2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Towards Deeper Neural Networks for Neonatal Seizure Detection

Aengus Daly[1,2,3], Alison O'Shea[3], Gordon Lightbody[1,2] and Andriy Temko[1]

Abstract — **Machine learning and more recently deep learning have become valuable tools in clinical decision making for neonatal seizure detection. This work proposes a deep neural network architecture which is capable of extracting information from long segments of EEG. Residual connections as well as data augmentation and a more robust optimizer are efficiently exploited to train a deeper architecture with an increased receptive field and longer EEG input. The proposed system is tested on a large clinical dataset of 4,570 hours of duration and benchmarked on a publicly available Helsinki dataset of 112 hours duration. The performance has improved from an AUC of 95.41% to an AUC of 97.73% when compared to a deep learning baseline.**

Clinical Relevance — **This research presents an improvement in performance over state-of-art neonatal seizure detection algorithms.**

I. Introduction

Neonatal seizure detection is a challenging task for clinicians as many neonates do not display any physiological indicators during seizure events [1]. The only way to detect all seizure events is to examine neonatal brainwaves – EEG. Interpretation of neonatal EEG (Fig. 1) is a challenging task, seizure events can be short and focal while multiple seizure-like artifacts are present during continuous EEG monitoring. Detection of seizures in neonatal EEG requires years of clinical training and this expertise is not available 24/7 in intensive care units. Machine learning techniques have been developed to automatically monitor EEG signal and trigger an alarm when a seizure is detected. Such automatic decision support tools can aid clinicians in detecting seizures [2], [3]. To achieve this, the EEG signal is usually characterised with a number of hand-engineered features which are pooled together by a classifier to make inferences [2], [4]. More recently deep learning based seizure detection algorithms have improved the algorithmic performance [5-9]. Deep learning methods exploit increasing amounts of data to learn signal representation patterns directly from raw EEG in an end-to-end optimisation paradigm without making assumptions about the signal.

A fully convolutional architecture for neonatal seizure detection was developed in [8]. The authors utilised an 8 second window of multichannel EEG as an input to the model thus removing the need for explicit feature extraction. The predictions of the model were smoothed to output a probability of seizure for every minute-long window. While the state-of-the-art results in detecting seizure events were achieved with this architecture, the postprocessing steps were still outside of the deep learning optimisation routine. Moreover, it was shown that increasing the input length and the receptive field by making an architecture deeper with more feature extraction blocks did not further improve the performance.

This study advances towards incorporating the postprocessing steps into the end-to-end optimisation by making the model operate on longer EEG segments. Resnet/skip connections [10] are incorporated into the previously developed system. Together with the proposed data augmentation and a more robust optimizer this allows for the increase of the input window to 16s and thus improve the performance.

II. Materials and Methods

A. Baseline model

A fully convolutional neural network is used as a baseline in this study [2], [11]. This architecture is shown in Fig. 2. The system takes an 8s window of 256 Hz multichannel EEG as input, down-samples it to 32 Hz using a band-pass filter of between 0.5Hz and 12.8Hz, transforms it through 3 feature extraction blocks, followed by a fully convolutional classification block. The use of a stacked convolutional layers with small sample-sized filters (3-sample wide) is similar to the VGG architecture, which was developed for image processing applications [12]. A one second shift is used between consecutive 8s window inputs.

Though the input to the model is two-dimensional, with stacked channels of temporal EEG, all convolutional and

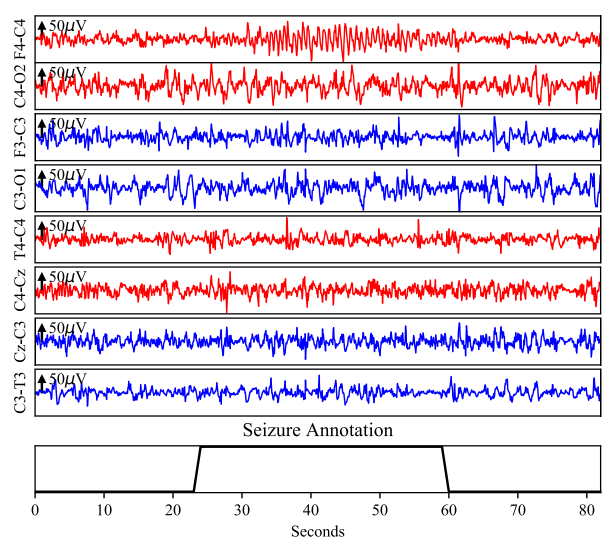

Figure 1. A segment of 80s of multichannel neonatal EEG with a seizure. Source: Adapted from [8].

[1] Department of Electrical and Electronic Engineering, University College Cork, Ireland. (corresponding author's e-mail: aengus.daly@ cit.ie)

[2] Irish Centre for Fetal and Neonatal Translational Research (INFANT), University College Cork, Ireland.
[3] Munster Technological University, Cork, Ireland.

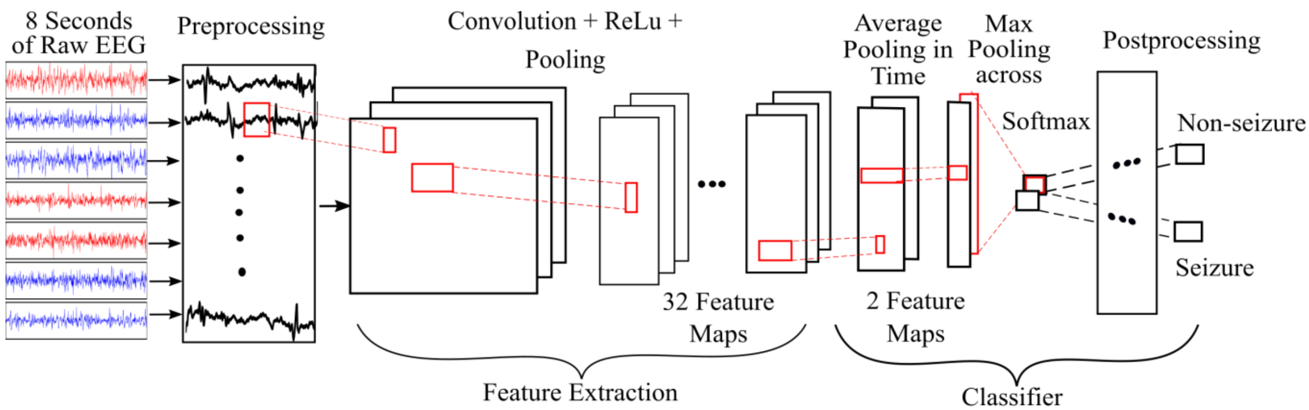

Figure 2. Schematic of the baseline model. Eight seconds of raw EEG from 8 channels are preprocessed and fed into the NN model. The probabilistic output is postprocessed. Source: Adapted from [8].

pooling operations are applied along the temporal dimension. The trained model is agnostic to the number of input EEG channels and to the location of the seizures. The classification block consists of a convolutional layer with two feature maps, representing seizure and non-seizure, followed by global average pooling operation across time and global maximum pooling across channels, to summarise the presence of seizure activity in the given 8s segment, in any of the input EEG channels.

Three postprocessing steps reduce the number of false alarms as developed in [13]. First a moving average filter of ~ 1 minute duration is applied, followed by a per patient adjustment for the level of background EEG by averaging the previous ~ 10 minutes of non-seizure EEG, finally detected seizures are extended on both sides by 30 seconds [13].

B. *The proposed approach with residual connections*

The architecture of the proposed approach is shown in Fig. 3. The input to the model is a 16s window of multichannel EEG. The main difference with the baseline is the integration of residual connections into the architecture. These skip connections allow each stack of 3 convolutional layers to learn a residual mapping – the difference between the desired underlying mapping for the stack and the input to the stack [10]. The residual functions enable the whole network to learn more efficiently with only a marginal increase in computational cost and no increase in the number of network parameters. Thus, a deeper model with an extra feature extraction block (13 convolutional layers in total) is trained.

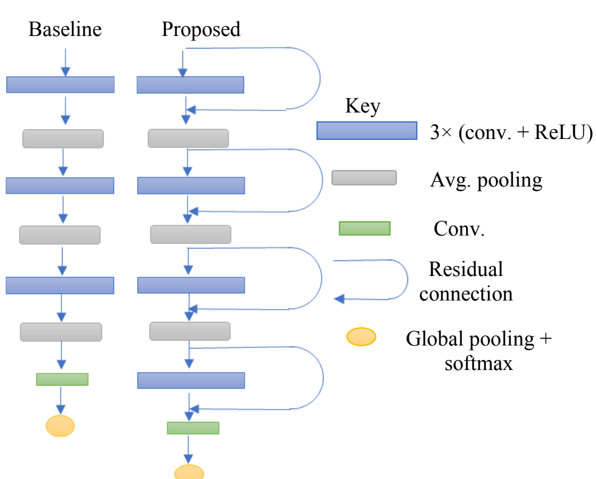

Figure 3. The baseline and the proposed models. The baseline consists of 3 feature extraction blocks. An extra feature extraction block is added to the proposed model along with the residual connections.

Data augmentation was utilised where amplitude rescaling was dynamically applied during training; the EEG signal was transformed at random by one of four transformations with equal probability. These four transformations were a vertical flip, increase the signal amplitude by a random number between 0.5 and 1.5, a combination of the previous 2 transformations, or no alteration.

A Rectified Adam (RAdam) optimizer [14] was used in training for the proposed model, compared to stochastic gradient descent in the baseline. RAdam adjusts the variance of the adaptive learning rate [14].

The inference stage of the proposed model includes the same smoothing filters but of shorter length since the probabilistic outputs are obtained for 16s windows rather than 8s windows in the baseline.

C. *Datasets*

The dataset used for training consists of EEG recordings taken at the Neonatal Intensive Care Unit of Cork University Hospital of 72 neonates. Eighteen of these neonates experienced seizures from hypoxic ischemic encephalopathy brain injury. The training dataset totals 835 hours of multichannel EEG recordings, with a total of 77.7 hours of annotated seizure activity from 1,389 seizure events. Eight channels of EEG were recorded at 256Hz. The 10-20 EEG placement system modified for neonates was used with the following bipolar montage F4-C4, C4-O2, F3-C3, C3-O1, T4-C4, C4-Cz, Cz-C3, C3-T3. One hour of background activity was included from each of the remaining non-seizure patients. The total length of the training dataset was of 889 hours. The same dataset was also used in [3], [8], [11], [13] and for the baseline system. This study had full ethical approval from the Clinical Research Ethics Committee of the Cork Teaching Hospitals.

Two datasets are used to test the performance. One dataset of continuous multichannel EEG consists of data from 78 neonates with 23 experiencing seizure events. This clinical dataset comprises 4,570 hours of unedited EEG recordings, with 57.7 hours of seizure activity from 1,704 seizure events.

Table I. The comparison of results of the proposed model on the test dataset of continuous EEG and public DB from Helsinki.

	Continuous EEG		Helsinki DB [15]	
	AUC	AUC90	AUC	AUC90
Helsinki SVM [8]	-	-	95.5%*	-
Baseline [5]	95.4%	79.0%	95.6%	70.4%
This study	**97.7%**	**83.6%**	**96.4%**	**72.2%**
Relative improvement wrt Baseline	**50.5%**	**21.9%**	**21.9%**	**6.5%**

* The Helsinki SVM AUC from a Leave One patient Out analysis is reported, all other metrics are the result of a single held-out test set analysis.

The same eight channel bipolar montage was used to record the data at 256Hz or 200Hz.

A publicly available database is also used as a test dataset in this study [15]. This dataset consists of short 1-2h excerpts from 79 neonates who were admitted to the NICU of Helsinki University Hospital. This dataset contains a total of 112 hours of EEG recordings with 11 hours of seizure activity from 342 seizure events. Eighteen channels of EEG were recorded at 256 Hz. The dataset is further described and analysed in [4].

D. Performance assessment

AUC was used as the primary performance metric in this work. Sensitivity is the percentage of seizure epochs correctly labelled as seizure by the algorithm. The specificity is the percentage of non-seizure epochs correctly labelled as non-seizure by the algorithm. While AUC is a commonly used metric to compare the performances of seizure detection algorithms [16] irrespective of the chosen operating point, the AUC90 metric which calculates the performance of an algorithm with Specificity>0.9, is more representative of clinically acceptable operating conditions with a low number of false alarms. This metric is also reported for comparison.

III. Results

Table I compares the performance of the proposed model to the baseline model when tested on the dataset of continuous EEG and the Helsinki dataset. It can be seen that the proposed model with residual connections gives a steady and consistent boost in performance with respect to the baseline on both datasets tested. In particular, a relative improvement of 50.5% was obtained on the dataset of continuous EEG, where the AUC increased from 95.4% to 97.7%, with a similar increase in AUC90. On a publicly available dataset, the proposed method outperforms both the baseline and the SVM system developed in [8], with a relative improvement of over 20%. Fig. 4. illustrates the receiver operating characteristic (ROC) curves for the baseline and the proposed model which are obtained from the dataset of continuous EEG. The shaded portion of the AUC corresponds to AUC90, highlighting the increased performance of the proposed model in this important area. Fig. 5 shows per-patient increase in AUC on the patients with seizures from the database of continuous EEG.

IV. Discussion

Previously developed methods of neonatal seizure detection have relied on clearly identifiable preprocessing, feature extraction, classification and postprocessing blocks [2]

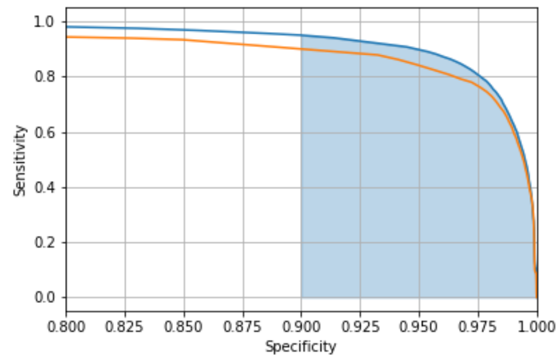

Figure 4. ROC with AUC90 performance (shaded blue) of the proposed model (blue) and the baseline model (orange).

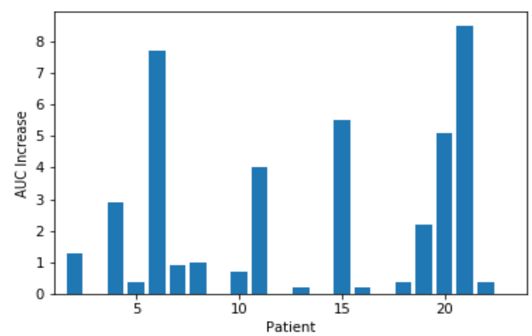

Figure 5. Per-patient AUC increase (absolute) of the proposed model vs. the baseline on the dataset of continuous EEG.

[3], [4]. The deep learning methods which employ convolutional filters have reduced the reliance on explicit feature engineering while significantly boosting the performance [8], [11]. More recently, architectures has been introduced which further reduced the reliance of the model on the availability of per-channel annotations, and can exploit available data to learn convolutional filters across time and EEG channels without learning patient-specific seizure location dependencies [8]. This method has shown a significant boost in generalisation abilities with respect to both the feature extraction based SVMs [2], [4] and the single EEG channel deep learning systems [8], [9], [11]. The postprocessing steps such as moving average smoothing filters were still designed and applied outside the deep learning methodology. One way to incorporate the postprocessing into the system is to learn longer convolutional filters by increasing the input length and the receptive field of the model. This study exploits an addition to the model architecture which is well-researched in deep learning community and allows a better gradient flow during training to enable efficient training of much deeper neural networks – residual connections [10].

When comparing the performance of deep learning methods, a large and diverse test corpus is required to increase the sensitivity of the performance to the architectural changes. In this work, a very large clinical dataset of continuous unedited EEG was utilised for testing. The corpus contains a large variety of artifacts naturally present in continuous recordings, a variety of seizure morphologies of different etiologies, and seizures of different lengths. This dataset can serve as a good representation of real-life clinical scenario which employs AI-assisted long term EEG monitoring in NICU.

The results on the continuous EEG database indicate that with the help of residual connections, data augmentation and a more robust optimizer, the proposed model trained on input windows of 16s has improved the relative performance by over 50%, with respect to the baseline which operates on 8s windows [8]. The proposed model has an increased receptive field of ~10s compared to the baseline model with the receptive field of ~5s, thus giving the proposed model the ability to learn longer temporal seizure characteristics. A neonatal seizure is often defined as an episode of abnormal electrographic activity with a minimum duration of ten seconds.

From the differences in AUC between the proposed model and the baseline model over the 23 patients with seizures from the database of continuous EEG (Fig. 5), it can be seen that the designed architecture consistently improves over the majority of patients with an increase from an average AUC of 96.0% to 97.8%. Interestingly, with the new model the performance has not decreased for a single patient.

The AUC90 metric captures the performance which is more relevant in a clinical setting, since the specificity of over 90% corresponds to a more clinically acceptable operating point where the number of false detections is smaller than one false detection per hour. Moving to a longer input with the proposed model, the AUC90 has similarly improved from an AUC of 79.0% to an AUC of 83.6%, as reflected in the shaded area in Fig. 4.

Benchmarking on the publicly available dataset, the performance has also improved from 95.6% to 96.4%, setting the new highest AUC score on this dataset, to date. Though this dataset consists of pre-selected 1-2h segments of multichannel neonatal EEG, this is the only publicly available dataset of neonatal EEG and it is important to benchmark the performance of the developed methods on this dataset.

The improvement in performance of the proposed model over the baseline model on this publicly available Helsinki dataset is not as large as that on the continuous EEG dataset as shown in Table I. The proposed model allows for more efficient training and exploitation of the training dataset which is long term unedited continuous EEG. These characteristics are better captured in the proposed model. While the better training procedure improves the overall performance on both test datasets, the mismatch in dataset characteristics between training and testing will inevitably increase the absolute difference in improvements.

V. Conclusion

The study has demonstrated that with the usage of residual connections, data augmentation and a robust optimizer, a deeper convolutional architecture can be constructed to operate on longer EEG input and reduce the reliance on the external postprocessing steps. The deeper network designed is capable of learning longer patterns from multichannel neonatal EEG through the increased receptive field. The new model has improved the performance of neonatal seizure detection on a large database of continuous EEG and a publicly available dataset. Future work will concentrate on further increasing the input length to eliminate the postprocessing steps altogether.

Acknowledgment

The authors wish to thank all the clinicians involved in this research particularly Prof. Geraldine Boylan and Dr Sean Mathieson of INFANT, University College Cork, Ireland.

References

[1] D. M. Murray, G. B. Boylan, I. Ali, C. A. Ryan, B. P. Murphy, and S. Connolly, “Defining the gap between electrographic seizure burden, clinical expression and staff recognition of neonatal seizures,” *Arch. Dis. Child. Fetal Neonatal Ed.*, vol. 93, no. 3, pp. F187-191, May 2008, doi: 10.1136/adc.2005.086314.

[2] A. Temko, E. Thomas, W. Marnane, G. Lightbody, and G. Boylan, “EEG-based neonatal seizure detection with Support Vector Machines,” *Clin. Neurophysiol.*, vol. 122, no. 3, pp. 464–473, Mar. 2011, doi: 10.1016/j.clinph.2010.06.034.

[3] A. M. Pavel *et al.*, “A machine-learning algorithm for neonatal seizure recognition: a multicentre, randomised, controlled trial,” *Lancet Child Adolesc. Health*, p. S235246422030239X, Aug. 2020, doi: 10.1016/S2352-4642(20)30239-X.

[4] K. T. Tapani, S. Vanhatalo, and N. J. Stevenson, “Time-Varying EEG Correlations Improve Automated Neonatal Seizure Detection,” *Int. J. Neural Syst.*, May 2019, doi: 10.1142/S0129065718500302.

[5] A. Craik, Y. He, and J. L. Contreras-Vidal, “Deep learning for electroencephalogram (EEG) classification tasks: a review,” *J. Neural Eng.*, vol. 16, no. 3, p. 031001, Jun. 2019, doi: 10.1088/1741-2552/ab0ab5.

[6] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, “Deep learning-based electroencephalography analysis: a systematic review,” *J. Neural Eng.*, vol. 16, no. 5, p. 051001, Aug. 2019, doi: 10.1088/1741-2552/ab260c.

[7] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, “Deep learning for time series classification: a review,” *Data Min. Knowl. Discov.*, vol. 33, no. 4, pp. 917–963, Jul. 2019, doi: 10.1007/s10618-019-00619-1.

[8] A. O’Shea, G. Lightbody, G. Boylan, and A. Temko, “Neonatal seizure detection from raw multi-channel EEG using a fully convolutional architecture,” *Neural Netw.*, vol. 123, pp. 12–25, Nov. 2019, doi: 10.1016/j.neunet.2019.11.023.

[9] A. O’Shea *et al.*, “Deep Learning for EEG Seizure Detection in Preterm Infants,” *Int. J. Neural Syst.*, p. 2150008, Dec. 2020, doi: 10.1142/S0129065721500088.

[10] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” *ArXiv151203385 Cs*, Dec. 2015, Accessed: Feb. 05, 2020. [Online]. Available: http://arxiv.org/abs/1512.03385.

[11] A. O’Shea, G. Lightbody, G. Boylan, and A. Temko, “Neonatal seizure detection using convolutional neural networks,” in *2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP)*, Tokyo, Sep. 2017, pp. 1–6, doi: 10.1109/MLSP.2017.8168193.

[12] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *ArXiv14091556 Cs*, Apr. 2015, Accessed: Mar. 30, 2021. [Online]. Available: http://arxiv.org/abs/1409.1556.

[13] A. Temko, G. Boylan, W. Marnane, and G. Lightbody, “Robust Neonatal EEG Seizure Detection through Adaptive Background Modeling,” *Int. J. Neural Syst.*, vol. 23, no. 04, p. 1350018, Aug. 2013, doi: 10.1142/S0129065713500184.

[14] L. Liu *et al.*, “On the Variance of the Adaptive Learning Rate and Beyond,” *ArXiv190803265 Cs Stat*, Apr. 2020, Accessed: Mar. 30, 2021. [Online]. Available: http://arxiv.org/abs/1908.03265.

[15] N. J. Stevenson, K. Tapani, L. Lauronen, and S. Vanhatalo, “A dataset of neonatal EEG recordings with seizure annotations,” *Sci. Data*, vol. 6, no. 1, Art. no. 1, Mar. 2019, doi: 10.1038/sdata.2019.39.

[16] A. Temko, E. Thomas, W. Marnane, G. Lightbody, and G. B. Boylan, “Performance assessment for EEG-based neonatal seizure detectors,” *Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol.*, vol. 122, no. 3, pp. 474–482, Mar. 2011, doi: 10.1016/j.clinph.2010.06.035.