Seizure Classification Using BERT NLP and a Comparison of Source Isolation Techniques with Two Different Time-Frequency Analysis

S. Davidson[1], N. McCallan[1], K.Y. Ng[1, 2], P. Biglarbeigi[3], D. Finlay[1], B.L. Lan[2] and J. McLaughlin[1]

1. Nanotechnology and Integrated BioEngineering Centre (NIBEC), Ulster University, UK.
2. School of Engineering, Monash University, Malaysia.
3. School of Science and Engineering, University of Dundee, UK.

{davidson-s18, mccallan-n2, mark.ng}@ulster.ac.uk, pbiglarbeigi001@dundee.ac.uk, d.finlay@ulster.ac.uk, lan.boon.leong@monash.edu, jad.mclaughlin@ulster.ac.uk

I. Introduction

Epilepsy is one of the most common neurological disorders in the world [1], affecting about 50 million people worldwide [2]. Epileptic seizures occur when millions of neurons are synchronously excited, resulting in a wave of electrical activity in the cerebral cortex [3]. Electroencephalography (EEG) is a noninvasive tool that measures cortical activity with millisecond temporal resolution. EEGs record the electrical potentials generated by the cerebral cortex nerve cells [4]. Therefore, this tool is commonly used for the analysis and detection of seizures [5]. Epilepsy causes many difficulties in relation to the quality of life of the patient. It is therefore vital that automatic detection algorithms exist to aid neurologists to accurately classify the different types of seizures. Roy et al. [10] used different machine learning techniques to achieve an average F1-score of 0.561 using 2 s windows whilst Vanabelle et al. [11] used 1 s windows and achieved an accuracy of 51.33%, which shows that reducing the time window would also decrease the accuracy of classification. This paper aims to show that an NLP can be used for hierarchical classification, following upon an earlier work on combining simple partial and complex partial seizures [9]. The second aim is to show a pipeline that can be used to separate the seizures back into their original labels using neural networks. This method is quick, effective, and requires less training.

II. Dataset

The Temple University Hospital (TUH) Seizure Corpus (TUSZ) is an open-source dataset with specific seizure classes that features a clinician's report on the patients during seizure and non-seizure periods with each recorded session [6]. The text files include details about the patients as well as the clinical interactions. The seizure signals are extracted from the TUSZ in the form of annonated European Data Format (EDF) files. The seizures are defined by the International League Against Epilepsy (ILAE) through the ILAE 2017 Classification of Seizure Types Checklist [7].

III. Methodology

Figure 1 shows an overview of the proposed model. The use of Natural Language Processing (NLP) to reduce the dimensionality of the dataset with a modified neural network to separate the grouped EEG signals is the main novelty of this design. As the TUSZ contains text and EDF files, different processing stages are needed, which will be further described in this manuscript.

IV. BERT Classification

A Bidirectional Encoder Representations from Transformers (BERT) NLP was used to categorise seizures into different groups by combining different seizures — the complex partial and simple partial becoming the focal all seizure, whilst the tonic and tonic-clonic seizures are combined into the tonic all seizure. The absence seizure was correctly classified by the BERT model and did not need a separate classification system. It was reported that the joining of the focal seizures could lead to a more stable classification system [9]. The joining of the tonic seizures is based upon similar responses observed by the clinician and possibly within the EEG. To improve the ability of the NLP to recognise different seizures, a genetic algorithm was

978-1-6654-7029-2/22/$31.00 ©2022 IEEE IEEE SPMB 2022 December 3, 2022

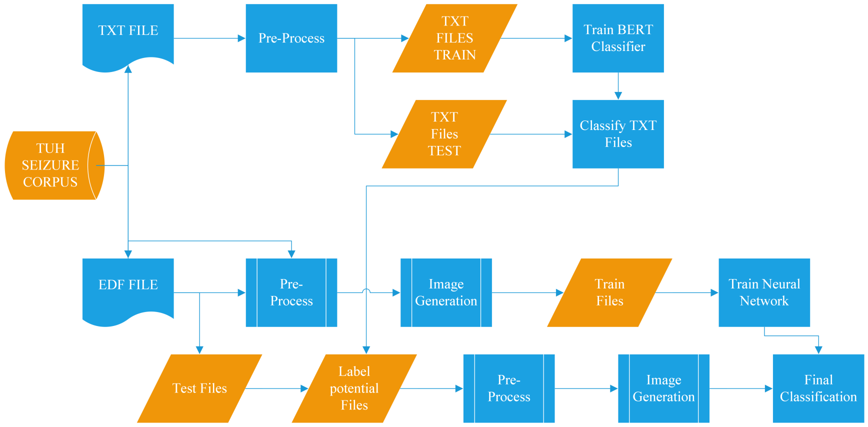


Figure 1. Overview of the system used in the proposed design in this research. In this case, the text files are used as a form of dimensionality reduction technique to reduce the need for image generation.

applied where instead of being utilised for feature selection, it now acts as a word selection tool.

V. Signal Classification

The EEGs were preprocessed using three source isolation techniques to remove noise: Empirical Mode Decomposition (EMD), Variable Mode Decomposition (VMD), and Empirical Wavelet Transform (EWT). Then, channel reduction was used to retain only 19 common EEG channels [9]. As the original signals were sampled at 250–1000 Hz, they were downsampled to 124 Hz using a Finite Impulse Response (FIR) anti-aliasing filter, which were then passed through one of the source isolation techniques, and subsequently windowed with a 1 s no overlap. The signals that were correctly classified by the BERT NLP were taken to the next stage. To separate the seizure groups, two commonly used time-frequency methods for seizure detection [8] were compared: CWT and the short-time Fourier transform (STFT) (see image generation stage in Figure 1). The absolute values of the outputs of the CWT and the STFT were obtained, with equal windowing length so that they can be combined into a matrix. The time-frequency representations are better at characterising the functions within the brain and hence, are better for analysing epileptic seizures [12]. This was then fed forward into a modified neural network, where the third channel of the matrix acts as the channel representation of the data in the format [frequency, time, channel]. The frequency and time act as the (x,y) components of the time-frequency value, which is then expanded for each of the 19 channels processed. The neural network was designed based on the Resnet-18 model, but downsized and matched with the outputs of the time-frequency representations, and then trained using the Adaptive Moment Estimation (ADAM) optimisation algorithm with 50 epochs, piecewise gradient decrease, and an initial learn rate of 0.0001.

VI. Results and Discussion

Figure 2 shows that the BERT NLP model can correctly separate the absence seizure from the tonic all and focal all seizures. The absence seizure is unique compared to the two other groups because it is always generalised and the age of the patients are significantly lower. As the BERT model can learn in dual directions, its understanding of the terminology used to describe absence seizures is clear. However, as focal all and tonic all seizures can both have focal onsets, they could be recorded using different descriptions in the clinicians' reports. This has a negative impact on the BERT classifier as it encodes texts with numerical values, i.e. words with similar meanings could be assigned different numerical values by the BERT classifier. As

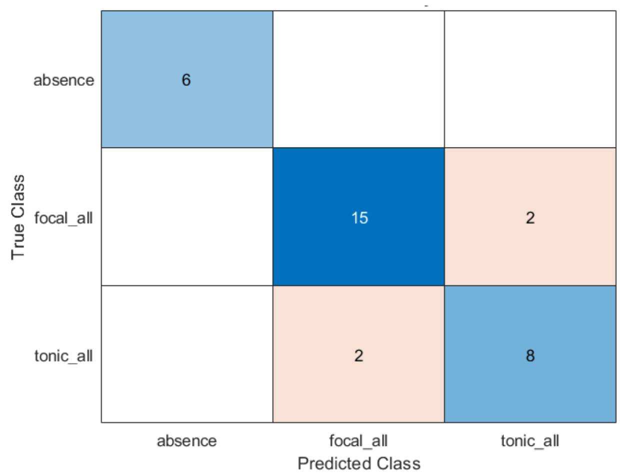


Figure 2. Confusion chart showing the BERT model and its ability to correctly classify seizures using a topographical approach.

such, misclassification of the same labels could occur in the model. Each model was measured using

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+TN+FP}, \qquad \text{F1}_{\text{score}} = \frac{TP}{TP+\frac{1}{2}(FP+FN)}, \tag{1}$$

where TP, TN, FN, and FP are the true positive, true negative, false negative, and false positive, respectively. Twelve models were compared; six for the separation of tonic and tonic-clonic, and six for the separation of simple partial and complex partial. In each separation, there are three CWTs and three STFTs analyses. Also, for each separation, there is one analysis for EMD, VMD, and EWT, respectively. This provides an overview of how each time-frequency representation can be used for the separation of the seizures. Table 1 is the results of the 12 models used in this experiment, where the results with the highest accuracy are highlighted in bold. Table 1 also shows that the STFT outperformed the CWT in two of the three investigations, which contradicts usual findings where CWT is the superior method. This could be due to the STFT using the same windowing boundary conditions for each seizure, which allowed the neural network to learn patterns faster compared to the CWT. Table 1 also shows that EMD outperformed the EWT and VMD in all cases except for the STFT with EWT, which could also possibly suggest that perhaps VMD is not recommended as a source isolation technique for seizure classification.

VII. Conclusions and Future Work

An NLP is needed to reduce dimensionality of the data, whilst also using a new pipeline for two seizure type classification. This paper has shown that an NLP can enhance seizure classification by reducing the need to convert original signals into a form more suited for either neural networks or machine learning. This reduction in the size of the dataset can enhance seizure detection in the future, especially if grouping individual seizure types will lead to a more general outlook on the patient's records. This paper has also shown that the STFT is more robust in extracting individual seizures from the group compared to CWT. Potential future work include using an NLP to classify all seizure types as there is currently a challenge in using NLP as some labels share a class in the dataset. Hence, a regression style network is needed.

References

[1] M.J. Brodie, S.C. Schachter, and P. Kwan, *Fast Facts: Epilepsy*, Karger Med. and Sci. Publishers, 2012.

[2] World Health Organisation, Key facts: Epilepsy, 2019. (Accessed: 11/06/2021)

Table 1. Results of the neural network using a modified Resnet-18 model.

Time-Frequency	Source Isolation	Combined Label	Seizure Type	Accuracy	F1-score
CWT	EWT	Tonic all		75.59%	
			Tonic		0.67
			Tonic-Clonic		0.81
		Focal		16.06%	
			Simple Partial		0.01
			Complex Partial		0.27
	EMD	Tonic all		**78.17%**	
			Tonic		0.72
			Tonic-Clonic		0.81
		Focal		75.12%	
			Simple Partial		0.72
			Complex Partial		0.76
	VMD	Tonic all		75.52%	
			Tonic		0.70
			Tonic-Clonic		0.78
		Focal		**75.83%**	
			Simple Partial		0.70
			Complex Partial		0.79
STFT	EWT	Tonic all		**80.38%**	
			Tonic		0.70
			Tonic-Clonic		0.83
		Focal		74.43%	
			Simple Partial		0.62
			Complex Partial		0.76
	EMD	Tonic all		80.02%	
			Tonic		0.76
			Tonic-Clonic		0.87
		Focal		**78.00%**	
			Simple Partial		0.72
			Complex Partial		0.80
	VMD	Tonic all		71.58%	
			Tonic		0.63
			Tonic-Clonic		0.76
		Focal		77.3%	
			Simple Partial		0.70
			Complex Partial		0.81

[3] A.T. Tzallas, M.G. Tsipouras, and D.I. Fotiadis, "Epileptic Seizure Detection in EEGs Using Time–Frequency Analysis", *IEEE Trans. Inf. Technol. Biomed.*, 13(5):703–710, 2009.

[4] A. Subasi and E. Erçelebi. "Classification of EEG Signals Using Neural Network and Logistic Regression", *Comput. Methods Programs Biomed.*, 78(2):87–99, 2005.

[5] H. Adeli, Z. Zhou, and N. Dadmehr, "Analysis of EEG Records in an Epileptic Patient Using Wavelet Transform", *J. Neurosci. Methods*, 123(1):69–87, 2003.

[6] V. Shah et al., "The Temple University Hospital Seizure Detection Corpus", *Front. Neuroinf.*, 12:83, 2018.

[7] R.S. Fisher et al., "Instruction Manual for the ILAE 2017 Operational Classification of Seizure Types", *Epilepsia*, 58(4):531–542, 2017

[8] M. Li, W. Chen, and T. Zhang, "Automatic epileptic EEG detection using DT-CWT-based non-linear features", *Biomed. Signal Process. Control*, 34:114–125, 2017.

[9] S. Davidson et al., "Epileptic Seizure Classification Using Combined Labels and a Genetic Algorithm", *IEEE MELECON*, pp. 430–435, 2022.

[10] S. Roy et al., "Seizure type classification using EEG signals and machine learning: Setting a benchmark", *IEEE SPMB*, pp. 1–6, 2020.

[11] P. Vanabelle et al., “Epileptic seizure detection using EEG signals and extreme gradient boosting”, *J. of Biomed. Res.*, 34(3):228, 2020.

[12] S. Morales and M.E. Bowers, “Time-frequency analysis methods and their application in developmental EEG data”, *Dev. Cognit. Neurosci.*, 54:101067, 2022.

Seizure Classification Using BERT NLP and a Comparison of Source Isolation Techniques with Two Different Time-Frequency Analysis

S. Davidson, N. McCallan, K.Y. Ng, P. Biglarbeigi, D. Finlay, B.L. Lan and J. McLaughlin

Ulster University | NIBEC Nanotechnology & Integrated Bioengineering Centre

2022 IEEE Signal Processing in Medicine and Biology Symposium

IEEE

Abstract

- The dataset used is the TUH seizure corpus
- Hierarchical classification using NLP and then a neural network.
- Combination of the simple partial and complex partial into focal-all and combining tonic and tonic-clonic into tonic-all along with the absence seizure. Performed to aid in the NLP classification strategy.
- A modified Neural Network is used to uncombine the signals
- A comparison of the STFT and the CWT with source isolation.
- STFT outperforms in two thirds of the experiments

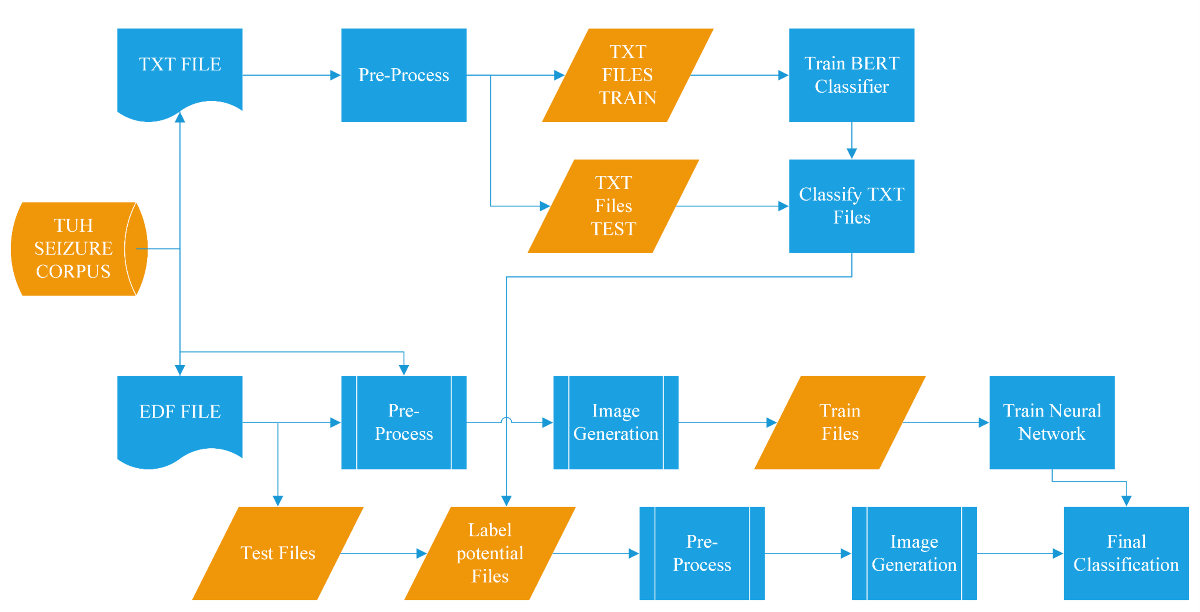
Introduction

- Epilepsy affects 50 million people worldwide[1].
- EEG's are non-invasive tools that measure the cortical activity with millisecond temporal resolution.
- Roy et al. [2] used different machine learning techniques with two second windows achieving an average F1-score of 0.561
- Vanbelle et al. [3] used 1 second windows to achieve an average accuracy of 51.33%.
- 12 models are compared each using the accuracy and the F1-score.

Dataset

- Neurologist reports are stored as text files and the EEG signals are stored as EDF files. Both are labelled as the TUH Seizure corpus.
- 5 labels are used; Absence, Complex Partial, Simple Partial, Tonic and Tonic-clonic.
- In the NLP the labels are Absence, Focal all and Tonic all

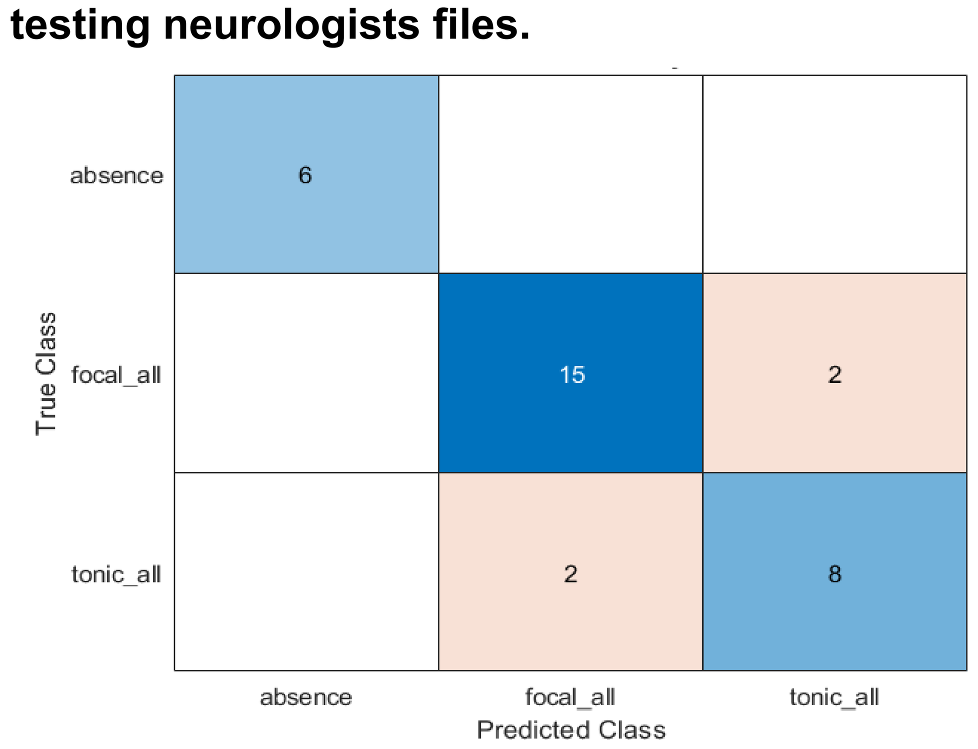
Overview Of the Model



- The model uses the same seizure corpus. In one direction is takes the text files pre-processes the words. Splits them into the pre-defined training and testing. to allow for the classification of the seizures.
- The second pipeline takes the EDF signals and pre-processes the training and testing signals in the same definition as the training and testing split of the data.
- Source isolation is a technique to remove noise within the signal. EEG's typically contain a lot of different artifacts.
- The pre-processing of the signal is firstly done by removing unwanted channels. The signal is then downsampled to 124 Hz. The downsampled signal is fed to three different source isolation techniques. Empirical Mode Decomposition (EMD), Variable Mode Decomposition (VMD) and Empirical Wavelet Transform (EWT). Then split into one second windows with no overlap.
- The labels the BERT classifier gives are then passed to the testing algorithm only the associated files that are correctly classified by the BERT model are then feed into the neural network.

NLP with BERT

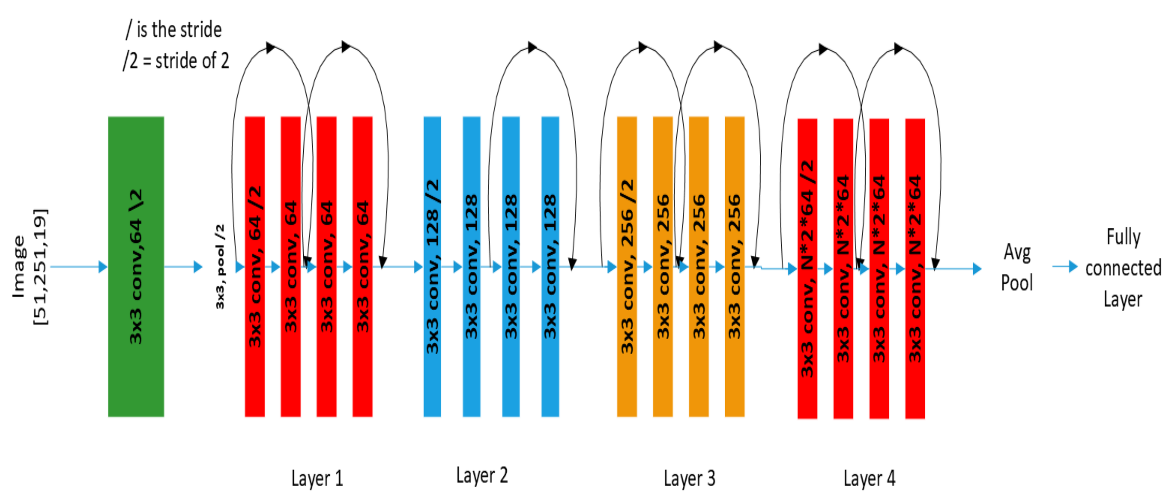
- Natural Language Processing (NLP) is a technique that combines text with Artificial Intelligence (AI).
- Bidirectional Encoder Representations from Transformers (BERT) is a new technique developed at Google.
- The genetic algorithm takes the words and places a token on them. The genetic algorithm then iterates through the words selection. The words that supply the highest classification on the training data are then used on the testing procedure.
- The training and testing split is performed on the pre-defined TUH training and testing split.
- The Confusion chart shows the classification of the testing neurologists files.

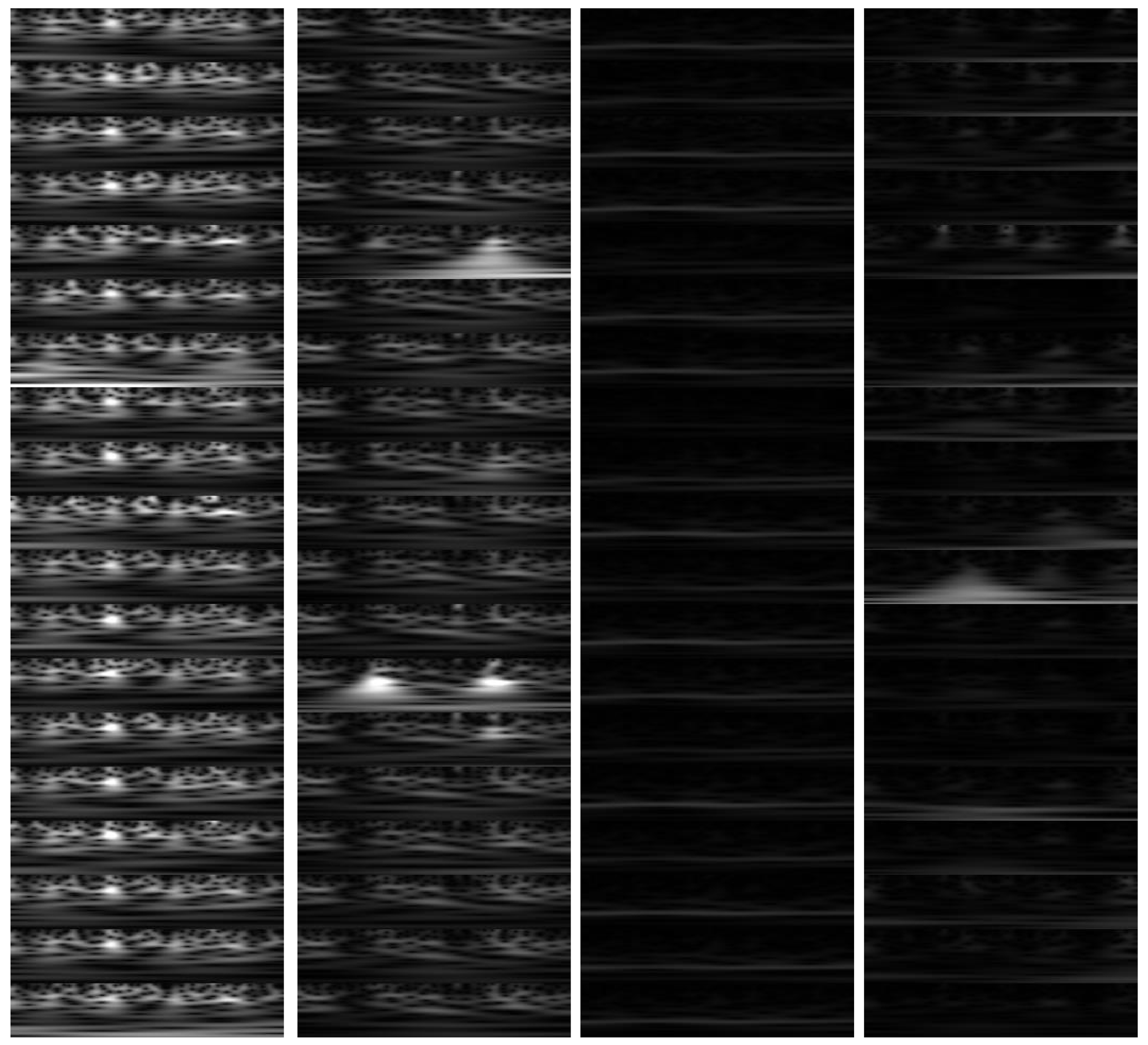


- Absence seizure is correctly classified. The age and waveform are possible words that allow for accurate classification. The tonic-all and focal-all are combined to allow for hierarchical classification.
- The dimensionality of the data can be reduced by using the BERT to classify the text before.
- The confusion between the tonic-all and focal-all could in the focal nature that exists in both of these labels. Motor responses also could be similar.

Neural Network

- The Neural Network (NN) is a shallower resnet-18 that has been modified to use the RBG channel to act as the channels of the EEG. This gives a representation of [frequency, time, channel].
- Adam Optimization, 50 epochs, Piecewise gradient descent, initial learn rate of 0.001.
- The images shown are an example of the cwt input where the channels are vertically stacked. In the model they are concatenated in the channel method. The normalization technique is done on a per window basis.
- The training and testing split is given by the TUH dataset. The testing files only contain files that the NLP has correctly classified.





Tonic-clonic | Tonic | Complex partial | Simple partial

- The reduction of the ResNet model increases the processing speed of the classification.
- ResNet uses skip connections that is one method to mitigate the vanishing gradient problem. These are shown in the arrows that pass over some of blocks.

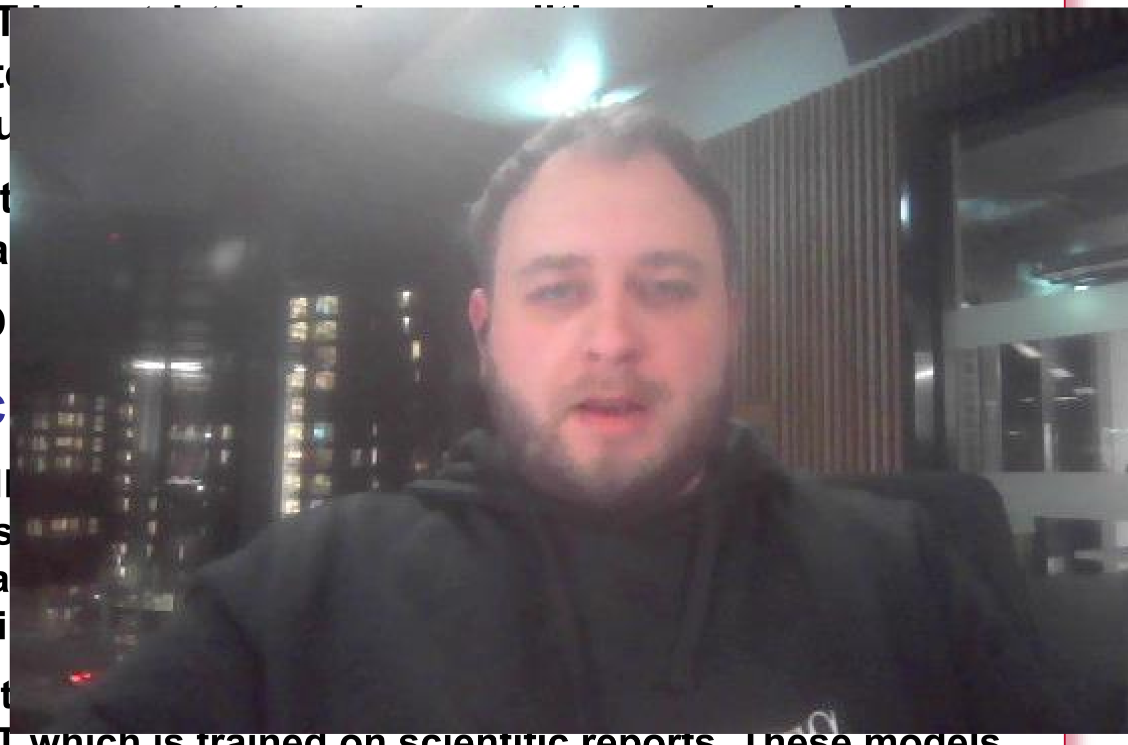
References

1. World Health Organisation, Key facts: Epilepsy, 2019. (Accessed: 11/06/2021)
2. S. Roy et al., "Seizure type classification using EEG signals and machine learning: Setting a bench-mark", IEEE SPMB, pp. 1–6, 2020.
3. P. Vanabelle et al., "Epileptic seizure detection using EEG signals and extreme gradient boosting",J. of Biomed. Res., 34(3):228, 2020

Results and Discussion

Time-Frequency	Source Isolation	Combined Label	Seizure Type	Accuracy (%)	F1-Score
CWT	EWT	Tonic all		75.59	
			Tonic		0.67
			Tonic-Clonic		0.81
		Focal all		16.06	
			Simple Partial		0.01
			Complex Partial		0.27
	EMD	Tonic all		**78.17**	
			Tonic		0.72
			Tonic-clonic		0.81
		Focal all		75.12	
			Simple partial		0.72
			Complex partial		0.76
	VMD	Tonic all		75.52	
			Tonic		0.70
			Tonic-Clonic		0.78
		Focal all		75.83	
			Simple Partial		0.70
			Complex Partial		0.79
STFT	EWT	Tonic all		80.38	
			Tonic		0.70
			Tonic-Clonic		0.83
		Focal all		74.43	
			Simple Partial		0.62
			Complex Partial		0.76
	EMD	Tonic all		80.02	
			Tonic		0.76
			Tonic-Clonic		0.87
		Focal all		**78.00**	
			Simple Partial		0.72
			Complex Partial		0.80
	VMD	Tonic all		71.58	
			Tonic		0.63
			Tonic-Clonic		0.76
		Focal all		77.30	
			Simple Partial		0.70
			Complex Partial		0.81

- $Accuracy = \frac{TP+TN}{TP+FN+TN+FP}, F1 = \frac{TP}{TP+\frac{1}{2}(FP+FN)}$
- Short Time Fourier Transform (STFT) outperformed the CWT in the EWT and EMD. In the VMD the CWT outperformed the STFT.
- There is variability in the results. For example in the STFT with the EWT outperformed the CWT with the EWT however, only the first IMF is taken.
- STFT plott freq
- The t sepa
- VMD



Conc

- An N datas to tra requi
- NLP t BERT which is trained on scientific reports. These models could train faster and provide a higher classification score.
- The input into the neural network is done with short windows. Where epochs of low energy can be represented in the entire duration of the signal.

Seizure Classification Using BERT NLP and a Comparison of Source Isolation Techniques with Two Different Time-Frequency Analysis

S. Davidson, N. McCallan, K.Y. Ng, P. Biglarbeigi, D. Finlay, B.L. Lan and J. McLaughlin

2022 IEEE Signal Processing in Medicine and Biology Symposium

Abstract

- The dataset used is the TUH seizure corpus
- Hierarchical classification using NLP and then a neural network.
- Combination of the simple partial and complex partial into focal-all and combining tonic and tonic-clonic into tonic-all along with the absence seizure. Performed to aid in the NLP classification strategy.
- A modified Neural Network is used to uncombine the signals
- A comparison of the STFT and the CWT with source isolation.
- STFT outperforms in two thirds of the experiments

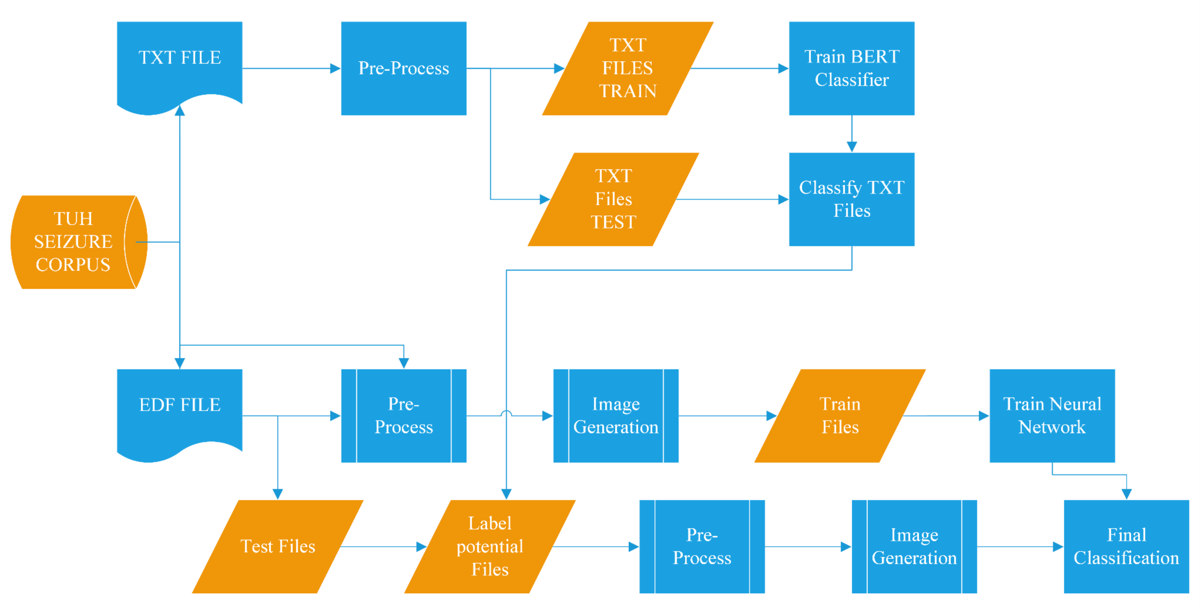
Introduction

- Epilepsy affects 50 million people worldwide[1].
- EEG's are non-invasive tools that measure the cortical activity with millisecond temporal resolution.
- Roy et al. [2] used different machine learning techniques with two second windows achieving an average F1-score of 0.561
- Vanbelle et al. [3] used 1 second windows to achieve an average accuracy of 51.33%.
- 12 models are compared each using the accuracy and the F1-score.

Dataset

- Neurologist reports are stored as text files and the EEG signals are stored as EDF files. Both are labelled as the TUH Seizure corpus.
- 5 labels are used; Absence, Complex Partial, Simple Partial, Tonic and Tonic-clonic.
- In the NLP the labels are Absence, Focal all and Tonic all

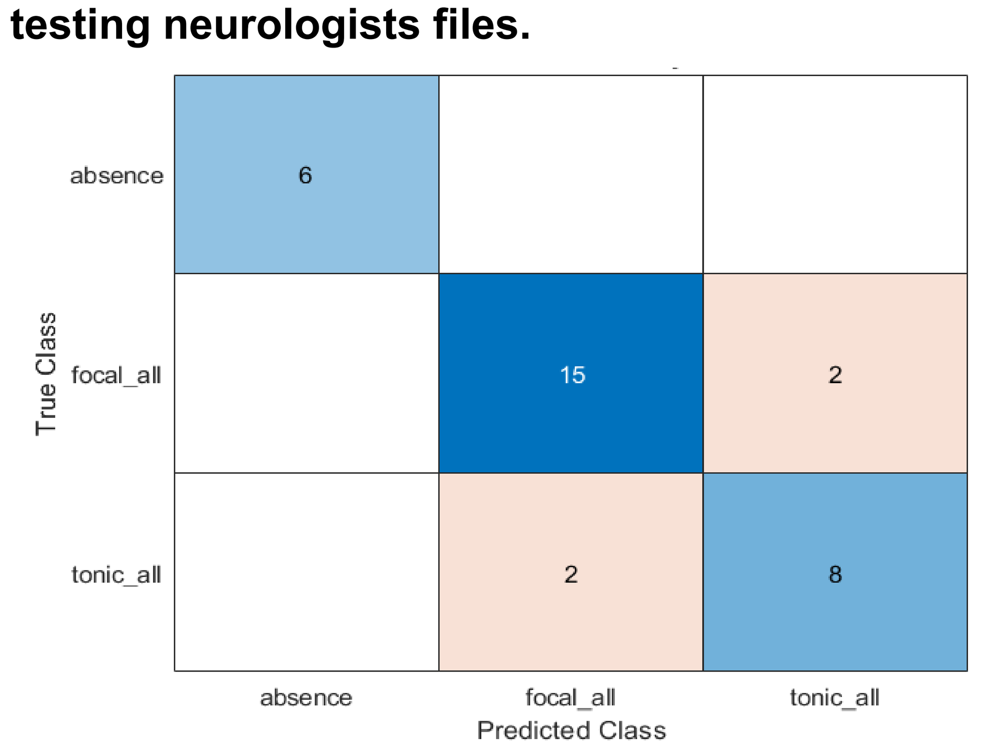
Overview Of the Model



- The model uses the same seizure corpus. In one direction is takes the text files pre-processes the words. Splits them into the pre-defined training and testing. to allow for the classification of the seizures.
- The second pipeline takes the EDF signals and pre-processes the training and testing signals in the same definition as the training and testing split of the data.
- Source isolation is a technique to remove noise within the signal. EEG's typically contain a lot of different artifacts.
- The pre-processing of the signal is firstly done by removing unwanted channels. The signal is then downsampled to 124 Hz. The downsampled signal is fed to three different source isolation techniques. Empirical Mode Decomposition (EMD), Variable Mode Decomposition (VMD) and Empirical Wavelet Transform (EWT). Then split into one second windows with no overlap.
- The labels the BERT classifier gives are then passed to the testing algorithm only the associated files that are correctly classified by the BERT model are then feed into the neural network.

NLP with BERT

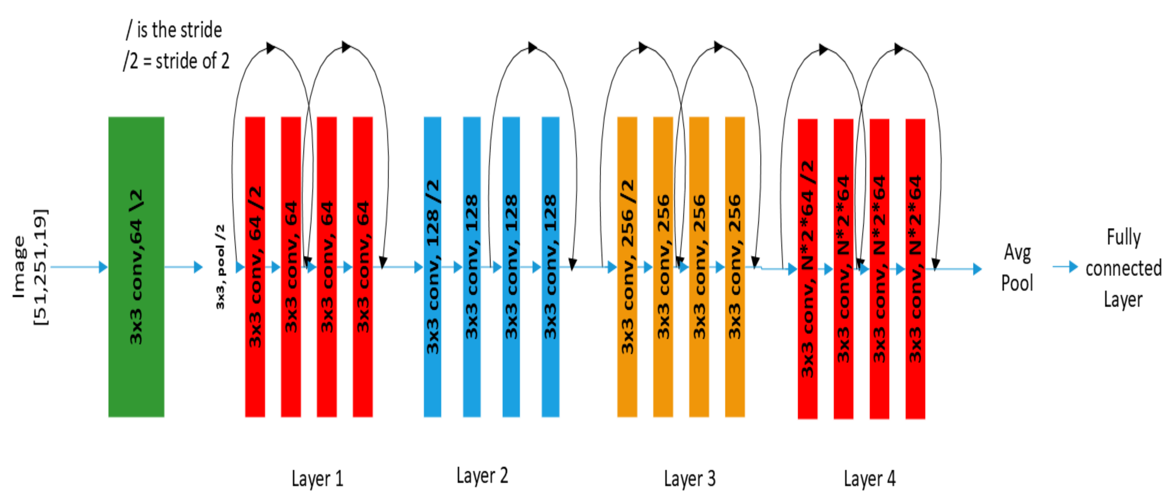
- Natural Language Processing (NLP) is a technique that combines text with Artificial Intelligence (AI).
- Bidirectional Encoder Representations from Transformers (BERT) is a new technique developed at Google.
- The genetic algorithm takes the words and places a token on them. The genetic algorithm then iterates through the words selection. The words that supply the highest classification on the training data are then used on the testing procedure.
- The training and testing split is performed on the pre-defined TUH training and testing split.
- The Confusion chart shows the classification of the testing neurologists files.

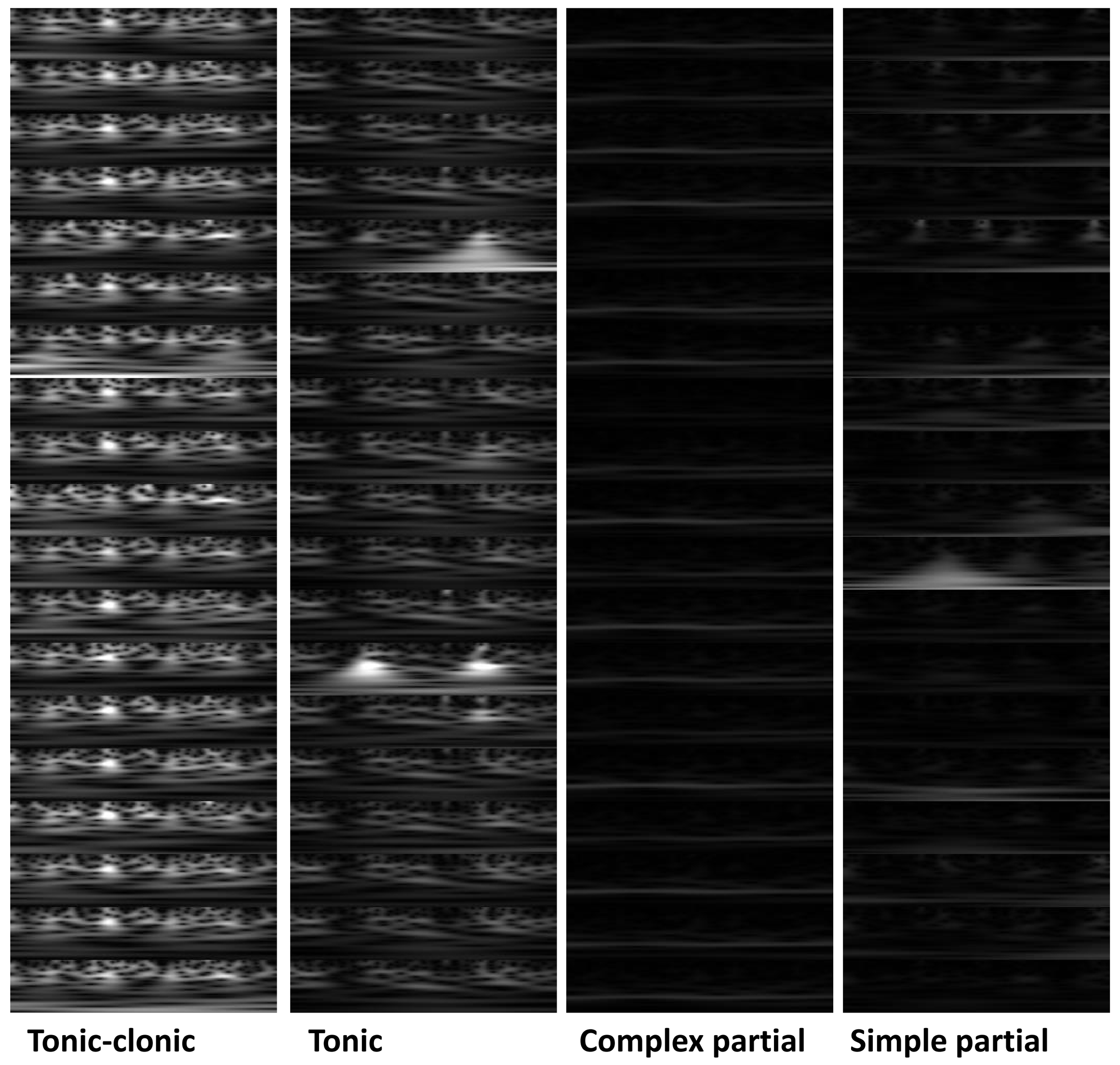


- Absence seizure is correctly classified. The age and waveform are possible words that allow for accurate classification. The tonic-all and focal-all are combined to allow for hierarchical classification.
- The dimensionality of the data can be reduced by using the BERT to classify the text before.
- The confusion between the tonic-all and focal-all could in the focal nature that exists in both of these labels. Motor responses also could be similar.

Neural Network

- The Neural Network (NN) is a shallower resnet-18 that has been modified to use the RBG channel to act as the channels of the EEG. This gives a representation of [frequency, time, channel].
- Adam Optimization, 50 epochs, Piecewise gradient descent, initial learn rate of 0.001.
- The images shown are an example of the cwt input where the channels are vertically stacked. In the model they are concatenated in the channel method. The normalization technique is done on a per window basis.
- The training and testing split is given by the TUH dataset. The testing files only contain files that the NLP has correctly classified.





- The reduction of the ResNet model increases the processing speed of the classification.
- ResNet uses skip connections that is one method to mitigate the vanishing gradient problem. These are shown in the arrows that pass over some of blocks.

References

1. World Health Organisation, Key facts: Epilepsy, 2019. (Accessed: 11/06/2021)
2. S. Roy et al., "Seizure type classification using EEG signals and machine learning: Setting a bench-mark", IEEE SPMB, pp. 1–6, 2020.
3. P. Vanabelle et al., "Epileptic seizure detection using EEG signals and extreme gradient boosting",J. of Biomed. Res., 34(3):228, 2020

Results and Discussion

Time-Frequency	Source Isolation	Combined Label	Seizure Type	Accuracy (%)	F1-Score
CWT	EWT	Tonic all		75.59	
			Tonic		0.67
			Tonic-Clonic		0.81
		Focal all		16.06	
			Simple Partial		0.01
			Complex Partial		0.27
	EMD	Tonic all		**78.17**	
			Tonic		0.72
			Tonic-clonic		0.81
		Focal all		75.12	
			Simple partial		0.72
			Complex partial		0.76
	VMD	Tonic all		75.52	
			Tonic		0.70
			Tonic-Clonic		0.78
		Focal all		75.83	
			Simple Partial		0.70
			Complex Partial		0.79
STFT	EWT	Tonic all		80.38	
			Tonic		0.70
			Tonic-Clonic		0.83
		Focal all		74.43	
			Simple Partial		0.62
			Complex Partial		0.76
	EMD	Tonic all		80.02	
			Tonic		0.76
			Tonic-Clonic		0.87
		Focal all		**78.00**	
			Simple Partial		0.72
			Complex Partial		0.80
	VMD	Tonic all		71.58	
			Tonic		0.63
			Tonic-Clonic		0.76
		Focal all		77.30	
			Simple Partial		0.70
			Complex Partial		0.81

- $Accuracy = \frac{TP+TN}{TP+FN+TN+FP}, F1 = \frac{TP}{TP+\frac{1}{2}(FP+FN)}$
- Short Time Fourier Transform (STFT) outperformed the CWT in the EWT and EMD. In the VMD the CWT outperformed the STFT.
- There is variability in the results. For example in the STFT with the EWT outperformed the CWT with the EWT however, only the first IMF is taken.
- STFT has strict boundary conditions when being plotted. The CWT has edge effects in the low frequencies.
- The tonic and tonic-clonic seizures are easier to separate because of the higher accuracies.
- VMD weakest performer in this analysis technique

Conclusion

- An NLP can be used for dimensionality reduction of the dataset. In the case of the absence seizure there is no need to transform the seizure. The combining of seizure does require a separation algorithm.
- NLP technology is still evolving with new models like Sci-BERT which is trained on scientific reports. These models could train faster and provide a higher classification score.
- The input into the neural network is done with short windows. Where epochs of low energy can be represented in the entire duration of the signal.