Epileptic Seizure Detection in EEG via Fusion of Multi-View Attention U-net Deep Neural Networks

C. Chatzichristos[1] ¶, J. Dan[1,2]¶, A. M. Narayanan[1]¶, N. Seeuws[1] ¶, K. Vandecasteele[1]¶, M. De Vos[1], A. Bertrand[1] and S. Van Huffel [1]

1. Dept. of Electrical Engineering (ESAT),
Stadius Center for Dynamical Systems, Signal Processing and Data Analytics (STADIUS),
KU Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium
2. Byteflies, Antwerp, Belgium
¶Contributed equally to this work
{cchatzic,jonathan.dan,abhijith,nick.seeuws,kaat.vandecasteele}@esat.kuleuven.be

Abstract— Electroencephalography (EEG) is an essential tool in clinical practice for the diagnosis and monitoring of people with epilepsy. Manual annotation of epileptic seizures is a time consuming process performed by expert neurologists. Hence, a procedure which automatically detects the seizures would be hugely beneficial for a fast and cost-effective diagnosis. Progress in machine learning techniques, especially deep learning methods, coupled with the availability of large public EEG seizure databases have enhanced the possibility, in the near future, of automatic seizure detection in EEG. We propose an epileptic seizure detection pipeline based on the fusion of multiple attention U-nets, each operating on a distinct view of the EEG data. The model uses a simple long short term memory (LSTM) network based fusion of the individual attention U-net outputs to detect seizures in EEG. The model developed outperformed the state-of-the-art models on the TUH EEG seizure dataset and achieved the first place in the Neureka™ 2020 Epilepsy Challenge. Even though the model in its current format is not suitable for clinical practice, the performance obtained within the time constraints of a challenge imparts confidence for further research and improvement.

I. Introduction

Electroencephalography (EEG) is an essential tool in clinical practice for the diagnosis of epilepsy, the hallmark of which are epileptic seizures. Manual annotation and interpretation of long-term EEG recordings is a time-consuming and expensive task. Therefore, automated EEG-based epileptic seizure detection systems would be a valuable clinical support tool. The recognition of epileptic and non-epileptic EEG signals is a classification problem. It involves extraction of the discriminatory features from EEG signals followed by a classification step. Many EEG-based seizure detection methods have been developed over the years [1]. These methods have relied on classical machine learning techniques such as neural networks and support vector machines [2]. Deep learning (DL) methods, currently used to solve a plethora of other machine learning problems, have seen limited use in EEG-based seizure detection due to lack of large annotated datasets [3]. However, public availability of large EEG seizure datasets such as the TUH EEG corpus [4] has led to a recent development of DL methods for solving the seizure detection problem [5]. DL has shown great promise in EEG based classification due to its capacity to learn good feature representations from raw data. Currently, convolutional neural networks (CNNs) seem to be the most popular approach for automatic seizure detection [6–10]. This observation is in line with recent discussions and findings regarding the effectiveness of CNNs for processing time series. However, recurrent neural networks (RNN) such as Long Short Term Memory (LSTM) or Gated Recurrent Unit (GRU), which are powerful tools for modelling sequential signals, have also been used [5, 6, 11–13].

U-net is a DL architecture originally developed for image segmentation [14]. The U-net is a convolutional autoencoder with skip connections to recover the local spatial information lost during compression. This architecture has found recent applications in the analysis of biomedical signals, for example, in electrocardiograms (ECGs) for arrhythmia diagnosis [15] and in EEG for the identification of sleep stages [16]. In the current work, we used a modified U-net architecture called attention U-net [17]. To the best of our knowledge, attention U-nets have never been applied to EEG based seizure detection.

To improve seizure detection in EEG, we fused the predictions of three separate attention U-nets. These three individual attention U-nets operated on three distinct views of the EEG data. These individual views were distinguished by the filtering or pre-processing applied to the EEG data. Specifically, re-referenced and bandpass filtered EEG data, data filtered using a set of multi-channel spatio-temporal Wiener filters and EEG data processed using the ICLabel toolbox [18]. Hence, we term the proposed model as fusion of multi-view U-nets.

Most of the popular DL models used in seizure detection detect seizures in a sliding time window [6–8, 10]. In our proposed approach, a view of the entire EEG recording is used as an input to the attention U-nets, which outputs a probability of being a seizure for each EEG time sample. To produce interpretable predictions, we fuse the multi-view attention U-net outputs with a simple LSTM neural network.

We developed the proposed model for the Neureka™ 2020 Epilepsy Challenge [19], a month-long challenge on seizure detection using the TUH EEG Seizure dataset [20]. Our submission to the challenge, based on the proposed model, was awarded the first place among 15 submissions from all over the world [21].

The paper is organized as following. In Section II a description of the pre-processing of raw EEG, the multi-view attention U-nets and the post-processing steps of the seizure detection model, are given. We describe the training, validation and testing procedure used on the model in Section III. We report and draw inferences on the performance of the proposed model on TUH EEG seizure dataset in Section IV. Finally a summary of our work and future improvements of the model used are presented in Section V.

II. Methods

The Neureka™ 2020 Epilepsy Challenge made use of the Temple University Hospital Seizure Detection Corpus v1.5.1 [20].

Figure 1 gives on overview of the architecture of our seizure detection pipeline. The raw EEG data are first standardized to form the input data of our pipeline. This was done by means of re-referencing to a bipolar montage, re-sampling to 200Hz and filtering the data (see Section III). Different views of the data were created through different pre-pocessing methods (Section II-A). Three U-nets with attention layers were trained on these different views (Section II-B). These U-nets were combined by means of an LSTM network (Section II-C). Finally, after some post-processing of the predicted labels, a list of detected seizure events is provided (Section II-D).

II-A. Pre-processing

Three different pre-processing pipelines of the input data were used as different data views to train three different U-nets (Fig. 1):

- The input data (without any further processing) were used as a first view of the data.

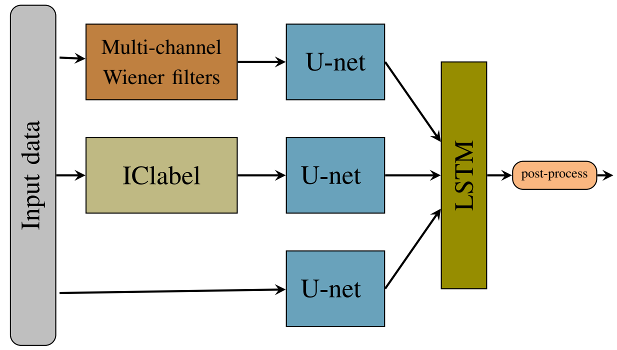

Figure 1. Seizure detection pipeline

- A set of multi-channel spatio-temporal Wiener filters that suppress the dominant artifacts were computed and applied to provide a second view of the data.
- The input data filtered with the use of ICLabel toolbox provided by EEGlab [18] was used as a third view.

II-A1. Multi-channel Wiener filter

Wiener filtering has proved an effective method for removing artifacts in EEG recordings [22–24]. In typical implementations of this method, a well identified type of artifact (e.g. eye blinks) is manually marked by an expert. A spatial or spatio-temporal co-variance matrix of the artifact is computed and the resulting Wiener filter is calculated. This is a semi-automated method as it first requires a manual labelling of paradigms of each type of artifact. We propose two main novelties with regards to established methods:

1) An automatic artifact identification followed by a an automatic clustering of these artifacts.
2) The use of filter-bank Wiener filters to remove the dominant artifacts identified in step 1.

II-A1a. Automatic artifact identification

Automatic artifact identification is done with the use of a spatio-temporal max Signal-to-Noise-Ratio filter (max-SNR) [25]. The max-SNR filter ($\mathbf{w}$) is a spatio-temporal filter that linearly combines the time lagged multi-channel input data ($\mathbf{y}(t)$) into a single-channel output ($o(t)$) : $o(t) = \mathbf{w}^T\mathbf{y}(t)$. The filter $\mathbf{w}$ is optimized in a data-driven fashion to maximize the SNR of $o(t)$ over a training set, where the target signal corresponds to the seizure epochs. A detailed description of this artifact selection procedure is described in [26].

II-A1b. Artifact clustering

The artifacts are clustered in groups based on the Euclidean distance between their spatio-temporal covariance matrices. The spatio-temporal covariance matrix of each individually identified artifact is computed. Dimensionality reduction, based on Principal Component Analysis (PCA), is applied to the covariance matrices, retaining 99% of the variance of the data. A K-means clustering is then applied on the compressed covariance matrices. This allows to group the artifacts in clusters of similar covariance matrices.

II-A1c. Wiener filtering

Spatio-temporal filters are computed based on the average covariance matrix of the two biggest clusters (which are considered the most representative clusters of the respective artifact). These two filters are pre-computed and do not take into consideration any statistics of the background EEG (i.e. background EEG is considered as white noise).

II-A2. IC Label pre-processing

Blind Source Separation (BSS) approaches for multichannel EEG processing have become popular, in view of their proven ability for artefact removal and source extraction. In particular Independent Component Analysis (ICA) makes use of different properties of the signal, such as non-Gaussianity, sample dependence, geometric properties, or non-stationarity in order to maximize the independence among the extracted components. In matrix-based BSS approaches the multichannel EEG signal forms a matrix, $\boldsymbol{T} \in \mathbb{R}^{I_t \times I_e}$ for which a decomposition is sought such that:

$$\boldsymbol{T} \approx \boldsymbol{M}\boldsymbol{A}^\top, \tag{1}$$

with $\boldsymbol{A} \in \mathbb{R}^{I_e \times R}$ containing the weights of the topographic maps and $\boldsymbol{M} \in \mathbb{R}^{I_t \times R}$ containing the time-courses. I_e repesents the total number of electrodes, I_t the total time in number of samples and R being the estimated number of sources [27]. Note that, in practice, the decomposition cannot be exact due to unmodeled phenomena including noise.

ICA solves Eq. (1) by assuming that the matrix $\boldsymbol{A}$ contains statistically independent topographic maps in its columns, each one corresponding to a time-course in the associated column of the (mixing) matrix $\boldsymbol{M}$. The, R, independent components (IC) obtained by ICA are manually inspected and interpreted in order to identify if they represent an artifact or source of interest; the artifact components are then removed and the signal is composed again from the remaining components.

Automated IC classifiers have been designed, speeding up the analysis of EEG studies with many subjects. IClabel [18, 28] is one of the most accurate automated classifiers available via EEGlab [29]. The IClabel classifier uses a fusion of convolutional neural networks of different depth for each of the feature-sets. Namely, the feature-sets included in the ICLabel dataset, are scalp topography images, channel-based scalp topography measures, power spectral densities (PSD) measures, plus features used in several published IC classifier approaches [18]. The ICLabel dataset, used for training comprised of spatio-temporal measures for over 200,000 ICs from more than 6000 EEG recordings, the biggest dataset with which such a classifier has ever been trained. The seven different clusters used for the ICs are: brain, muscle, eye, heart, line noise, channel noise, and other.

The input data were first high-passed filtered (0.25 to 0.75 Hz) and then possible "bad channels" were rejected. A channel was rejected either if it was flat for more than 20 seconds, or its SNR was lower than 0.25 standard deviations based on the total channel population, or if its Pearson correlation was less than 0.6 to an estimate based on other channels. The "cleaned" input data were decomposed with the Second Order Blind Identification (SOBI) [30] ICA algorithm. The resulted ICs were clustered based on IClabel probabilities. Any IC with probability higher than 0.6 for being in any of the following five classes, was disregarded: muscle, eye, heart, line noise, channel noise.

II-B. Attention-Gated U-net

A U-net neural network was used as the base learning algorithm in the seizure detection pipeline. The network architecture is based on [17], which was originally designed for medical imaging tasks.

II-B1. Architecture

The network processes multi-channel EEG data and outputs a single-channel signal indicating the likelihood of a seizure for each time sample. Figure 2 shows the architecture of the network. All convolutions operate along the temporal axis, i.e., all channels are processed in a parallel manner.

The "downward path" of the U-Net extracts information on different scales. It uses a maxpooling operation to down-sample the data. This down-sampling, similarly to the convolutions, works along the temporal axis. In the lowest part of the U-net the data channels are merged. This merging step performs maxpooling along the channel axis.

The "upward path" of the U-Net combines local and global information before outputting a signal. Skip connections in the upward path make use of Attention Gating [17]. Attention Gating assigns a weight to every feature fiber in the $(time \times channel \times feature)$ data tensor. These weights are used to compute a weighted average along the channel axis for data flowing from a skip connection. Every stage of the upward path concatenates this weighted average with up-sampled data from the lower stage. The entire "upward path" consists of single-channel data resulting in the final output signal of probabilities.

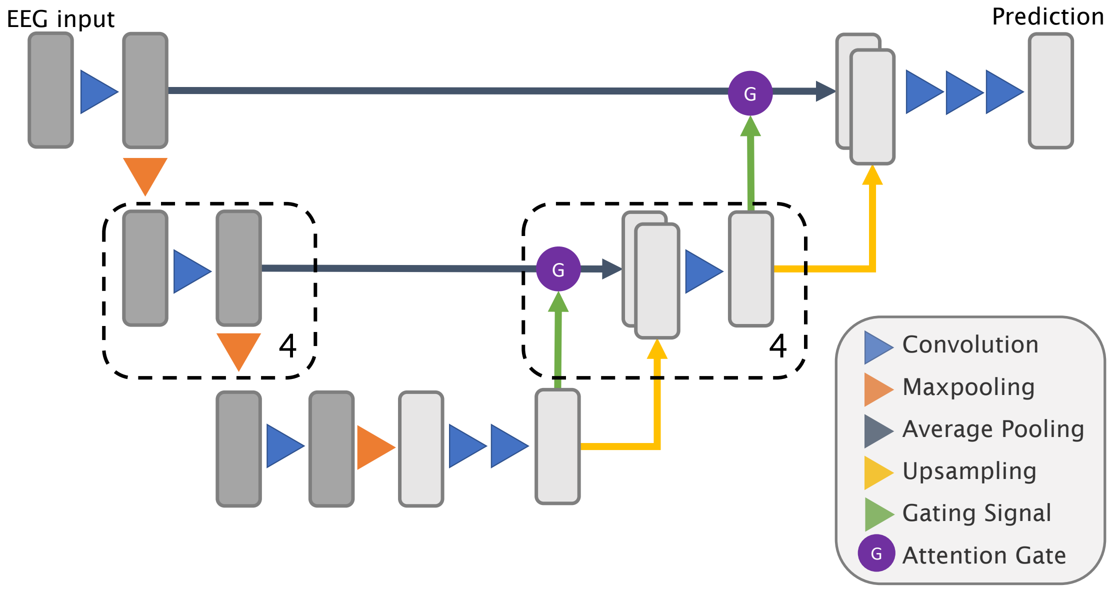

Figure 2. Illustration of the architecture of the U-nets used as base learners for the different data views. The "downward path" consists of five stages, down-sampling along the time axis at every stage. The last part performs max-pooling over the entire channel axis, reducing the data to a single channel. The "upward path" up-samples the data along the time axis in five stages, making use of attention-gated skip connections from the "downward path". The darker blocks represent multi-channel data and the lighter ones correspond to single-channel data.

II-B2. Attention Gating

The local information of channels are merged using the Attention Gating mechanism of [17], allowing the network to decide, at every time step, which channels should be focused on and which should be ignored. As explained above, Attention Gating calculates an attention weight, α, for a specific feature fiber, $\boldsymbol{x}$. The attention mechanism makes use of a gating signal, $\boldsymbol{g}$, another feature fiber originating from a lower stage in the upward path up-sampled to match the time resolution of the data flowing from the downward path. The attention weight α is calculated as follows:

$$\alpha = \sigma(\boldsymbol{w}^T \sigma(W_x \boldsymbol{x} + W_g \boldsymbol{g} + \boldsymbol{b}) + b, \qquad (2)$$

with $\boldsymbol{w}, W_x, W_g, \boldsymbol{b}$ and b all being trainable weights and $\sigma(\cdot)$ the element-wise sigmoid function. This calculation is based on the additive attention mechanism of [31], taking a form similar to a classical multilayer perceptron.

II-B3. Training

The training process makes use of multiple regularization methods to improve the generalization power of a single U-net. Firstly, regular training loss making use of the output of the network is extented by use of Deep Supervision [17]. This makes sure the network focuses on seizure information at every stage of the U-net instead of relying on one-off patterns it may find in the training set by accident. Secondly, label smoothing prevents over-confident predictions. The normal training labels, 0 or 1, are changed to values closer to 0.5. By doing so, the network should avoid saturating the sigmoid activation in its output and overfit less. Finally, a weighted cross-entropy loss is used to mitigate the effect of class imbalance (higher number of background EEG samples than seizure events).

II-C. U-net fusion

Each trained U-net predicts the likelihood of each time sample being part of a seizure. To include long-term memory and information on the probability of transitioning between a seizure and non-seizure state, a RNN is used to combine the different U-Nets. The RNN is implemented as a bidirectional LSTM node with a state vector of length 4 followed by a dense layer. The LSTM receives as an input a downsampled version of the U-net predictions at 1Hz and provides predictions at the same sampling rate.

II-D. Post-processing

The proposed seizure detection model designed for the Neureka challenge [19] was tuned with the objective to maximize a scoring function described in Section IV. To this end, the following set of post-processing rules was used to merge neighbouring events and remove short events:

1) Seizure events less than 30 seconds apart are merged together.
2) Merged seizure events, for which the proba-

bility of being a seizure is less than 82%, are rejected. The probability per event is calculated as the mean of all the output probabilities during that seizure event, normalized by the mean probability of the event with the highest probability.

3) Seizure events of duration less than 15 seconds are rejected.

III. Experiments

III-A. Dataset

We developed the seizure detection model for the Neureka ™ Challenge [19]. This dataset consists of routine EEG recordings performed on 692 patients. The seizures in the recordings are annotated by experts. The dataset contains more than 3500 seizures. The recordings in the dataset come in different EEG montages and with different sampling frequencies. A subset of this dataset, was used in the challenge [20].

The subset which was made available for the challenge was divided into a training, development and evaluation set. EEG recordings along with seizure annotations were made available for the training and development (validation) set but not for the evaluation set. The training set consisted of 4597 files of EEG recordings, with a total duration of approximately 752 hours, which contained 46.7 hours of seizure data (6.21%). The development set consisted of 1013 files of EEG recordings, with a total duration of approximately 170 hours, which contained 16.2 hours of seizure data (9.53%) [19]. For the evaluation set, only the final performance scores of the submitted detection models were made available.

III-B. Model training and validation

The recording montages, sampling frequencies and number of channels were not uniform across the recordings in the EEG dataset. Therefore, for uniformity, the following pre-processing steps were applied. First, only a subset of channels available in all the recordings were used in the development of the seizure detection algorithm. These channels were the following 16 channels: FP1, F7, T3, T5, O1, FP2, F8, T4, T6, O2, C3, CZ, C4, F3, P3, F4. The EEG measurements from these channels were re-referenced to 18 bipolar pairs from a double banana montage. Second, the re-referenced EEG was resampled to 200 Hz. The resampled EEG was high-pass filtered with a fourth order Butterworth filter with a cut-off frequency of 1 Hz and with two band-stop fourth order Butterworth filters with stop bands of respectively $[47.5, 52.5]$ Hz and $[57.5, 62.5]$ Hz.

For each of the three data views (Raw EEG, Wiener filtering and IClabel pre-processing) the network specifics are kept identical. For the wiener filters the number of time lags was set to $L = 50$ and 40 minutes of artifact are selected per 24h. The U-nets are trained on data with 4096 time samples, or about 20 seconds of EEG data. In every stage of the “downward path”, the network performs down-sampling along the time axis with a factor of 4. In total, the U-net contains five such down-sampling steps, resulting in 4 "time steps" at the lowest level of the U-net. The “upward path” up-samples data by a factor of 4 in five stages, similarly to the “downward path”. Training makes use of the Adam optimizer [32] with 0.0001 as learning rate. Training was stopped by early stopping monitoring the performance of the network on a separate validation set.

We used a bidirectional LSTM with 4 hidden nodes followed by a dense layer to fuse the attention U-Net outputs. The LSTM based network was trained and tested on the development set using 10−fold cross validation. The LSTM state vector was reset for each recording. EEG recordings which contained seizures were given a larger weight during LSTM training by using 15 times more epochs from these recordings compared to those not containing any seizures.

IV. Results and Discussion

The evaluation of the submissions in the Neureka™ Epilepsy challenge was based on the Time-Aligned Event Scoring (TAES) metric. The TAES metric weighs each seizure event predicted by a model equally. For each event a partial score based on its overlap with a true seizure [33] is assigned. The metric is designed to be a compromise between the fraction of a true seizure events correctly detected as well as the number of false detections (background EEG detected as a seizure event). Using the TAES metric, true positives (TPs) and false alarms (FAs) are calculated. The following formula was used to compute the “TAES score”:

$$\text{TAES score} = Sens - \alpha * FAs_{24hr} - \beta * \frac{N}{19}, \quad (3)$$

where $Sens$ is the sensitivity in %, FAs_{24hr} is the number of FAs per 24 hours, N is the mean number of EEG channels used for seizure detection and $\alpha = 2.5$ and $\beta = 7.5$ are constants defined by the challenge organizers.

In Fig. 3, the sensitivity is plotted as a function of FA rate. As expected, there is a compromise between sensitivity and FA rate, based on the threshold (τ) set on the predicted probabilities (a segment with probability higher than τ is considered a seizure). In order to select the optimal τ value we computed the TAES score (on the validation set) for different threshold values Fig. 4 (using a 10−fold cross-validation). The bold line indicates the median score and the edges of the shaded area represent the maximum and minimum scores across 10 folds. The optimal probability threshold, to be used for the evaluation set, was selected based on this figure.

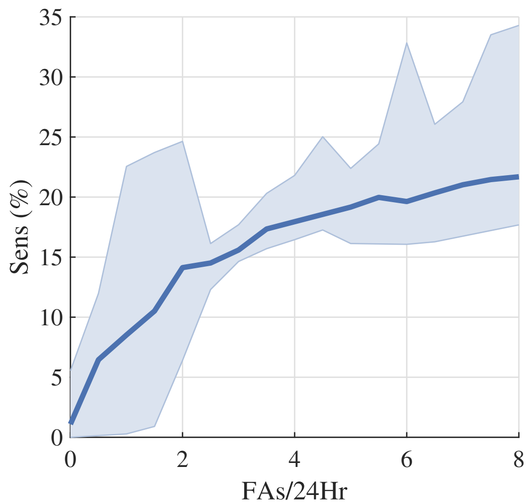

Figure 3. Development set performance: Sensitivity in function of FA rate using a 10-fold cross-validation. The bold line indicates median sensitivity and the edges of the shaded area represent minimum and maximum sensitivity across the 10 folds.

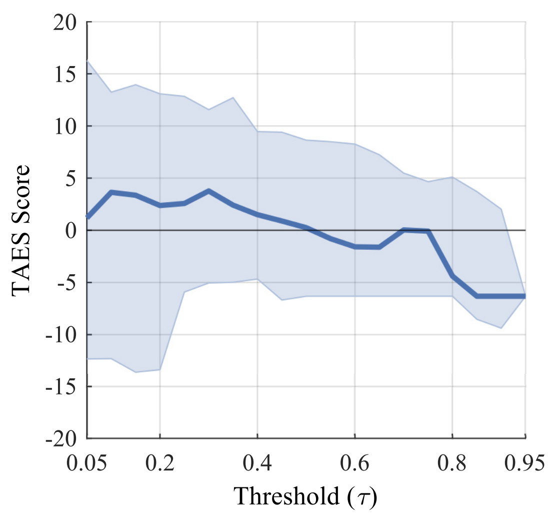

Figure 4. Development set performance: TAES score obtained for different thresholds on the development set using a 10-fold cross-validation. The bold line indicates median score and the edges of the shaded area represent minimum and the maximum score across the 10 folds.

After selecting a threshold [1], we applied the proposed model on the evaluation set to detect seizures. The results were submitted for the challenge, in the name of team Biomed Irregulars. The scores obtained in the challenge by the top 5 teams, from 15 worldwide submissions, is tabled in Table 1. As can be noted, our submission achieved the top position, by a considerable margin.

[1]We would like to point out that for the model used to generate the submission for the challenge a threshold of 0.55 was used. Due to time constraints, the threshold was selected after training and testing of the LSTM on the development set without the use of 10-fold cross validation. A threshold equal to 0.35 would have probably resulted in an even better performance, as can be noted from Fig. 4.

Rank	Team	Sensitivity	FAs/24Hr	Channels	Score
1	**Biomed Irregulars**	12.37	1.44	16	**2.46**
2	NeuroSyd	2.04	0.17	2	0.82
3	USTC-EEG	8.93	0.71	17	0.45
4	RocketShoes	5.98	3.36	3	-3.6
5	Lan Wei	20.00	15.59	4	-20.56

Table 1. Neureka challenge results: The top 5 teams and the performance scores of their submissions on the test set of the Neureka challenge.

While our seizure detection algorithm showed better results than competition in the Neureka challenge, it still needs to be better optimized for use in clinical practice. Sensitivity remains below 25% even for high FAs/24hr. This is partially due to the nature of the dataset which contains seizures of different type, some of them challenging to identify. The architecture and training of the model was the same for all types of seizures. The use of a different model for every subset of seizure type may result in better performance, which should be examined as future work. Furthermore, the impact of each step of our detection pipeline must be studied and documented. In the view of the success of the multi-view approach, alternative views (other than different pre-processing methods) will also be examined.

V. Conclusion

Automatic seizure detection is highly beneficial for the quick and efficient diagnosis of patients. The availability of large public EEG seizure databases have enhanced the possibility for developing DL approaches. We propose an epileptic seizure detection model based on the fusion of multiple attention U-nets, each operating on a distinct view of the EEG data. The outputs of the U-nets were combined with an LSTM. This model achieved the highest performance in the Neureka challenge competition. However, it still remains insufficient for use in clinical practice.

Acknowledgements

We would like to thank Novela Neurotech and NeuroTechX for conducting the Neureka™ 2020 Epilepsy Challenge. We also thank Temple University for providing the TUH EEG Corpus.
The authors acknowledge the financial support of the KU Leuven Research Council for project C14/16/057, FWO (Research Foundation Flanders) for projects G.0D75.16N and G.0A49.18N. The researchers have also received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 802895 and No 766456). This research received funding from the Flemish Government (AI Re-

search Program), Bijzonder Onderzoeksfonds (BOF) KU Leuven (Prevalence of Epilepsy and Sleep Disturbances in Alzheimer Disease) (C24/18/097), EIT Health: 19263 SeizeIT2 (Discreet Personalized Epileptic Seizure Detection Device).

References

[1] C. Baumgartner and J. P. Koren, "Seizure detection using scalp-EEG," *Epilepsia*, vol. 59, pp. 14–22, Jun. 2018.

[2] T. N. Alotaiby, S. A. Alshebeili, T. Alshawi, I. Ahmad, and E. A. Fathi, "EEG seizure detection and prediction algorithms: a survey," *EURASIP Journal on Advances in Signal Processing*, vol. 2014, no. 1, Dec. 2014.

[3] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: a systematic review," *J. Neural Eng.*, pp. 1–37, Aug. 2019.

[4] I. Obeid and J. Picone, "The temple university hospital EEG data corpus," *Frontiers in neuroscience*, vol. 10, p. 196, 2016.

[5] M. Golmohammadi, V. Shah, I. Obeid, and J. Picone, "Deep learning approaches for automated seizure detection from scalp electroencephalograms," *Signal Processing in Medicine and Biology*. Springer, 2020, pp. 235–276.

[6] P. Thodoroff, J. Pineau, and A. Lim, "Learning Robust Features using Deep Learning for Automatic Seizure Detection," *arXiv:1608.00220*, Jul. 2016.

[7] M. Zhou, C. Tian, R. Cao, B. Wang, Y. Niu, T. Hu, H. Guo, and J. Xiang, "Epileptic Seizure Detection Based on EEG Signals and CNN," *Frontiers in Neuroinformatics*, vol. 12, pp. 95–109, Dec. 2018.

[8] I. Ullah, M. Hussain, E.-u.-H. Qazi, and H. Aboalsamh, "An automated system for epilepsy detection using EEG brain signals based on deep learning approach," *Expert Systems App.*, vol. 107, pp. 61–71, Oct. 2018.

[9] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," *Computers Biology Medicine*, vol. 100, pp. 270–278, Sep. 2018.

[10] A. H. Ansari, P. J. Cherian, A. Caicedo, G. Naulaers, M. De Vos, and S. Van Huffel, "Neonatal Seizure Detection Using Deep Convolutional Neural Networks," *Int. J. Neural Systems*, vol. 29, no. 04, May 2019.

[11] A. Petrosian, D. Prokhorov, R. Homan, R. Dasheiff, and D. Wunsch, "Recurrent neural network based prediction of epileptic seizures in intra- and extracranial EEG," *Neurocomputing*, vol. 30, no. 1-4, pp. 201–218, Jan. 2000.

[12] M. Golmohammadi, S. Ziyabari, V. Shah, E. Von Weltin, C. Campbell, I. Obeid, and J. Picone, "Gated recurrent networks for seizure detection," *IEEE Signal Process. Medicine Biology Symposium (SPMB)*, Philadelphia, USA, Dec. 2017.

[13] R. Hussein, H. Palangi, R. K. Ward, and Z. J. Wang, "Optimized deep neural network architecture for robust detection of epileptic seizures using EEG signals," *Clinical Neurophysiology*, vol. 130, no. 1, pp. 25–37, Jan. 2019.

[14] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Int. Conf. Medical Image computing computer-assisted intervention (MICCAI)*, Munich, Germany, Nov. 2015.

[15] S. L. Oh, E. Y. Ng, R. San Tan, and U. R. Acharya, "Automated beat-wise arrhythmia diagnosis using modified u-net on extended electrocardiographic recordings with heterogeneous arrhythmia types," *Computers Biology Medicine*, vol. 105, pp. 92–101, Jan 2019.

[16] M. Perslev, M. Jensen, S. Darkner, P. J. Jennum, and C. Igel, "U-time: A fully convolutional network for time series segmentation applied to sleep staging," pp. 4417–4428, 2019.

[17] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, and D. Rueckert, "Attention gated networks: Learning to leverage salient regions in medical images," *Medical Image Analysis*, vol. 53, pp. 197–207, Apr. 2019.

[18] L. Pion-Tonachini, K. Kreutz-Delgado, and S. Makeig, "ICLabel: an automated electroencephalographic independent component classifier, dataset, and website," *NeuroImage*, vol. 198, pp. 181–197, Sep. 2019.

[19] N. Neurotech and NeuroTechX. Neureka™ 2020 epilepsy challenge. (available at: *https://neureka-challenge.com/*).

[20] V. Shah, E. von Weltin, S. Lopez, J. R. McHugh, L. Veloso, M. Golmohammadi, I. Obeid, and J. Picone, "The Temple University Hospital Seizure Detection Corpus," *Frontiers Neuroinformatics*, vol. 12, pp. 83–89, Nov. 2018.

[21] N. Neurotech and NeuroTechX. Neureka™ 2020 epilepsy challenge. (available at: *https://neureka-challenge.com/results*).

[22] H. Maki, T. Toda, S. Sakti, G. Neubig, and S. Nakamura, "EEG signal enhancement using multi-channel wiener filter with a spatial correlation prior," *IEEE Int. Conf. Acoustics, Speech Signal Process. (ICASSP)*, S. Brisbanne, Australia, Aug. 2015, pp. 2639–2643.

[23] B. Somers, T. Francart, and A. Bertrand, "A generic EEG artifact removal algorithm based on the multi-channel Wiener filter," *J. Neural Engin.*, vol. 15, no. 3, Jun. 2018.

[24] A. Borowicz, "Using a multichannel Wiener filter to remove eye-blink artifacts from EEG data," *Biomedical Signal Processing and Control*, vol. 45, pp. 246–255, Aug. 2018.

[25] B. D. Van Veen and K. M. Buckley, "Beamforming: A Versatile Approach to Spatial Filtering," *IEEE ASSP Magazine*, vol. 5, no. 2, pp. 4–24, Apr. 1988.

[26] J. Dan, B. Vandendriessche, W. Van Paesschen, D. Weckhuysen, and A. Bertrand, "Computationally-efficient algorithm for real-time absence seizure detection in wearable electroencephalography," *Int. J. Neural Systems*, To appear.

[27] S. Sanei and J. A. Chambers, *EEG Signal Processing*. John Wiley & Sons Ltd, 2007.

[28] L. Pion-Tonachini, K. Kreutz-Delgado, and S. Makeig, "The ICLabel dataset of electroencephalographic (EEG) independent component (IC) features," *Data in Brief*, vol. 25, Jun. 2019.

[29] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neuroscience Methods*, vol. 134, pp. 9–21, Mar. 2004.

[30] A. Belouchrani, K. Abed-Meraim, J. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Proc.*, vol. 45, no. 2, pp. 434–444, Feb. 1997.

[31] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[32] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[33] S. Ziyabari, V. Shah, M. Golmohammadi, I. Obeid, and J. Picone, "Objective evaluation metrics for automatic classification of eeg events," *arXiv preprint arXiv:1712.10107*, 2017.