Multi-label Classification for Epileptic Seizure Recognition: Deep Neural Network Ensemble versus Choquet Fuzzy Integral Fusion

Simone A. Ludwig
North Dakota State University
Fargo, ND, USA
simone.ludwig@ndsu.edu

***Abstract*—Epilepsy is a chronic neurological disorder. The cause are unprovoked recurrent seizures that patients experience. The most commonly used tool for the diagnosis of epilepsy is the electroencephalogram (EEG). The EEG measures the electrical activity of the brain. The patients have to be monitored as to detect an epileptic episode early on in order to prevent associate risks. Research in the past has used a combination of time and frequency features for the automatic recognition of epileptic seizures. An classification approach has been used to automatically detect epileptic seizures. In this paper, the epileptic seizure recognition data set is used for the investigation. Two fusion methods, ensemble and Choquet fuzzy integral, are compared using different deep neural network architectures. To aid the comparison, evaluation measures such as confusion matrix, AUC and accuracy are used in conjunction with MSE and RMSE. The results of the experiments show that the Choquet fuzzy integral fusion method outperforms the ensemble method. In addition, other state-of-the-art classification methods are also outperformed by the Choquet fuzzy integral fusion method.**

I. Introduction

Epilepsy is a very common neurological disorder affecting one in every 100 persons around the world [1]. Paroxysmal abnormal ultra-synchronized electrical activity can be measured in the brain, which usually occurs suddenly when a person has an epileptic episode. Researchers are looking to automate the monitoring and detection of an epileptic episode such that the patient can be made aware of early on to prevent potential risks [2]. However, one of the challenges remains that is not easy to detect and that is that the time frequencies of epileptic episodes are uncertain.

Different types of sensors [3] have been used to collect biological data from the patients' surface such as electrocardiogram (ECG), electromyography (EMG) [4], motion data [5], and electrodermography (EDG) [6]. Most of the sensors are commonly integrated into clothing such as E-textiles [7], capacitive sensing [8], polymer materials, such as carbon nanotube (CNT)-polydimethylsiloxane (PDMS) [9], Ag/AgCl electrodes [10], and micro-needle arrays [11]. The advantage of the "wearable" sensor systems is that they can monitor the signals of epileptic patients for long periods of time. On the other hand, measuring brain signals can provide faster and more usable information and are less likely to contain noisy signals.

Given that EEG is more accuarate, many researchers have looked at different approaches to directly measure 'epilepsy signals'. The different approaches include positron emission tomography (PET), single photon emission computed tomography (SPECT), magnetic resonance imaging (MRI), and functional magnetic resonance imaging (fMRI) [12]. At the point in time, most research studies make use of video-electroencephalograms (EEGs) [13], [14]. EEG signals provide temporal information and spatial information of the electrical brain activity. The video-EEG technique is currently seen as the best approach for studying epilepsy. The differentiation of rhythmic discharges from non-stationary processes provides challenges to the analysis of the EEG signals since the physiological processes of a seizure are typically non-stationary, dynamic, and nonlinear.

The automated way to detect EEG signals are feature extraction and classification whereby the extracted features can be divided into four categories: (1) statistical features, (2) fractal dimension features, (3) entropy features, and (4) time-frequency domain features. Different research studies have primarily used a combination of time and frequency features for the recognition of epileptic seizures.

The classification task to automatically detect EEG uses machine learning approaches that are mostly supervised-learning based. In this paper, the 5-class epileptic seizure recognition data set is investigated. In particular, two fusion methods are compared, which are an DNN ensemble method and the Choquet fuzzy integral fusion method.

II. Related Work

A lot of related work has used the Bonn datasets [15]. For example, a neural network (NN) classification technique was applied in the field of brain science as outlined in [16]. Another machine learning method called support vector machines (SVMs) was used to identify the EEG signals of epilepsy patients showing good recognition performance [17], [18], [14]. A modification based on the least squares support vector machine (LS-SVM) was proposed in [19] to classify two-class seizure and non-seizure EEG signals. The results reported a 98.0-99.5% accuracy using the radial basis function (RBF) kernel, and a 99.5-100% accuracy using the Morlet kernel.

Another approach used an Ada-Boost classifier to identify spike detection of epileptic seizures [20]. Given the implications of the no-free-lunch theorem [21], several classification algorithms have been applied to seizure detection, including random forests (RF), K-nearest neighbors (KNN) [22], and Bayesian neural networks [23]. These approaches achieved accuracy values ranging from 93% to 99.66%. One shortcoming was that these accuracy results only used binary classification and are also too time consuming for some practical clinical applications.

A three-label classification problem was studied in [15] whereby the distinction was made between continuous ictal epilepsy patients, intermittent epilepsy patients, and healthy subjects. The researchers used a SVM-based recognition system and obtained an accuracy of 93.9%.

Related work that focuses on the data set that is being used in this research study, the following can be listed. Researchers use deep learning methods to predict epileptic seizures [24]. The authors used a deep learning classifier to identify the signals that occur before and after a seizure. Afterwards, the classifier performance was tested on data from all patients and was compared against the performance of a random predictor.

A deep learning model with automatic learning features was built in [25]. More specifically, a CNN (convolutional neural network) as the deep learning method was used to train the model. Different types of interictal epileptiform discharges (IEDs) within the group as well as features invariant to time differences between the IEDs were identified. Please note that IEDs are pathological patterns of activity between seizures the brain of patients with epilepsy create.

Authors in [26] trained deep neural networks with EEG data for predicting seizures by simultaneously collecting spectral, temporal and spatial information for the analysis of seizures. Their study mostly focused on the cross-patient study of predicting the seizure and the outcome showed that the deep learning model generalizes well among different patients.

III. Approaches

A. Deep Neural Networks

Deep learning is a term used to describe the different learning approaches/architecture of artificial neural networks. The different architectures that are included in deep learning are deep neural networks (DNN), deep belief networks, recurrent neural networks and convolutional neural networks. These architectures have been widely applied to many different research areas. The research areas include audio recognition, speech recognition, natural language processing, computer vision, bioinformatics, gaming. In particular, a DNN [28] can be described as containing an input layer, several hidden layers, and an output layer. The DNN is trained using backpropagation whereby the error between the actual output and the desired output is to be minimized.

B. Ensemble

Ensemble learning was first introduced in 1979 [29] when an ensemble system in a divide-and-conquer fashion was applied to a feature space that was partitioned using two or more classifiers. Another ensemble system later was introduced showcasing the generalization performance whereby similar neural network configurations can be improved using ensembles by introducing the variance reduction property [30]. However, it was research reported in [31] that placed ensemble systems at the center of machine learning research. The proof was that a strong classifier can be generated by combining weak classifiers through a procedure called boosting.

Nowadays, ensemble methods are heavily used due to their success primarily when applied to classification tasks. Ensemble methods train multiple learning algorithms, and thus achieving significantly higher accuracy than a single learner [32]. The common methods that are used are for ensemble learning are boosting, bagging, stacking.

Boosting uses a model that was trained on data and incrementally constructs new models that focus on the errors in the classification made by the previous model, thus making incremental improvements. An example of a boosting algorithm is XGBoost [33].

Bagging on the other hand involves the training of models based on random subsamples. Each model votes with equal weight on the classification task. For example, Random forest uses a bagging approach to allow the selection of a random set of features to be used [34].

The third method referred to as stacking takes the output of a set of models and feeds them into another algorithm that combines them and thus making a final prediction. Any set of base learners and combiner algorithm can be used for this. The combiner algorithm can either use a simple or a weighted average approach.

C. Choquet Integral

Figure 1 shows the fusion process whereby the data set is provided to the three different DNN. Then, the fusion is performed by taking the learned densities as well as the classification performance of the three DNN models. The result reports on the classification performance that is based on unseen test set.

The Choquet integral (ChI) [35] - [38] is a well-known parametric function used for data and information fusion. Moreover, ChI is a generator function that is parametrized by a fuzzy measure (FM). This FM is monotone and normal. Once the FM has been determined the ChI turns this into a specific aggregation operator [39].

The basic idea of a fusion algorithm is that the algorithm should prioritize the most accurate evidence among the different inputs however also considering the contribution any input makes. The Choquet fuzzy integral achieves this by making use of a non-linear weighted average of all data sources. The defined fuzzy measure, takes the incoming evidence that is weighted by a fuzzy measure value, and this is summed over to produce a single confidence value. In most cases, the Sugeno λ-measure is used, which needs to be initialized for the data source subsets. These data source subsets can be thought of as different values of importance that represent each data

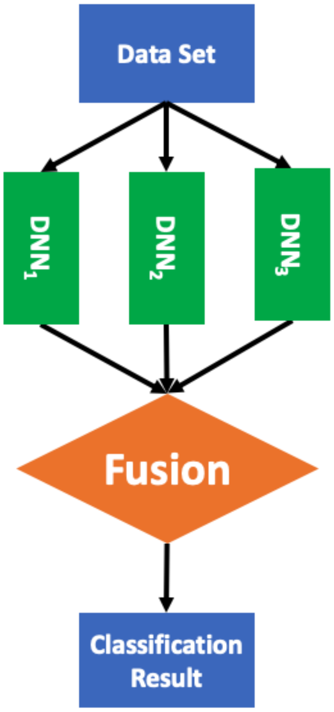

Fig. 1. Choquet Fuzzy Integral Fusion Process

source. These values of importance are referred to as densities and are often defined by experts or by some metric. For our experiments, the AUC (Area Under the Curve) of ROC curves is used as the measure. For more detailed information as well as the equations please refer to [40].

IV. Experiments and Results

A. Data Set Description

The Epileptic Seizure Recognition data set [41] is used for this study. It contains 4,097 data points collected from a EEG recording. Each data point represents the value of the EEG recording at a particular time period. The data set contains recordings from 500 individuals. The data was divided and shuffled into 23 chunks whereby each chunk contains 178 data points (features). Therefore, 23 chunks times 500 individuals results in an overall value of 11,500 rows and 178 columns plus the last column, which represents the class label. The following class label are used:

- 1: Recording of seizure activity
- 2: Recording of tumor area
- 3: Recording from the healthy brain area
- 4: Recording when patient had their eyes closed
- 5: Recording when patient had their eyes open

In past research investigations, this data set was primarily used as a binary data set where classes 2, 3, 4 and 5 were categorized as not having an epileptic seizure, versus class 1 having an epileptic seizure. However, since binary classification is an easier task, in this paper we are using this data set for a multi-label classification investigation.

TABLE I
AUC and Accuracy Results in %

	DNN1	DNN2	DNN3	Ensemble	Choquet
AUC	41.29	42.07	36.93	36.93	**50.72**
Accuracy	61.95	57.95	64.21	63.37	**89.57**

TABLE II
MSE and RMSE Results

	DNN1	DNN2	DNN3	Ensemble	Choquet
MSE	0.958957	1.273391	0.988870	0.716870	**0.392000**
RMSE	0.979263	1.128446	0.994419	0.846682	**0.626099**

There are a total of 8,627 samples/rows with the following distribution:

- Class 1: 1,735
- Class 2: 1,732
- Class 3: 1,693
- Class 4: 1,744
- Class 5: 1,726

B. Models

Figure 2 shows the three DNN models that were used for the investigation. The three models contain three, two and four hidden layers with a softmax function at the output.

C. Results

The three DNN models were trained first. Once this was done, then the results from these models were applied to obtain the ensemble model by average ranking. In a similar fashion, the Choquet integral fusion was applied using the three DNN models together with the learned densities to build the classification model.

Fig. 3 shows the confusion matrices obtained from the three DNN models. Model DNN2 seems to be best followed by DNN1, and DNN3 scores worst. Table I shows the results in form of AUC and accuracy. Only comparing the three DNN models we can see that DNN2 has the highest AUC score with 42.07%. In terms of accuracy though DNN3 outperforms the other two models. Surprisingly, the Ensemble method does not achieve better results, however, the Choquet method outperforms all by far with values of 50.72% and 89.57% for AUC and accuracy, respectively. Table II shows the MSE and RMSE results confirming the superior results of the Choquet method. The confusion matrices of the Ensemble method and the Choquet method are shown in Fig. 4.

Table III shows the results of applying standard machine learning techniques. The accuracy results are provided for the following approaches Support Vector Machines (SVM), Decision Tree (DT), Logistic regression (LR), Gaussian Nearest Neighbour (GNN), Random Forest Classifier (RFC), Extra Tree Classifier (ETC), and Gradient Boosting Classifier (GBDT). The best AUC result was achieved by the GNN algorithm with 59.59% followed by SVM and DTC with 57.08% and 56.01%, respectively. As for the accuracy, ETC scored best with 69.08% followed by RC and GBC with 97.86% and 62.47%, respectively. However, Choquet method

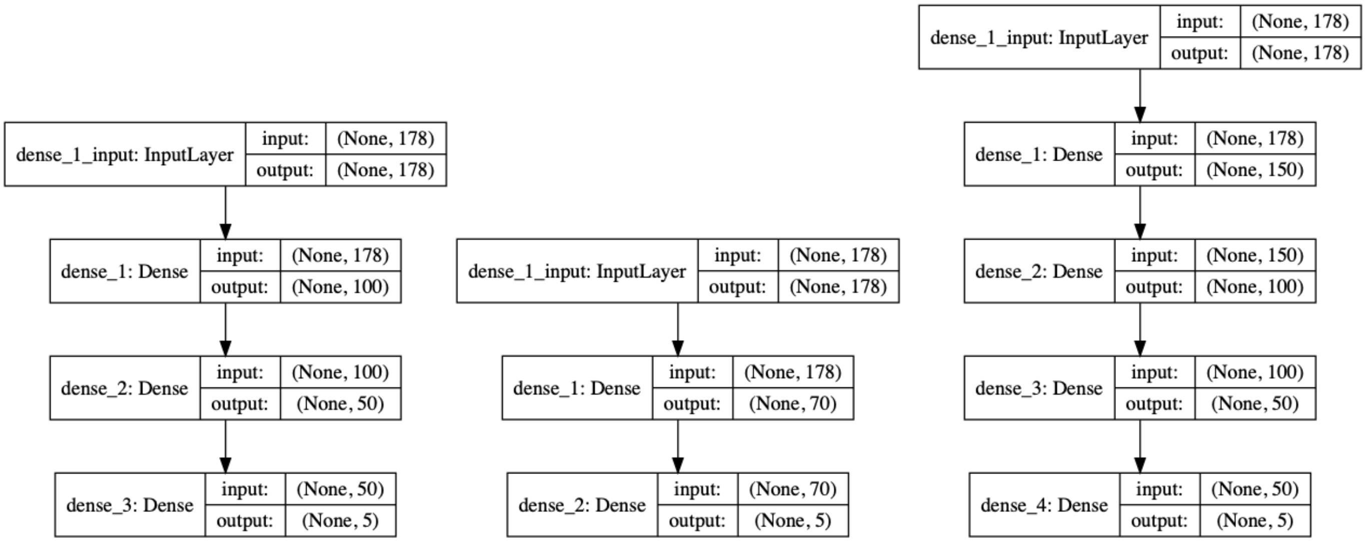

Fig. 2. Three DNN Models

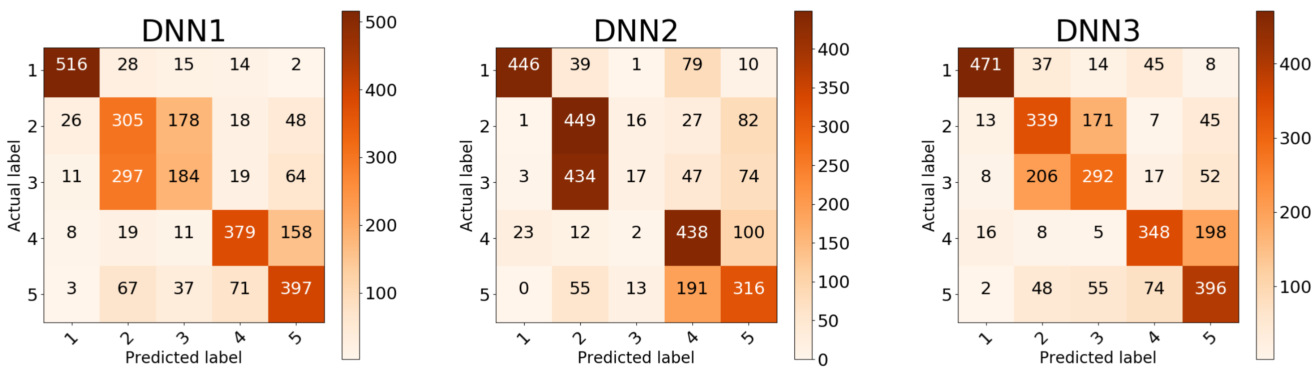

Fig. 3. Confusion Matrices of DNN Models

outperforms all the other approaches in terms of accuracy and provides the second best value in terms of AUC.

V. Conclusion

In the past, the Choquet fuzzy integral fusion method had shown very good performance applied to fused CNN model results that were trained on image data. This paper investigated the epileptic seizure recognition data set using a DNN ensemble method. This ensemble method was compared with the Choquet fuzzy integral fusion method. For this, three different DNN methods were trained with 3, 4, and 5 hidden layers. Afterwards, the ensemble method and the Choquet fusion method were applied. In terms of evaluation measures, confusion matrix, AUC, accuracy, MSE and RMSE were used.

The results show that the Choquet fuzzy integral fusion method outperforms the DNN ensemble method as well as other state-of-the-art classification methods. In particular, the AUC and accuracy results with values of 50.72% and 89.57%, respectively compared to 36.93% and 63.93% achieved by the ensemble method. Furthermore, a comparison with state-of-the-art classification algorithms showed that the Choquet fusion method is far superior to the best performing ML algorithm that achieved an accuracy of 69.08%. Unfortunately, no AUC value was provided with the accuracy result of the comparison approaches, and thus, further analysis is not possible.

References

[1] S.L. Mosh, E. Perucca, P. Ryvlin, T. Tomson. Epilepsy: New advances. Lancet. 385:884-898, 2015. doi: 10.1016/S0140-6736(14)60456-6.

[2] F. Leijten. Multimodal seizure detection: A review. Epilepsia. 59:42-47, 2018. doi: 10.1111/epi.14047.

[3] D. Sukumaran, Y. Enyi, S. Sun, A. Basu, D. Zhao, J. Dauwels. A low-power, reconfigurable smart sensor system for EEG acquisition and classification; Proceedings of the 2012 IEEE Asia Pacific Conference on Circuits and Systems; Kaohsiung, Taiwan, pp. 9-12, 2-5 December 2012.

[4] T.M.E. Nijsen, R.M. Aarts, P.J.M. Cluitmans, P.A.M. Griep. Time-Frequency Analysis of Accelerometry Data for Detection of Myoclonic Seizures. IEEE Trans. Inf. Technol. Biomed. 14:1197-1203, 2010. doi: 10.1109/TITB.2010.2058123.

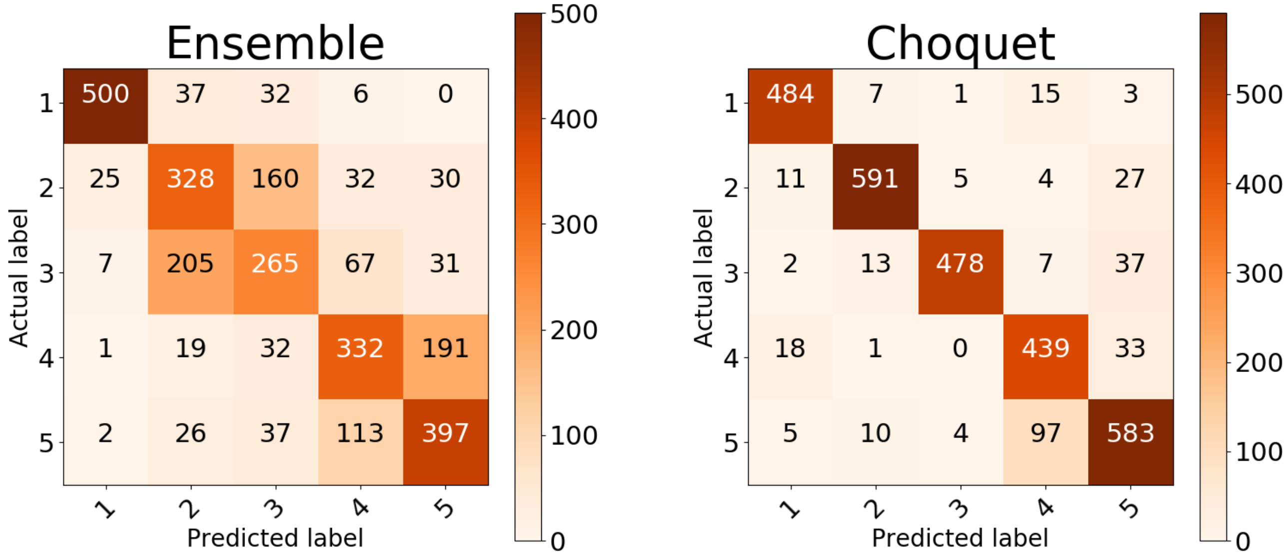

Fig. 4. Confusion Matrices of Ensemble and Choquet Methods

TABLE III
MACHINE LEARNING RESULTS [27]

	SVM	DTC	LR	GNN	KNC	RFC	ETC	GBC
AUC	57.08	56.01	51.56	**59.59**	44.45	39.84	38.87	44.04
Accuracy	54.85	33.88	22.47	43.06	47.41	67.86	**69.08**	62.47

[5] H. Chen, M. Xue, Z. Mei, S.B. Oetomo, W. Chen. A Review of Wearable Sensor Systems for Monitoring Body Movements of Neonates. Sensors. 16:2134, 2016. doi: 10.3390/s16122134.

[6] J.L. Andreassi. Psychophysiology: Human behavior and physiological response (Fifth ed.). Mahwah, NJ: Lawrence Erlbaum Associates, 2006.

[7] J. Coosemans, B. Hermans, R. Puers. Integrating wireless ECG monitoring in textiles; Proceedings of the International Conference on Solid-State Sensors, Actuators and Microsystems; pp. 228-232. Digest of Technical Papers, Transducers, Seoul, Korea. 5-9 June 2005.

[8] B. Yang, C. Yu, Y. Dong. Capacitively Coupled Electrocardiogram Measuring System and Noise Reduction by Singular Spectrum Analysis. IEEE Sens. J. 16:3802-3810, 2016. doi: 10.1109/JSEN.2016.2532599.

[9] H.C. Jung, J.H. Moon, D.H. Baek, J.H. Lee, Y.Y. Choi, J.S. Hong, S.H. Lee. CNT/PDMS composite flexible dry electrodes for long-term ECG monitoring. IEEE Trans. Biomed. Eng. 59:1472-1479, 2012. doi: 10.1109/TBME.2012.2190288.

[10] M.A. Yokus, J.S. Jur. Fabric-Based Wearable Dry Electrodes for Body Surface Biopotential Recording. IEEE Trans. Bio-Med. Eng. 63:423-430, 2016. doi: 10.1109/TBME.2015.2462312.

[11] R. Lei, Q. Jiang, K. Chen, Z. Chen, C. Pan, L. Jiang. Fabrication of a Micro-Needle Array Electrode by Thermal Drawing for Bio-Signals Monitoring. Sensors. 16:908, 2016.

[12] S.S. Spencer. MRI, SPECT, and PET imaging in epilepsy: Their relative contributions. Epilepsia. 35:S72-S89, 1994. doi: 10.1111/j.1528-1157.1994.tb05990.x.

[13] C.. Szab, L.C. Morgan, K.M. Karkar, L.D. Leary, O.V. Lie, M. Girouard, J.E. Cavazos. Electromyography-based seizure detector: Preliminary results comparing a generalized tonic-clonic seizure detection algorithm to video-EEG recordings. Epilepsia. 56:1432-1437, 2015. doi: 10.1111/epi.13083.

[14] Y. Gu, E. Cleeren, J. Dan, K. Claes, W.V. Paesschen, S.V. Huffel, B. Hunyadi. Comparison between Scalp EEG and Behind-the-Ear EEG for Development of a Wearable Seizure Detection System for Patients with Focal Epilepsy. Sensors. 18:29, 2018. doi: 10.3390/s18010029.

[15] Y. Wang, Z. Li, L. Feng, H. Bai, C. Wang. Hardware design of multiclass SVM classification for epilepsy and epileptic seizure detection. IET Circuits Devices Syst. 12:108-115, 2018. doi: 10.1049/iet-cds.2017.0216.

[16] Q. He, B. Wu, H. Wang, L. Zhu. VEP Feature Extraction and Classification for Brain-Computer Interface; Proceedings of the 8th International Conference on Signal Processing; Guilin, China. 16-20 November 2006.

[17] B.E. Boser, I.M. Guyon, V.N. Vapnik. A Training Algorithm for Optimal Margin Classifiers; Proceedings of the Annual Workshop on Computational Learning Theory; Pittsburgh, PA, USA, pp. 144-152. 27-29 July 1992.

[18] K. Fu, J. Qu, Y. Chai, T. Zou. Hilbert marginal spectrum analysis for automatic seizure detection in EEG signals. Biomed. Signal Process. Control. 18:179-185, 2015. doi: 10.1016/j.bspc.2015.01.002.

[19] K.D. Brabanter, P. Karsmakers, F. Ojeda, C. Alzate, J.D. Brabanter, K. Pelckmans, B.D. Moor, J. Vandewalle, J.A.K. Suykens. LS-SVMlab Toolbox Users Guide: Version 1.7. Ku Leuven Leuven; Leuven, Belgium: 2010.

[20] Y.C. Liu, C.C.K. Lin, T. Jing-Jane, Y.N. Sun. Model-Based Spike Detection of Epileptic EEG Data. Sensors. 13:12536-12547, 2013. doi: 10.3390/s130912536.

[21] D.H. Wolpert, W.G. Macready, No Free Lunch Theorems for Optimization, IEEE Transactions on Evolutionary Computation 1, 67, 1997.

[22] R.M. Isa, I. Pasya, M.N. Taib, A.H. Jahidin, W.R.W. Omar, N. Fuad, H. Norhazman. EEG brainwave behaviour due to RF Exposure using kNN classification; Proceedings of the IEEE International Conference on System Engineering and Technology; Shah Alam, Malaysia. 19-20 August, pp. 385-388, 2013.

[23] R. Chai, Y. Tran, G.R. Naik, T.N. Nguyen, S.H. Ling, A. Craig, H.T. Nguyen. Classification of EEG based-mental fatigue using principal component analysis and Bayesian neural network; Proceedings of the

Engineering in Medicine and Biology Society; p. 4654, Orlando, FL, USA, 16-20 August 2016.

[24] I. Kiral-Kornek, S. Roy, E. Nurse, B. Mashford, P. Karoly, T. Carroll, D. Payne, S. Saha, S. Baldassano, T. O'Brien, D. Grayden. Epileptic seizure prediction using big data and deep learning: toward a mobile system. EBioMedicine, 27, pp.103-111. Vancouver, 2018.

[25] A. Antoniades, L. Spyrou, C.C. Took, S. Sanei. Deep learning for epileptic intracranial EEG data. In 2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP) (pp. 1-6). IEEE, 2016.

[26] P. Thodoroff, J. Pineau, A. Lim. Learning robust features using deep learning for automatic seizure detection. In Machine learning for healthcare conference (pp. 178-190), 2016.

[27] H. Liu, L. Xi, Y. Zhao, Z. Li. Using Deep Learning and Machine Learning to Detect Epileptic Seizure with Electroencephalography (EEG) Data, arXiv, eprint:1910.02544, 2019.

[28] S. Aiello et al. Machine Learning with Python and H20. H2O. ai Inc., 2016.

[29] B.V. Dasarathy and B.V. Sheela, Composite classifier system design: concepts and methodology, Proceedings of the IEEE, vol. 67, no. 5, pp. 708-713, 1979.

[30] L.K. Hansen and P. Salamon, Neural network ensembles, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, no. 10, pp. 993-1001, 1990.

[31] R.E. Schapire, The Strength of Weak Learnability, Machine Learning, vol. 5, no. 2, pp. 197-227, 1990.

[32] G. Seni, J.F. Elder. Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions. In Grossman, R., ed.: Synthesis Lectures on Data Mining and Knowledge Discovery. Morgan & Claypool, 2010.

[33] T. Chen, C. Guestrin. XGBoost: A Scalable Tree Boosting System. In Krishnapuram, Balaji; Shah, Mohak; Smola, Alexander J.; Aggarwal, Charu C.; Shen, Dou; Rastogi, Rajeev (eds.). Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016. ACM. pp. 785-794, 2016.

[34] J. Xie, V. Rojkova, S. Pal, S. Coggeshall. A Combination of Boosting and Bagging for KDD Cup 2009 - Fast Scoring on a Large Database. The Journal of Machine Learning Research (JMLR) 7, 35-43, 2009.

[35] M. Sugeno, Theory of fuzzy integrals and its applications, Ph.D. thesis, Tokyo Institute of Technology, 1974.

[36] G. Choquet, Theory of capacities, in Annales de linstitut Fourier, vol. 5. Institut Fourier, 1954, pp. 131-295.

[37] M. Grabisch, H.T. Nguyen, and E.A. Walker, Fundamentals of uncertainty calculi with applications to fuzzy inference. Springer Science & Business Media, 2013, vol. 30.

[38] D.T. Anderson, T.C. Havens, C. Wagner, J.M. Keller, M.F. Anderson, and D.J. Wescott, Extension of the fuzzy integral for general fuzzy set-valued information, IEEE Trans. Fuzzy Syst., vol. 22, no. 6, pp. 1625-1639, 2014.

[39] D.T. Anderson et al., Binary fuzzy measures and Choquet integration for multi-source fusion, International Conference on Military Technologies (ICMT), Brno, 2017, pp. 676-681, 2017.

[40] J.M. Keller, D. Liu, D.B. Fogel, [Fundamentals of Computational Intelligence], John Wiley & Sons, Inc., Hoboken, New Jersey, 2016.

[41] R.G. Andrzejak, K. Lehnertz, C. Rieke, F. Mormann, P. David, C.E. Elger. Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state, Phys. Rev. E, 64, 061907, 2001.