *The Lancet* **357**:9251:183-188 (2001).
10. C. Cortes, V. Vladimir, Support-vector networks, *Machine Learning* **20**(3):273-297 (1995).
11. C. A. Teixeira et al., EPILAB: A software package for studies on the prediction of epileptic seizures, *Journal of Neuroscience Methods* **200**(2):257-271 (2011)
12. S. Minaee, W. Yao, Fingerprint recognition using translation invariant scattering network, *Proc. IEEE Signal Processing Medicine and Biology Symposium (SPMB)* (2015)
13. S. Lawrence et al., Face recognition: A convolutional neural-network approach, *IEEE Transactions on Neural Networks* **8**(1) 98-113 (1997)
14. M. Parashar, S. Hariri, Autonomic computing: an overview, in J.-P. Banatre et al. (Eds.): UPP 2004, LNCS **3566**:247–259 (2005).
15. R. Sterritt, Automatic computing, *Innovations in Systems and Software Engineering*, **1**:79-88 (2005) DOI : 10.1007/s11334-005-0001-5.
16. M. Simsek, A. Aijaz, M. Dohler, J. Sachs, G. Fettweis, 5G-Enabled tactile internet, *IEEE Journal on Selected Areas in Communications*, **34**(3):460-473 (March 2016)
17. M. Patel et al., Mobile-edge computing, Eur. Telecommun. Standards Inst., Sophia Antipolis, France, *White Paper* (2014).
18. M.-P. Hosseini, T. X. Tran, D. Pompili, K. Elisevich, H. Soltanian-Zadeh, Deep learning with edge computing for localization of epileptogenicity using multimodal RS-fMRI and EEG big data, *Proc. IEEE International Conference on Autonomic Computing*, 83-92 (2017)
19. K. Schindler et al., Assessing seizure dynamics by analysing the correlation structure of multichannel intracranial EEG, *Brain* **130**(1):65-77 (2006)
20. T. J. Archdeacon, *Correlation and Regression Analysis: A Historian's Guide*. Univ. of Wisconsin Press (1994)
21. P. Bashivan et al., Learning representations from EEG with deep recurrent-convolutional neural networks, *arXiv preprint arXiv:1511.06448* (2015).
22. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556* (2014)
23. A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet classification with deep convolutional neural networks, *Advances in Neural Information Processing Systems* (2012)
24. Chang, Y.-W., et al. Training and testing low-degree polynomial data mappings via linear SVM, *Journal of Machine Learning Research*, **11**:1471-1490 (April 2010)
25. I. Obeid, J. Picone, The Temple University Hospital EEG data corpus, *Frontiers in Neuroscience*, **10**(196):1-5 (2016)
26. M. Stead et al., Microseizures and the spatiotemporal scales of human partial epilepsy, *Brain*, **133**(9): 2789-2797 (2010)